Sparse Portfolios and Benchmarking in Corporate Bond Markets

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

First draft: March 4, 2026

Abstract

We use detailed data on fixed-income benchmark indexes in Canada and the United States to provide systematic evidence of how benchmarking shapes corporate bond ownership and prices. Funds hold sparse portfolios, and index weights strongly influence which bonds active and passive funds select. We rationalize these patterns in a model with benchmarked managers who face portfolio management costs, which predicts which assets managers optimally include in their portfolios. In the model, a bond's price increases with its benchmarking intensity (BMI)—a measure of the amount of fund capital benchmarked against the bond—while portfolio sparsity attenuates this price impact for excluded bonds. Exploiting discontinuities in benchmarked assets around bond maturity cutoffs, we show that increases in bonds' BMIs lead to reductions in yield spreads and increases in fund ownership—but only for bonds predicted to enter sparse portfolios.

JEL Classification: G11, G12, G23

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

---

*We thank seminar participants at the 6th MBS – Southampton – Surrey Virtual Research Seminar, Bocconi University, DAM Working Group, Georgia State University, Johns Hopkins University, Michigan State University, Warwick Business School, Università della Svizzera italiana, and WAPFIN at NYU Stern for their helpful comments. We are grateful to Yuan Gao, Nilesh Hegde, Austin Liu, Daniel Yoon, and especially Jinrui Wang for excellent research assistance.

*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 United States, 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. It is common in the literature to model the asset allocation decisions of fixed-income funds using a standard portfolio optimization framework. Yet, funds face unique incentives and constraints, influencing both their portfolio composition and, ultimately, asset prices. In this paper, we highlight two key aspects of fixed-income portfolio 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.

Our research question is to assess how benchmarking affects investment-grade corporate bond ownership and prices. Using data covering most fixed-income benchmarks in Canada and the United States, 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 and higher benchmark weight. 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 only for bonds predicted to enter sparse portfolios.

We first present stylized facts on fixed-income benchmark indexes used by mutual funds and ETFs in the United States and Canada. We find that in both countries, there is an increasing heterogeneity of benchmarks, catering to different investment mandates and styles. In the United States, most fixed-income assets are benchmarked to Bloomberg indexes, while Canadian funds primarily use FTSE benchmarks. To our knowledge, ours is the first study to systematically describe benchmarks in these markets. In contrast to equity benchmarks, bond benchmarks so far have received limited attention in the literature. 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 prohibitively expensive for fund managers to hold every bond in their benchmark index. Despite the high turnover levels, fund managers steer performance of their funds close to their benchmark indexes, as evidenced by their tracking error levels.

Our next set of stylized facts concerns fund portfolios. 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—up to 26% in the U.S. and 46% in Canada. 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 that even within narrowly defined categories, there is sparsity.

In the second part of the paper, we propose a model that rationalizes these new stylized facts and derives testable asset pricing implications in an equilibrium with sparse portfolios. The model features fund managers who are benchmarked to an index and face portfolio management costs, which we interpret as rebalancing, monitoring, and other portfolio management costs. These costs increase with the number of assets included in the portfolio, preventing managers from holding all benchmark constituents and resulting in sparse portfolios. When adding an asset, a manager trades off the marginal cost of managing a larger portfolio against the benefit of reducing tracking error variance and increasing expected return. Accordingly, the manager excludes assets with low benchmark weights, given their small impact on tracking error.

The model highlights asset substitutability as another determinant of which assets are excluded. Specifically, managers are more likely to exclude assets for which there are many close substitutes. We find that, within each category of closely substitutable assets, managers are likely to hold ones with higher benchmark weights, larger size, and higher idiosyncratic risk.

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 empirically test these predictions for investment-grade corporate bonds in Canada and the United States. 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—particularly 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 within-bond variation in BMI comes from crossing maturity-based cutoffs. Our identification strategy exploits index reconstitutions, which create quasi-exogenous variation in this measure. Specifically, we adopt a research design inspired by Bretscher, Schmid, and Ye (2024) that leverages discontinuities arising when bonds cross maturity cutoffs specified in funds' mandates.

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 United States. In contrast, we find no such effects for excluded bonds. While this may seem at odds with the literature documenting stronger demand effects among smaller assets (Wurgler and Zhuravskaya, 2002), it aligns with our theoretical framework, in which included assets are less substitutable.

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. Whereas the 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 observed only for bonds predicted to be included in sparse portfolios.

To address alternative interpretations of the importance of bond size for our results, we exploit sharp ownership discontinuities around benchmark size-eligibility cutoffs in both Canada and the United States. Even for the smallest bonds, there is a discrete jump in fund ownership exactly at the threshold for benchmark inclusion. This pattern holds for both active and passive funds, which provides additional evidence in support of our mechanism.

Related Literature. Our paper contributes to several strands of 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), Buffa, Vayanos, and Woolley (2022), Kashyap, Kovrijnykh, Li, and Pavlova (2021), and Pavlova and Sikorskaya (2023) study the implications of benchmarking for portfolio management and asset prices. However, none of these models produces portfolio sparsity and explores how it affects asset prices. Furthermore, our model is related to the literature on portfolio underdiversification, as sparsity may result from underdiversification. Previous research proposes various mechanisms to explain underdiversification, such as ambiguity aversion (Dow and Werlang, 1992) or limited information-processing capacity (Van Nieuwerburgh and Veldkamp, 2010). Our mechanism is distinct. The paper closest to ours in this strand is Merton (1987), which assumes that investors exogenously neglect certain stocks and studies the resulting implications for asset prices. In contrast, our contribution is to endogenously determine which assets managers exclude from their portfolios and analyze how benchmarking considerations influence this decision.

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 ETFs context by Shim and Todorov (2022), Brogaard, Heath, and Huang (2024), Koont, Ma, Pastor, and Zeng (2025), who note that ETF creation/redemption baskets exclude some (illiquid) assets and attribute this to a tradeoff 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. Dick-Nielsen and Rossi (2019), Bretscher, Schmid, Sen, and Sharma (2023), Bretscher, Schmid, and Ye (2024), Chaudhary, Fu, and Li (2024), Gabaix, Koijen, Richmond, and Yogo (2025) shed light on the nature of institutional investors' portfolios in this market, but do not discuss the importance of fixed-income benchmarks in asset allocation. One exception is Ottonello (2019), whose focus is on high-yield bond prices and funds' liquidity management.[1] We document results consistent with Ottonello's findings but utilize a fundamentally different identification strategy based on plausibly exogenous changes in BMIs of investment-grade bonds. Several papers in this literature examine corporate bond mutual funds and ETFs. Dannhauser and Hoseinzade (2022), and Ma, Xiao, and Zeng (2022) examine active choices of these investors and explore how these choices may contribute to financial fragility. The paper closest to ours in this literature is Bretscher, Schmid, and Ye (2024), which attributes bond price changes around index maturity cutoffs to the inelastic demand of passive funds. We depart from their analysis in several important ways. First, we show that active funds also contribute to inelastic demand. Second, we show that pricing and ownership effects are concentrated among bonds predicted to be included in sparse portfolios. Finally, methodologically, rather than using a maturity-cutoff dummy

[1] Recent evidence in market microstructure 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).

as an instrument for fund ownership, we employ BMI—a continuous measure that provides substantially richer variation than a cutoff indicator.

2 Stylized Facts

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

2.1 Corporate bond investment management

Mutual funds and exchange-traded funds (ETFs) have become increasingly important holders of corporate bonds, particularly in the investment-grade (IG) 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. Fund managers in this sector almost universally operate relative to benchmark indexes, which serve two purposes: they define the investment universe for eligible holdings and provide the yardstick 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 ratings 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 saliently 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 their white papers and fund materials, emphasizing that their 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 (London Stock Exchange Group) 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 using Refinitiv to further include variables such as Return Index and bid-ask spread. We use Bank of Canada's bond yield curves to construct maturity-matched credit spreads.[3] We collect mutual fund and ETF holdings and fund static and dynamic characteristics from Morningstar, including a snapshot of fund 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

[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/.

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 significantly contribute to the AUM (Huang, Lu, Song, and Xiang (2023)). We use CRSP to classify funds into active and passive, as described in Appendix A.10.2. We further extract a dynamic panel of each funds' primary prospectus benchmark over time directly from regulatory filings reported on U.S. SEC EDGAR and augmented with a Morningstar snapshot and U.S. SEC Mutual Fund Risk and Return database. More details and validation exercises on this are reported in Appendix A.10. 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.

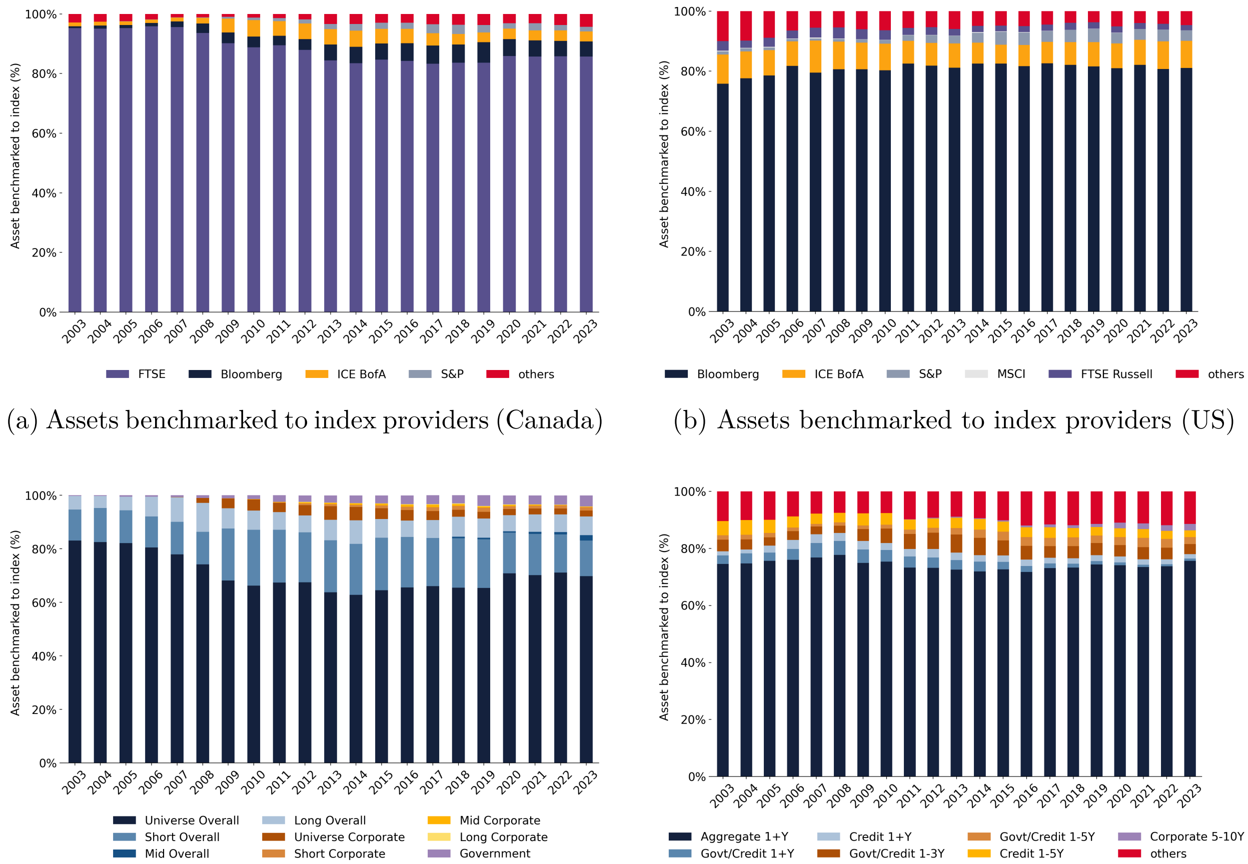
In Canada, we only consider corporate bonds by selecting 'corporate' FTSE industry sector. Within that sector, there are 7 industry groups (communication, energy, financial, industrial, infrastructure, real estate, and securitization) and 21 finer industry subgroups. In the U.S., we filter for corporate 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 United States.

2.3 Stylized facts

In this section, we document new stylized facts on corporate bond benchmarks and funds portfolios in Canada and the United States.

Among the universe of fixed-income benchmarks adopted by mutual funds and ETFs in Canada, over 90% of total assets is benchmarked to FTSE indexes (Figure 1a). Therefore, our detailed benchmark data from FTSE covers most of the assets of those institutional investors. Importantly, within FTSE Canada bond index family, there is an entire suite of indexes—such as, for example, short-term, mid-term, long-term, and aggregate—catering

Figure 1: Fund assets benchmarked to bond indexes in Canada and the United States



(a) Assets benchmarked to index providers (Canada)

(b) Assets benchmarked to index providers (US)

(c) Assets benchmarked to FTSE indexes (Canada)

(d) Assets benchmarked to Bloomberg indexes (US)

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.

to different investment objectives. The assets under management benchmarked to each of them are shown in Figure 1c. In the US, on the other hand, FTSE Russell is very small, and the market is dominated by Bloomberg indexes. Despite slightly more heterogeneity than demonstrated in Canada, Bloomberg still accounts for the dominating 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 5–10Y 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 46% (71%), on average, for the Canadian (U.S.) bond indexes in our sample, which is a magnitude difference from the turnover of around 10% for S&P 500 index, as shown in Table 1. 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 46%, but still fewer than passive equity funds would. Active bond funds, on the other hand, hold very concentrated portfolios with only 10% of the benchmark bonds. As Table 1 illustrates, the corporate bond universe in the U.S. is a magnitude larger than Canada, so we see even more sparsity. US funds on average hold only 7% of bonds in their benchmark indexes. Passive funds hold 26%, while active funds hold only around 4% of benchmark bonds.

This sparsity is also very persistent. As shown in the transition matrix in Table 2, an asset that is not held by a fund on average has over 96% (89%) 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 quicker, with 54–61% of bonds still being held at 3-year horizon, which likely reflects high turnover documented above.

Despite the high turnover and sparsity levels, fund managers in our sample manage to steer performance of their funds close to their benchmark indexes. The average annualized tracking error of passive funds in our Canadian (U.S.) sample is well below 1% at 59 bps (92 bps). Even for active funds, tracking errors are well below their equity fund counterparts at 116 bps (207 bps) for Canada (U.S.). Within all types of funds, there is considerable cross-sectional variation, with standard deviations of tracking error always exceeding the averages we report here. Details of our tracking error calculation as well as additional statistics are reported in Appendix A.5.

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

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.

Table 2: 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	97.8%	2.2%	96.1%	3.9%	93.5%	6.5%	89.2%	10.8%
Held in last period	7.6%	92.4%	13.7%	86.3%	22.7%	77.3%	43.2%	56.8%
Panel B: US								
Not held in last period	99.4%	0.6%	98.9%	1.1%	98.1%	1.9%	96.5%	3.5%
Held in last period	8.0%	92.0%	13.6%	86.4%	22.6%	77.4%	45.6%	54.4%

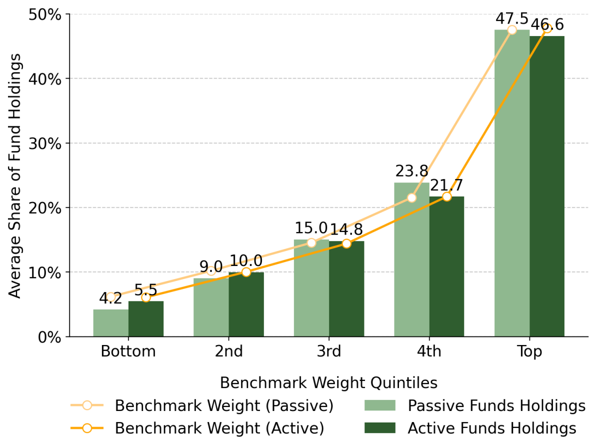
This table reports transition probabilities for corporate bond holdings of mutual funds and ETFs (Canada and U.S.). Rows are conditioned on last period's holding status; columns show the status after the indicated horizon. The reference universe is the broadest benchmark available in each geography,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

how close that portfolio is to the benchmark. Even at this aggregate level, portfolios are sparse, with active funds holding, on average, 66%–73% of benchmark bonds and passive funds holding 76%–90%. Appendix A.6 reports these statistics as well as active shares and tracking errors of the aggregate portfolios.

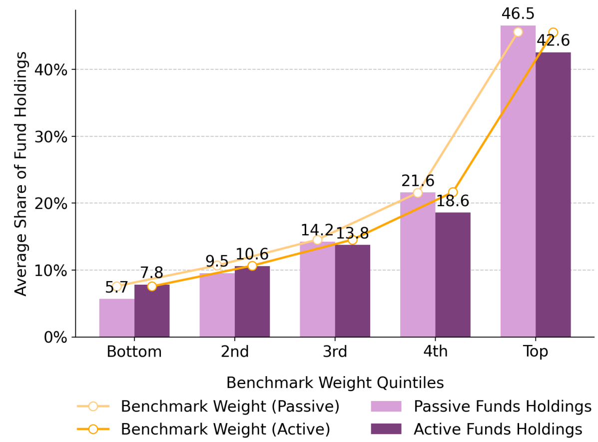
Next, we describe how benchmark index weights affect bond portfolio holdings and their sparsity. Figure 2 characterizes asset allocation across benchmark weight quintiles. We see that funds typically allocate larger shares of assets to bonds with larger index weights, with over 40% of fund assets allocated 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 the smaller fractions of benchmark bonds in the lower benchmark weight quintiles. Moreover, benchmark weights are as 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 3 reports the

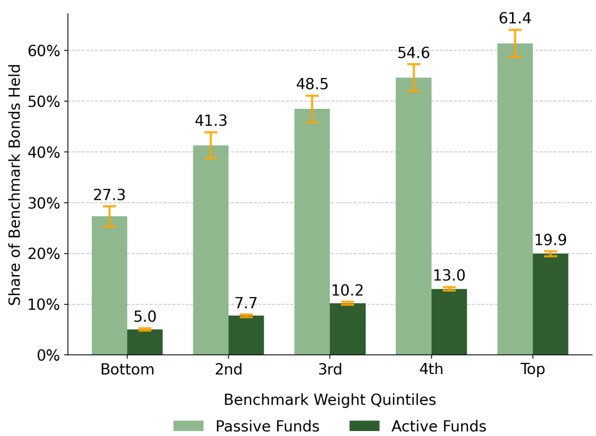
Figure 2: Fund holdings across benchmark weight quintiles



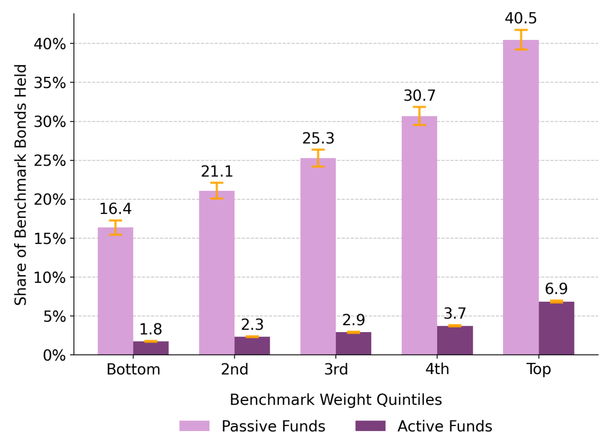
(a) Share of fund holdings across benchmark weight quintiles (Canada)



(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 corporate bond part of fund and benchmark portfolios.

results of panel regressions of fund holdings on 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, so 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 bond). We take the broadest benchmark index bonds as a reference universe for the extensive margin analysis, that is, all corporate bonds

Table 3: Regressions of fund holdings on benchmark index membership and weights for Canada and US

	Fund weight | fund weight $\neq 0$			**Dummy(fund weight > 0)**		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Canada						
Dummy(within benchmark)		0.002***	0.0003		0.264***	0.155***
		(0.0003)	(0.0003)		(0.009)	(0.008)
Benchmark weight			0.727***			47.87***
			(0.073)			(2.95)
Observations	436,651	436,651	436,651	3,582,170	3,582,170	3,582,170
R^2	74.6%	74.6%	74.7%	24.6%	28.3%	28.8%
Panel B: United States						
Dummy(within benchmark)		0.090***	0.030***		0.154***	0.129***
		(0.005)	(0.007)		(0.003)	(0.004)
Benchmark weight			0.586***			45.57***
			(0.053)			(2.686)
Observations	5,088,906	5,088,906	5,088,906	117,069,613	117,069,613	117,069,613
R^2	8.2%	8.2%	8.2%	15.3%	18.2%	18.3%

This table reports estimates of the conditional correlation between fund weight or propensity to hold and benchmark membership dummy variable and benchmark index weight. Panel A reports the results for Canada, and Panel B for the US. 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. All regressions are saturated by including fund-by-date and bond-by-date fixed effects. Standard errors clustered at the bond and year-month levels are presented in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

in 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 3 show that both Canadian and US 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 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 A8 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 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 the funds' portfolio choice decisions, on both whether or not to include a bond in the portfolio and 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

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

where $Z_m \sim \mathcal{N}(0, \sigma_m^2)$ are common shocks (factors) and $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is an idiosyncratic one. All shocks are orthogonal to each other. For simplicity, there are only two common shocks. The $N \times 1$ vectors $D \equiv (D_1, \ldots, D_N)'$ and $p \equiv (p_1, \ldots, p_N)'$ denote vectors of period-1 cash flows and period-0 risky asset prices, respectively. Period-1 risky asset prices equal D. 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). We denote $\mu \equiv (\mu_1, \ldots, \mu_N)'$ and $\boldsymbol{\beta}_m \equiv (\beta_{m,1}, \ldots, \beta_{m,N})'$.

There are two types of investors: direct investors and fund managers. Direct investors, who represent a fraction λ_D of the investor population, manage their own portfolios. In contrast, fund managers manage portfolios on behalf of fund investors. Fund investors can directly purchase the riskless asset but do not trade stocks themselves; instead, they delegate portfolio selection to fund managers, who receive compensation 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, and the manager selects a portfolio consisting of θ shares to maximize 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. The term Cn in (2) represents portfolio management costs that scale with the number of assets (n) included in the fund's portfolio. 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—independent of the number of shares traded. Such costs are presumably lower for index funds than for actively managed funds, given the lower frequency of trades and reduced monitoring requirements. Finally, the parameters a and b capture the manager's compensation sensitivity to absolute and relative portfolio performance, respectively.

3.1 Investor Optimization Problems

The direct investor's problem at time $t = 0$ is

$$\max_{\theta_D} E[-\exp\{-\gamma W\}]. \tag{3}$$

It is well-known that this problem reduces to the following mean-variance optimization (see Appendix B.1):

$$\max_{\theta_D} \quad \theta_D'(\mu - p) - \frac{1}{2}\gamma \underbrace{\theta_D' \Sigma \theta_D}_{\text{Variance}}. \tag{4}$$

It follows that the optimization problem of a fund manager can be represented as $\max_{\theta,n} E[-\exp\{-\gamma(a+b)\theta'(D-p) - b\omega'(D-p))\}] - Cn$, which is equivalent to the following mean-tracking-error-variance optimization problem (see Appendix B.1):

$$\max_{\theta} \quad (a+b)\theta'(\mu - p) - \frac{1}{2}\gamma \underbrace{((a+b)\theta - b\omega)'\Sigma((a+b)\theta - b\omega)}_{\text{Tracking error variance}} - Cn. \tag{5}$$

The first term implies 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. 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 tradeoff is the portfolio management cost, which discourages including too many assets in the fund's portfolio.

3.2 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 size $\overline{\theta}$—a realistic assumption given that bond indexes are typically value-weighted. We normalize the weights in the fund's benchmark to be $\omega = 1/W_F\overline{\theta}$, where W_F denotes the fund's initial assets under management delegated by investors.

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*

$$\theta_{D,i} = \frac{1}{\gamma\sigma_\epsilon^2}(\mu_i - p_i), \tag{6}$$

$$\theta_i = \begin{cases} \frac{1}{\gamma\sigma_\epsilon^2(a+b)}(\mu_i - p_i) + \frac{b}{a+b}\omega_i, & \textit{if } i \leq n^* \quad \textit{(included by managers)} \\ 0, & \textit{if } n^* < i \leq N \quad \textit{(excluded by managers),} \end{cases} \tag{7}$$

and the equilibrium asset prices are

$$p_i = \begin{cases} \mu_i - \gamma\frac{1}{\lambda_D + \frac{\lambda_F}{a+b}}\sigma_\epsilon^2\left(\overline{\theta}_i - \frac{b}{a+b}\underbrace{\lambda_F\omega_i}_{BMI_i}\right), & i \leq n^* \quad \textit{(included by managers)} \\ \mu_i - \gamma\frac{1}{\lambda_D}\sigma_\epsilon^2\overline{\theta}_i, \quad n^* < i \leq N & \textit{(excluded by managers),} \end{cases} \tag{8}$$

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

The optimal number of assets in the funds' portfolios, n^*, satisfies the following inequalities:

$$\frac{\gamma}{2}\sigma_\epsilon^2\left(\frac{\overline{\theta}_{n^*}}{\lambda_D} + b\,\omega_{n^*}\right)^2 - C \geq 0, \tag{9}$$

$$\frac{\gamma}{2}\sigma_\epsilon^2\left(\frac{\overline{\theta}_{n^*+1}}{\lambda_D} + b\,\omega_{n^*+1}\right)^2 - C < 0. \tag{10}$$

To develop an 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 (7), 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.

To understand the nuances of the tradeoff the manager faces when deciding whether to add a marginal asset to the portfolio, it is useful to examine the value function of a fund manager whose portfolio size is fixed at n assets. For any subset $\mathsf{I} \subset \{1, \ldots, N\}$, let $V(\mathsf{I})$

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

denote the fund manager's maximized objective value in (5) subject to $\theta_i = 0$ for all $i \notin \mathsf{I}$. Because assets are ordered by decreasing benchmark weights and managers optimally include the top-ranked assets, we write $V(n) \equiv V(\{1, \ldots, n\})$.

$$V(n) \equiv \max_{\theta} \;\; (a+b)\underbrace{\theta'(\mu - p)}_{\text{Expected return}} - \frac{1}{2}\gamma \underbrace{((a+b)\theta - b\,\omega)'\Sigma((a+b)\theta - b\omega)}_{\text{Tracking error variance}} - \underbrace{C\,n}_{\text{Cost}} \tag{11}$$

When adding an asset, a manager balances the marginal benefit of increasing the portfolio's expected return and reducing its tracking error variance with the marginal cost of managing a larger portfolio. The inequalities determining the optimal number of assets n^* presented in (9)–(10) are equivalent to

$$V(n^*) - V(n^* - 1) = \frac{\gamma}{2}\sigma_\epsilon^2 \left(\frac{\overline{\theta}_{n^*}}{\lambda_D} + b\,\omega_{n^*}\right)^2 - C \geq 0, \tag{12}$$

$$V(n^* + 1) - V(n^*) = \frac{\gamma}{2}\sigma_\epsilon^2 \left(\frac{\overline{\theta}_{n^*+1}}{\lambda_D} + b\,\omega_{n^*+1}\right)^2 - C < 0, \tag{13}$$

respectively. The first terms in (9)–(10) represent the marginal benefit of including an asset, while the last term captures the marginal cost, which is simply a per-asset portfolio management cost C. As (9)–(10) 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. We have already discussed that assets with higher benchmark weights contribute more to reducing the fund's tracking error. Idiosyncratic risk acts in a similar way: if an asset has a high idiosyncratic risk, including it in the portfolio reduces the fund's tracking error.[5] Furthermore, an asset's size is positively related to the likelihood of its inclusion in the portfolio because larger assets in our model have higher expected returns. Larger assets constitute a greater share of the market and therefore have higher betas, making them riskier and requiring higher expected returns as compensation for risk.

We next discuss asset prices in the presence of portfolio sparsity. Since funds optimally

[5] Although we assume for parsimony that all assets share the same idiosyncratic risk, equations (9)–(10) 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.

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 (8). 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. Specifically, there may exist an equilibrium in which a fraction $\phi \in (0, 1)$ of managers holds $n^* + 1$ assets (i.e., $1, \ldots, n^* + 1$) while the remaining fraction $1 - \phi$ holds n^* assets (i.e., $1, \ldots, n^*$). In this mixed equilibrium, the first n^* assets are held by all managers, whereas asset $n^* + 1$ is included only by a fraction ϕ of managers. Equilibrium prices coincide with Proposition 1 for assets $i \leq n^*$ and for $i > n^* + 1$, while the price of the marginal asset $n^* + 1$ reflects the reduced effective mass $\phi\lambda_F$ of managers demanding it. The mixing probability ϕ is pinned down by the indifference condition $V(n^*) = V(n^* + 1)$ (equivalently, the incremental value of adding asset $n^* + 1$ equals zero). Appendix B.3 provides the derivation and verification.

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.

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.

3.3 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 $\beta_{m,i} > 0$.

The problem of 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:

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

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

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*

$$\theta_i = \begin{cases} \frac{\mu_1 - p_1}{\gamma\sigma_1^2(a+b)} + \frac{b}{a+b}\omega_1 + \underbrace{\frac{b}{a+b}\frac{\rho_{12}\sigma_1\sigma_2}{\sigma_1^2}\omega_2}_{\text{Additional demand}}, & \text{if } i = 1, \\ 0, & \text{if } i = 2. \end{cases} \tag{15}$$

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

$$\theta = \frac{1}{\gamma(a+b)}\Sigma^{-1}(\mu - p) + \frac{b}{a+b}\omega \tag{16}$$

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 3.2, 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*

$$p_1 = \mu_1 - \gamma A\left[\sigma_1^2\,\overline{\theta}_1 + \rho_{12}\sigma_1\sigma_2\,\overline{\theta}_2 - \frac{b}{a+b}\left(\sigma_1^2\underbrace{\lambda_F\omega_1}_{BMI_1} + \rho_{12}\sigma_1\sigma_2\,\lambda_F\omega_2\right)\right], \tag{17}$$

$$\begin{aligned} p_2 = \mu_2 - \gamma A&\left[\rho_{12}\sigma_1\sigma_2\,\overline{\theta}_1 + \rho_{12}^2\sigma_2^2\,\overline{\theta}_2 - \frac{b}{a+b}\left(\rho_{12}\sigma_1\sigma_2\,\lambda_F\omega_1 + \rho_{12}^2\sigma_2^2\underbrace{\lambda_F\omega_2}_{BMI_2}\right)\right] \\ &- \frac{\gamma}{\lambda_D}\,\sigma_2^2(1-\rho_{12}^2)\,\overline{\theta}_2, \end{aligned} \tag{18}$$

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 (17)*. 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 3.2, 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 (15), which depends on Asset 2's benchmark weight ω_2.

However, assets in 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 tradeoff 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 maximized objective value in (5) 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 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 (12) for the uncorrelated case flows. Specifically, the marginal benefit of including Asset 2 in the portfolio is higher if the asset has (i) higher benchmark weights, (ii) larger size, (iii) higher idiosyncratic risk. The marginal cost of including Asset 2 is C. The additional consideration that the correlated case brings is the influence of asset substitutability, as measured by the correlation ρ_{12}. Lemma 3 reveals that the more substitutable Assets 1 and 2 are, the lower the marginal benefit to the manager of adding Asset 2 to her portfolio. This is intuitive because the manager can substitute for holding Asset 2 by holding more of Asset 1, which saves on the portfolio management costs, while still delivering a similar expected return and tracking error variance reduction.

For illustrative purposes, we have assumed in this section that fund managers first include Asset 1 in their portfolio 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 tradeoff 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.)

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & 0 & 0 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & \rho_{34}\sigma_3\sigma_4 \\ 0 & 0 & \rho_{34}\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}, \tag{19}$$

where $\sigma_i^2 \equiv \mathrm{Var}(D_i)$ is the (unconditional) cash-flow variance of asset i, $\rho_{12} \equiv \mathrm{Corr}(D_1, D_2)$, and $\rho_{34} \equiv \mathrm{Corr}(D_3, D_4)$. 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).[6] 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.

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 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.

[6] 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.

4.1 Empirical measure of benchmarking intensity

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

$$BMI_{i,t} = \frac{\sum_{j=1}^{J} \lambda_{j,t} \times \omega_{i,j,t}}{MV_{i,t}}, \tag{20}$$

where $\lambda_{j,t}$ is the total assets under management (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 is therefore the same as how BMI was first defined for equities (Pavlova and Sikorskaya (2023)).

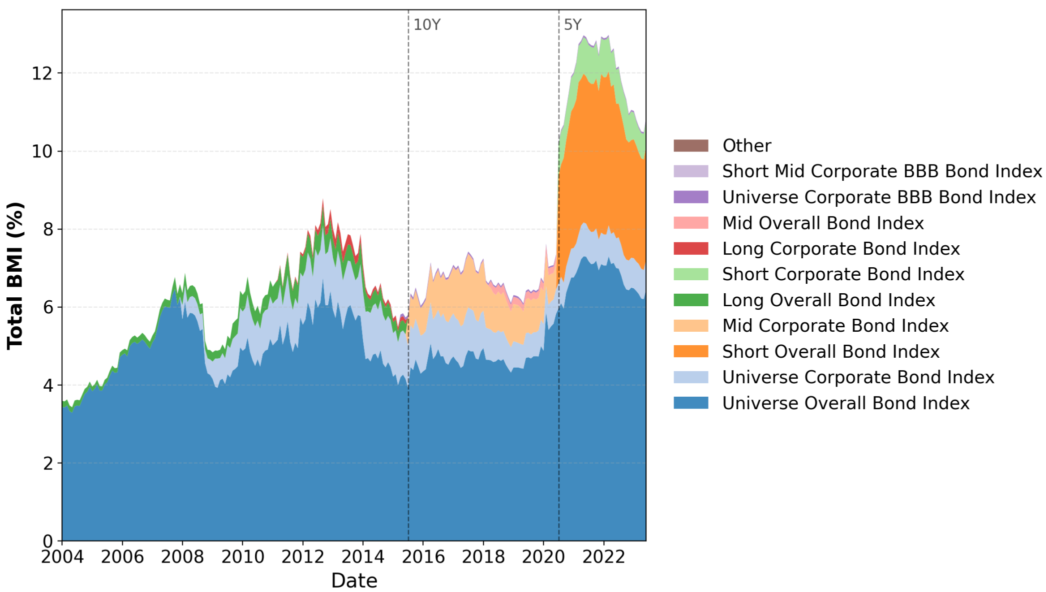
This measure captures the amount of benchmarked capital effectively allocated to a bond and reflects heterogeneity in benchmark index coverage. Economically, BMI measures the fraction of a bond's outstanding market value that would be held by benchmark-constrained capital under full index replication. 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 10%. These statistics mask considerable heterogeneity by bond maturity, as we show below.

Figure 3 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.

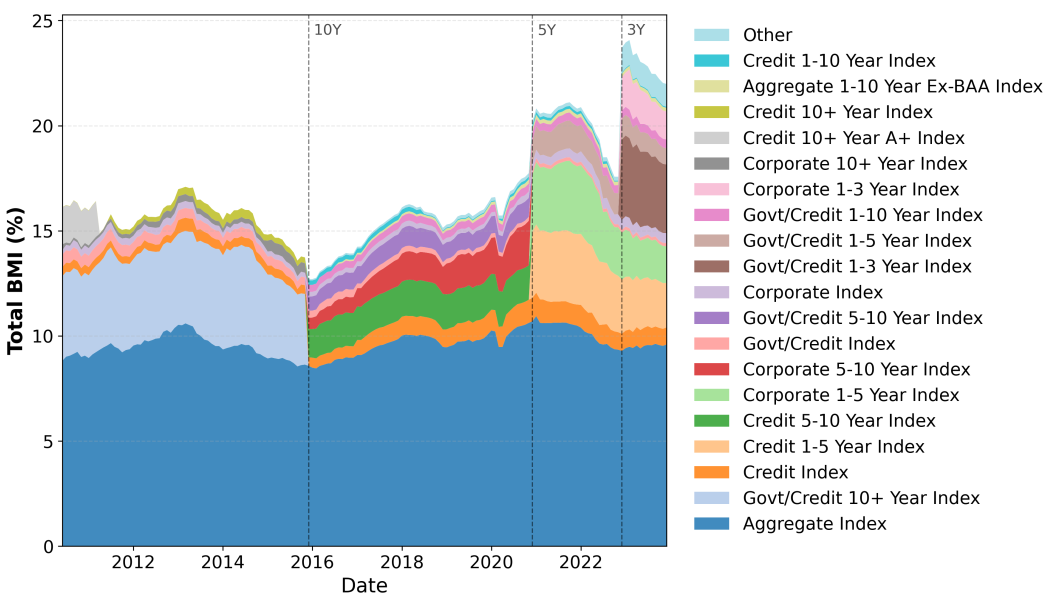
Because indexes in our sample are value-weighted and do not use float factors, equation (20) simplifies mechanically. Substituting the definition of index weights and canceling the bond's own market value yields

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

Figure 3: BMI composition of individual bonds



(a) Canada



(b) United States

This figure illustrates the composition of BMI by bond's time to maturity in (a) Canada and (b) the United States. 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.

where $IndexMV_{j,t}$ is the market value of index j, $\lambda_{j,t}$ is benchmarked AUM as defined around (20), and $D_{i,j,t} = 1$ 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 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 United States. 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.

Both in Canada and the U.S., the benchmarked AUM in the mutual fund and ETF sector strongly varies by bond maturity. Figure 4 shows that this variation in assets generates discontinuities in BMI around the 5-year maturity cutoff: an average bond sees a 2%–3% increase in BMI as it crosses the cutoff. These discontinuities are sizeable and strongly significant statistically. In Appendix C.1, we formally estimate their size by regressing the changes in BMI on a 5-year switch dummy variable.

An experiment that exploits 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. corporate bond market. We use a similar 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

$$\Delta Y_{i,t+h} = \boldsymbol{\beta} \Delta BMI_{i,t} + \zeta' \bar{X}_{i,t} + \nu_i + \mu_t + \epsilon_{i,t+h}. \tag{21}$$

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 $t+1$.[7] Bond-level controls in $\bar{X}_{i,t}$ include logarithm of 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 bond returns from WRDS. ν_i and μ_t are bond and year-month fixed effects. We cluster standard errors by issuer and year-month. Finally, we estimate this specification only in the neighborhood of the 5-year cutoff, namely, for bonds with between 4 and 6 years to maturity, and only consider bonds that were traded in a given month.[8]

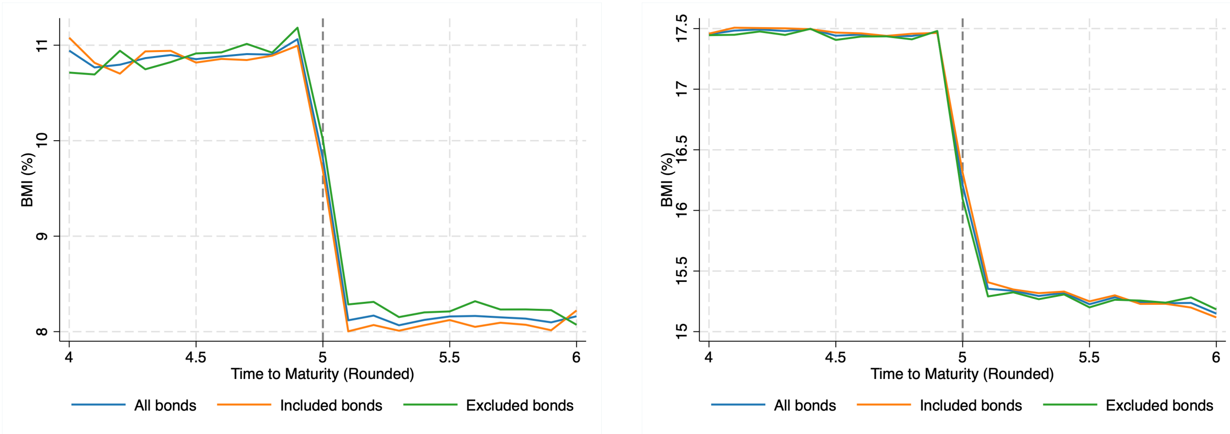
Our identifying assumption is that, conditional on this set of controls and fixed effects, changes in BMI are plausibly exogenous to bond returns and changes in fund ownership. Since changes in bond index membership arouns maturity cutoffs are predetermined, our design effectively exploits treatment intensity around a sharp treatment cutoff. 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.[9] Other types of bonds experience sharp drops

[7]Bloomberg bond indexes get rebalanced at the start of the month, while FTSE indexes get 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.

[8]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 Li (2023) and Chaudhary (2024)), 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.

[9]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.

Figure 4: Discontinuity in BMI around the 5-year maturity cutoff



(a) Canada

(b) United States

This figure illustrates the average benchmarking intensity by bond's time to maturity around the 5-year mark in (a) Canada and (b) the United States. We report the average BMIs within 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.

in their option-adjusted duration at the 5-year maturity cutoff. One potential reason is that a 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 an average effect of crossing the cutoff across all bonds is attenuated to zero. We address this challenge by estimating specification (21) in subsamples of included and excluded bonds.

Following our 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 consider 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 it belongs to and as *included* otherwise. 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 our 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.

Table 4 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 United States.

Table 4: Changes in BMI and bond pricing (all bonds)

	Change in the pricing variable over horizon				
	$t+1$	$t+3$	$t+6$	$t+9$	$t+12$
Panel A: United States					
Total return	0.033**	0.039**	0.030	0.049*	0.047*
	(0.013)	(0.016)	(0.022)	(0.026)	(0.028)
Observations	86,217	85,930	85,034	84,200	83,337
R-squared	0.731	0.744	0.812	0.862	0.892
OAS	−0.519***	−0.854***	−0.526	−0.555	−0.427
	(0.143)	(0.241)	(0.323)	(0.372)	(0.394)
Observations	86,217	85,930	85,043	84,216	83,357
R-squared	0.638	0.676	0.677	0.699	0.714
Panel B: Canada					
Total return	0.024	0.063***	0.072***	0.074**	0.087***
	(0.015)	(0.020)	(0.026)	(0.032)	(0.031)
Observations	21,783	21,411	20,875	20,325	19,763
R-squared	0.760	0.812	0.864	0.887	0.903
Yield spread	−0.069	−0.372	−1.230***	−1.440***	−1.950***
	(0.185)	(0.264)	(0.454)	(0.550)	(0.575)
Observations	21,848	21,476	20,940	20,390	19,828
R-squared	0.662	0.766	0.779	0.799	0.809

This table reports the ΔBMI coefficient estimates from regression (21) in the U.S. (Panel A) and Canadian (Panel B) samples of bonds 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 quintile, time to maturity, duration, and include bond and year–month fixed effects. Standard errors, clustered by issuer and year–month, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

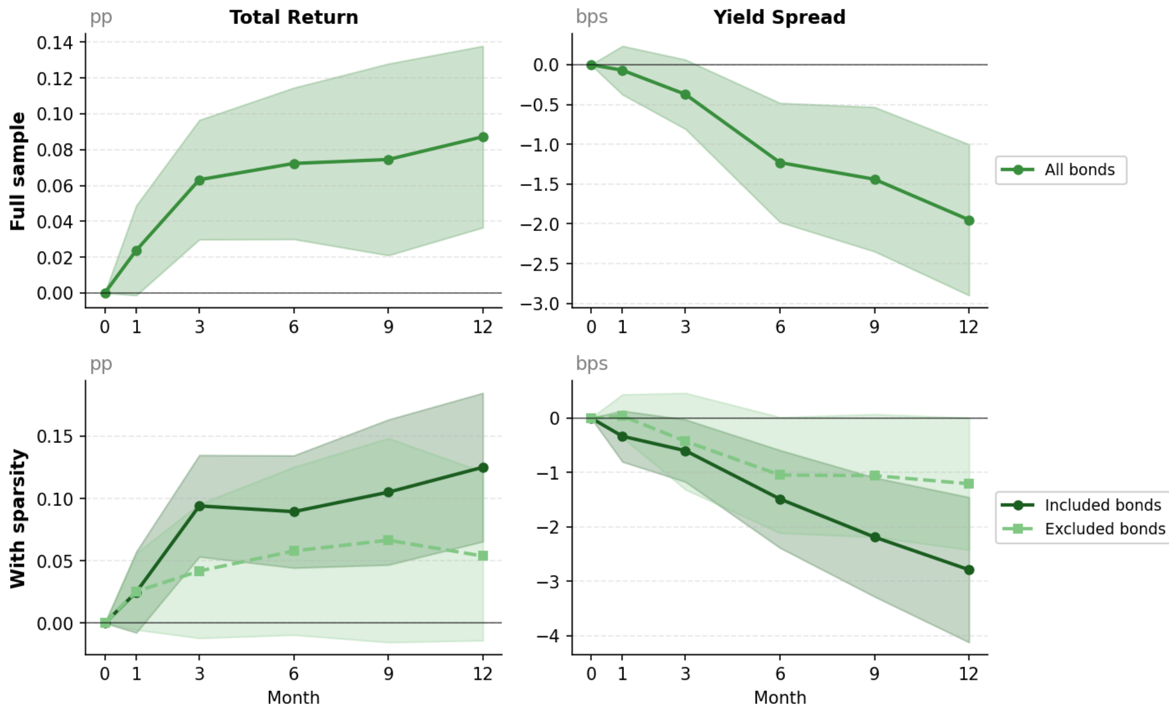
Importantly, these full-sample effects are attenuated because fund portfolios are sparse. In line with our model, the response of yield spreads to changes in BMI is significantly larger and more persistent in the set of included bonds, as documented in Table 5. The estimates are plotted in Figure 5 and demostrate that the spread response of included bonds is around two times stronger at all horizons.

The estimated effects are economically modest but clearly detectable, particularly for the included bonds. Upon crossing of the 5-year cutoff, BMI on average increases by 2.3 p.p. in the U.S. and 1.8 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 United States and 5 bps in Canada.

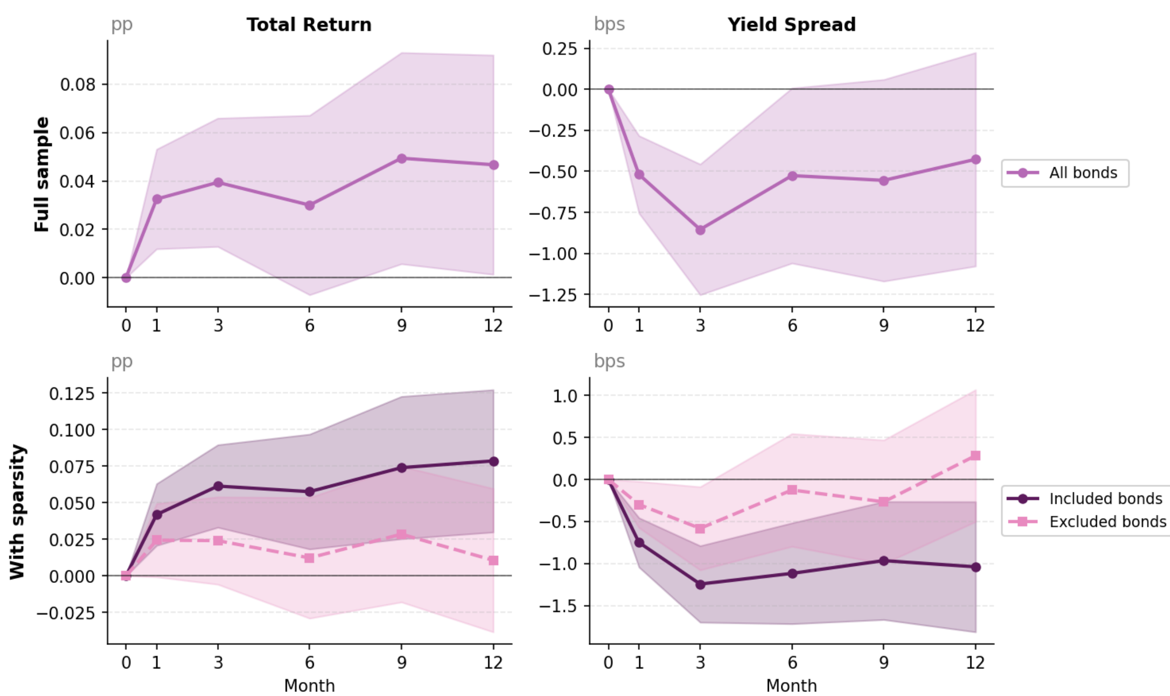
These estimates capture equilibrium price responses to changes in BMI and therefore incorporate general-equilibrium reallocation across substitute bonds. Existing price-impact studies (e.g., Chaudhary, Fu, and Li (2024)) typically define substitutability through similarities in cash-flow characteristics. In our framework, however, bonds with nearly identical fundamentals need not be close substitutes. For fund managers, bonds with higher index weights are less substitutable than bonds with lower weights or those outside the index. Dropping a high-weight bond from their portfolio increases tracking error the most, all else equal—a direct implication of our model. This adds another dimension to the discussion of substitutability. Consistent with the model's predictions, we find stronger price effects around the 5-year cutoff for included bonds. These bonds have higher benchmark weights within each bucket of substitutable bonds.

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 persistent effects. These results are reported in Appendix C.2.

Figure 5: The estimated effect of BMI on corporate bond prices



(a) Canada



(b) United States

This figure plots the estimates and 90% confidence intervals based on Tables 4 and 5 in (a) Canada and (b) the United States. 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 it belongs to and *included* otherwise.

Table 5: Changes in BMI and bond pricing (with sparsity adjustment)

	Change in the pricing variable within the sample of									
	included bonds					**excluded bonds**				
	$t+1$	$t+3$	$t+6$	$t+9$	$t+12$	$t+1$	$t+3$	$t+6$	$t+9$	$t+12$
Panel A: US										
Total return (	(0.013)	(0.017)	(0.024)	(0.029)	(0.029)	(0.015)	(0.018)	(0.025)	(0.028)	(0.030)
Observations	48,626	48,508	48,133	47,777	47,405	37,358	37,185	36,667	36,188	35,704
R-squared	0.731	0.748	0.810	0.860	0.891	0.750	0.763	0.837	0.882	0.910
OAS (bps)	−0.750***	−1.245***	−1.118***	−0.966**	−1.040**	−0.298*	−0.582*	−0.125	−0.266	0.284
	(0.178)	(0.276)	(0.364)	(0.426)	(0.470)	(0.167)	(0.300)	(0.407)	(0.446)	(0.476)
Observations	48,626	48,508	48,139	47,787	47,415	37,358	37,185	36,670	36,194	35,714
R-squared	0.656	0.686	0.690	0.708	0.720	0.641	0.695	0.701	0.730	0.751
Panel B: Canada										
Total return	0.025	0.094***	0.089***	0.105***	0.125***	0.025	0.042	0.058	0.067	0.054
	(0.020)	(0.025)	(0.027)	(0.035)	(0.036)	(0.019)	(0.033)	(0.041)	(0.050)	(0.041)
Observations	13,276	13,033	12,703	12,365	12,023	8,460	8,331	8,127	7,912	7,694
R-squared	0.764	0.827	0.886	0.907	0.924	0.772	0.812	0.847	0.874	0.886
Yield spread	−0.337	−0.602*	−1.490***	−2.190***	−2.790***	0.036	−0.429	−1.050	−1.060	−1.210
	(0.284)	(0.347)	(0.544)	(0.663)	(0.808)	(0.240)	(0.538)	(0.645)	(0.684)	(0.736)
Observations	13,307	13,064	12,734	12,395	12,053	8,495	8,366	8,162	7,948	7,730
R-squared	0.703	0.796	0.807	0.826	0.826	0.625	0.753	0.774	0.798	0.816

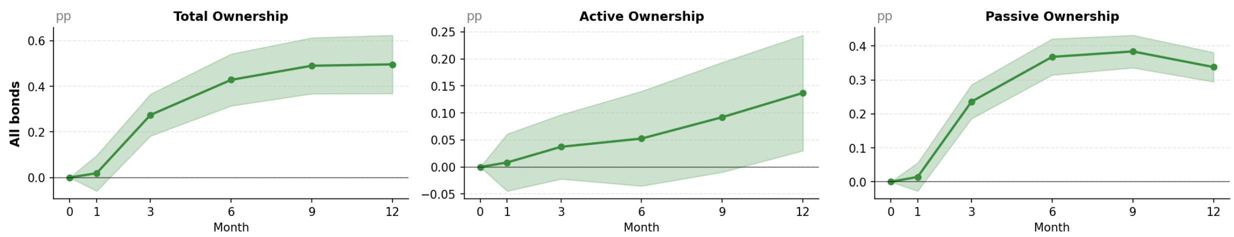
This table reports the ΔBMI coefficient estimates from regression (21) in the U.S. (Panel A) and Canadian (Panel B) samples of included and excluded bonds with time to maturity between 4 and 6 years. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes it belongs to and *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 quintile, time to maturity, duration, and include bond and year–month fixed effects. Standard errors, clustered by issuer and year–month, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

4.4 BMI and fund holdings

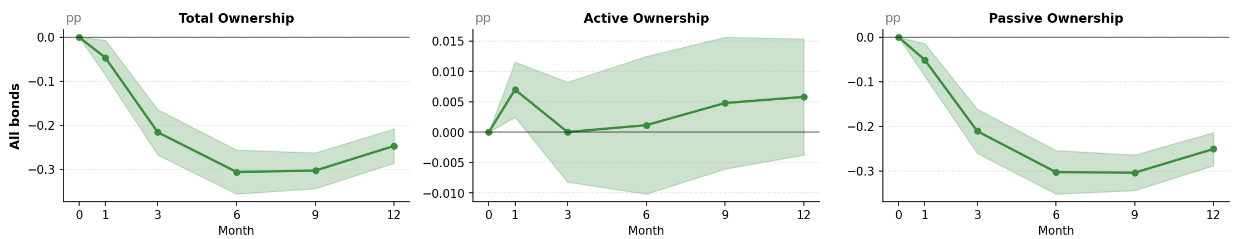
According to our model, equilibrium yields are lower for bonds with larger BMIs because funds hedge against underperforming their benchmark indexes. The model predicts that funds, both active and passive, rebalance in line with their mandates and, therefore, their assets contribute to BMI. In this section,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

First, our granular data allows 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 (21) 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

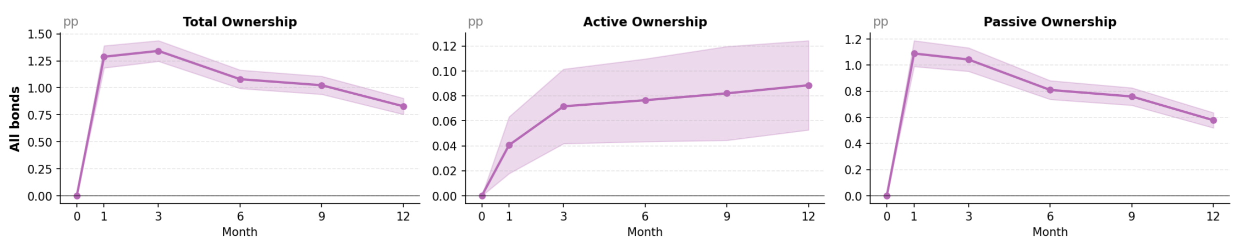
Figure 6: Changes in benchmarked ownership around the 5-year maturity cutoff



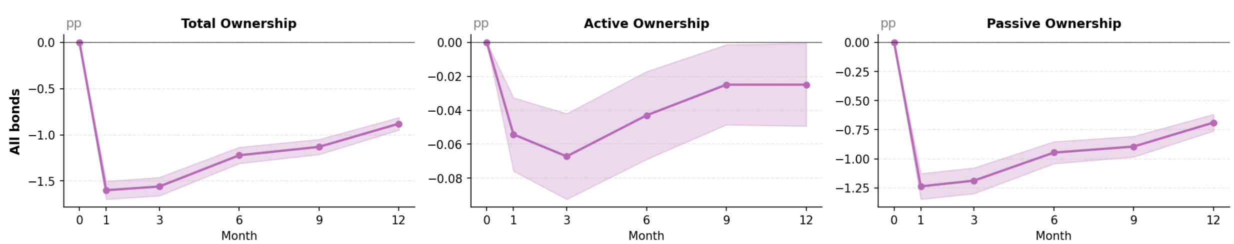
(a) Short-term fund ownership (Canada)



(b) Mid-term fund ownership (Canada)



(c) Short-term fund ownership (United States)



(d) Mid-term fund ownership (United States)

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 United States (c and d). We report total fund ownership changes of funds with short (1–5 years, panels a and c) and mid-term (5–10, panels b and d) years benchmark indexes, as well as their active and passive components. All estimates and confidence intervals are from estimating specification (21) 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% confidence intervals based on standard errors clustered by issuer and year-month.

Figure 6. The figure highlights that, when a bond crosses the 5-year cutoff, it is sold by funds benchmarked to the mid-term 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.[10]

Second, we study how aggregate fund ownership changes with BMI, both for all bonds and in the subsamples of included and excluded bonds. Table 6 presents the results of estimating the regression (21) for total, active and passive fund ownership. Consistent with the model, all funds' ownership significantly rises with an increase in BMI.

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 Table 7 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 7.

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 also get 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 level of sparsity, and our estimates reflect very little rebalancing in the set of bonds predicted to be excluded. In constrast, the magnitude of active ownership increase 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.,

[10] 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 6: Changes in BMI and bond ownership (all bonds)

	Change in the ownership variable over horizon				
	$t+1$	$t+3$	$t+6$	$t+9$	$t+12$
Panel A: US					
Total ownership	0.055***	0.168***	0.192***	0.244***	0.264***
	(0.020)	(0.030)	(0.034)	(0.033)	(0.036)
Observations	86,217	85,930	85,043	84,216	83,357
R-squared	0.131	0.187	0.287	0.379	0.464
Active ownership	0.003	0.070***	0.084***	0.132***	0.131***
	(0.014)	(0.024)	(0.028)	(0.028)	(0.030)
Observations	86,217	85,930	85,043	84,216	83,357
R-squared	0.118	0.183	0.287	0.379	0.464
Passive ownership	0.049***	0.097***	0.107***	0.110***	0.126***
	(0.010)	(0.013)	(0.015)	(0.015)	(0.014)
Observations	86,217	85,930	85,043	84,216	83,357
R-squared	0.171	0.229	0.341	0.449	0.544
Panel B: Canada					
Total ownership	0.089**	0.165**	0.222**	0.265**	0.289**
	(0.042)	(0.076)	(0.100)	(0.111)	(0.117)
Observations	21,848	21,476	20,940	20,390	19,828
R-squared	0.224	0.312	0.422	0.499	0.574
Active ownership	0.070*	0.124*	0.174*	0.209**	0.227**
	(0.038)	(0.070)	(0.093)	(0.101)	(0.110)
Observations	21,848	21,476	20,940	20,390	19,828
R-squared	0.204	0.299	0.413	0.491	0.567
Passive ownership	0.021***	0.045***	0.054***	0.061***	0.066***
	(0.008)	(0.013)	(0.016)	(0.018)	(0.016)
Observations	21,848	21,476	20,940	20,390	19,828
R-squared	0.254	0.292	0.382	0.465	0.534

This table reports the ΔBMI coefficient estimates from regression (21) in the U.S. (Panel A) and Canadian (Panel B) samples of bonds 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 quintile, time to maturity, duration, and include bond and year–month fixed effects. Standard errors, clustered by issuer and year–month, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

an increase in BMI upon crossing the 5-year cutoff (around 2.3 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

Table 7: Changes in BMI and bond ownership (with sparsity adjustment)

	Change in the ownership variable within the sample of									
	included bonds					**excluded bonds**				
	t+1	t+3	t+6	t+9	t+12	t+1	t+3	t+6	t+9	t+12
Panel A: US										
Total ownership	0.055**	0.197***	0.248***	0.319***	0.345***	0.045**	0.116***	0.101**	0.115***	0.124***
	(0.026)	(0.038)	(0.041)	(0.045)	(0.047)	(0.022)	(0.038)	(0.046)	(0.044)	(0.046)
Observations	48,626	48,508	48,139	47,787	47,415	37,358	37,185	36,670	36,194	35,714
R-squared	0.147	0.200	0.303	0.399	0.482	0.143	0.224	0.331	0.421	0.501
Active ownership	−0.003	0.086***	0.121***	0.190***	0.191***	0.005	0.035	0.016	0.026	0.018
	(0.020)	(0.033)	(0.037)	(0.040)	(0.042)	(0.018)	(0.033)	(0.041)	(0.041)	(0.043)
Observations	48,626	48,508	48,139	47,787	47,415	37,358	37,185	36,670	36,194	35,714
R-squared	0.132	0.195	0.300	0.396	0.481	0.132	0.221	0.334	0.423	0.502
Passive ownership	0.054***	0.111***	0.126***	0.128***	0.145***	0.037***	0.078***	0.084***	0.086***	0.103***
	(0.011)	(0.015)	(0.016)	(0.016)	(0.016)	(0.009)	(0.013)	(0.015)	(0.015)	(0.014)
Observations	48,626	48,508	48,139	47,787	47,415	37,358	37,185	36,670	36,194	35,714
R-squared	0.199	0.257	0.376	0.475	0.556	0.168	0.251	0.371	0.492	0.594
Panel B: Canada										
Total ownership	0.135**	0.242**	0.346**	0.362**	0.457***	0.024	0.037	0.060	0.142	0.136
	(0.056)	(0.107)	(0.144)	(0.149)	(0.152)	(0.048)	(0.091)	(0.118)	(0.141)	(0.135)
Observations	13,307	13,064	12,734	12,395	12,053	8,495	8,366	8,162	7,948	7,730
R-squared	0.268	0.360	0.463	0.534	0.602	0.198	0.283	0.413	0.504	0.591
Active ownership	0.109**	0.170*	0.253*	0.264*	0.357**	0.024	0.057	0.087	0.149	0.119
	(0.052)	(0.099)	(0.134)	(0.135)	(0.144)	(0.046)	(0.088)	(0.115)	(0.142)	(0.132)
Observations	13,307	13,064	12,734	12,395	12,053	8,495	8,366	8,162	7,948	7,730
R-squared	0.249	0.349	0.455	0.525	0.595	0.179	0.272	0.403	0.498	0.585
Passive ownership	0.031***	0.076***	0.095***	0.102***	0.103***	0.000	−0.014	−0.015	−0.001	0.021
	(0.011)	(0.017)	(0.022)	(0.025)	(0.024)	(0.005)	(0.012)	(0.018)	(0.019)	(0.021)
Observations	13,307	13,064	12,734	12,395	12,053	8,495	8,366	8,162	7,948	7,730
R-squared	0.291	0.322	0.414	0.487	0.543	0.239	0.312	0.411	0.510	0.597

This table reports the ΔBMI coefficient estimates from regression (21) in the U.S. (Panel A) and Canadian (Panel B) samples of included and excluded bonds with time to maturity between 4 and 6 years. 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 it belongs to and *included* otherwise. 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 quintile, time to maturity, duration, and include bond and year–month fixed effects. Standard errors, clustered by issuer and year–month, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

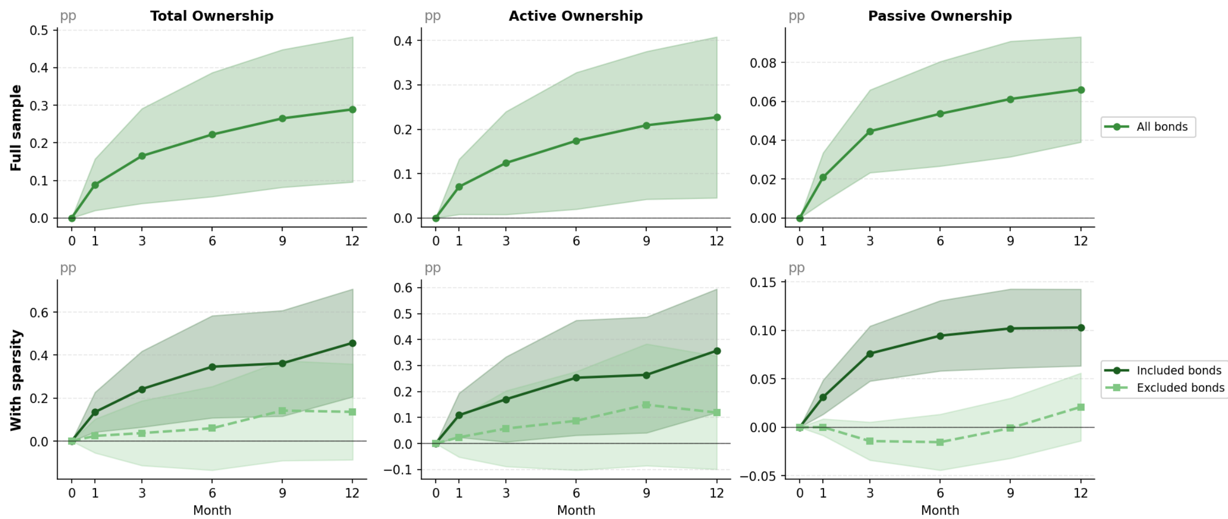
fund rebalancing, given their share of the market.

4.5 Discussion

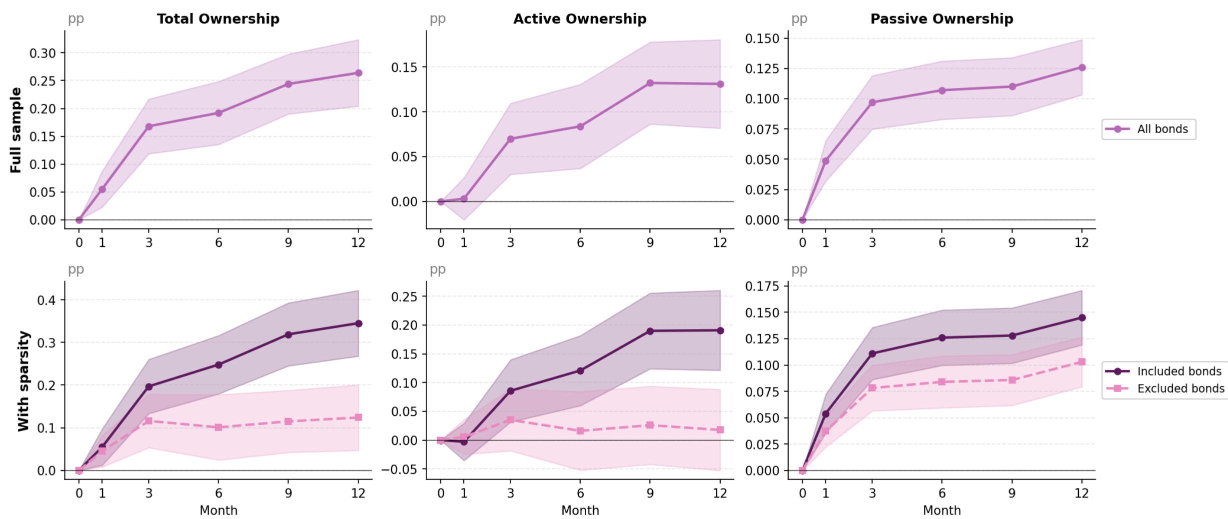
Our findings contribute to the growing literature on institutional demand in fixed-income markets, while offering several novel angles relative to prior work. In this section, we discuss these differences and also address several potential caveats in our research design.

Our approach builds on the maturity-based discontinuities in corporate bond markets

Figure 7: The estimated effect of BMI on corporate bond fund ownership



(a) Canada



(b) United States

This figure plots the estimates and 90% confidence intervals from Tables 6 and 7 in (a) Canada and (b) the United States. 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 it belongs to and *included* otherwise.

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.

Third, we show that the effects around maturity cutoffs are concentrated among bonds predicted to enter managers' sparse portfolios. The existing literature on institutional ownership in corporate bond markets 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 the magnitudes relative to the demand shift predicted by benchmarking, and to account for heterogeneity across maturities, fund types, and time. For comparison, estimates based on the cutoff switch dummy are reported in Appendix C.3 and are quantitatively similar to our baseline results.

We note that our research design allows us to detect the effects on prices and ownership because it leverages 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 the corresponding BMI—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 these maturity-driven shocks, even thouh the sparsity is measured imperfectly and subject to some noise.

One striking feature of the demand shocks in our setting is that the timing of a maturity crossing is perfectly predictable. Market participants can observe a bond's remaining maturity and anticipate when it will enter or exit a given maturity bucket. This stands in contrast with the index inclusion effects in equity markets, where index additions and deletions need to be predicted. However, the magnitude of the demand shock is perhaps even 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 that, we do not observe significant pretrends in bond prices before crossing the 5-year maturity cutoff, despite some anticipatory rebalancing of passive funds in the United States. We report those estimates in Appendix C.4.

Corporate bond markets are characterized by limited liquidity and infrequent trading, which raises concerns about stale prices and delayed price discovery. To mitigate these issues, we (i) use wide event windows to capture gradual adjustments, and (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, suggesting that our findings are not driven by liquidity effects. Consistent with industry practice, we rely on index provider prices from Bloomberg and FTSE, which are widely used for portfolio valuation by institutional investors and have been adopted in academic studies.

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. However, our included bonds are represented across all quintiles of the bid–ask spread distribution. To further address this interpretation, we exploit sharp ownership discontinuities around benchmark size-eligibility cutoffs in both Canada and the United States. In Appendix C.5, 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 (2025)). The presence of similar discontinuities for active funds provides additional support for our mechanism: the results are driven by benchmarking rather than by bond size mechanically proxying for liquidity or other bond characteristics.

5 Concluding Remarks

In this paper, we present evidence from Canada and the United States that corporate bond ownership by both active and passive corporate bond funds is strongly influenced by their mandates, and in particular their benchmark indexes. The tracking errors of active funds in our comprehensive sample of funds and their benchmarks are not substantially higher than those of passive, indicating that the two classes of funds still stay relatively 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, with an average fund in Canada holding only 15% of bonds in its benchmark and in the U.S. only 7%. We hypothesize that the reason for this sparsity is a high turnover of bond indexes and high rebalancing costs, as well as other portfolio management costs. We introduce portfolio management costs and benchmarking considerations in a simple equilibrium asset pricing model and show that it produces sparse portfolios that include assets with higher benchmark weights, higher size, and higher idiosyncratic risk within each bond category. The model also implies that bond prices depend on sparse benchmarking intensity, a measure of inelastic benchmark-driven demand of 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 only for bonds predicted to be included in sparse portfolios.

Whereas numerous studies in the literature documented index inclusion effects in equity markets, there are few corresponding studies for bonds. The lack of index effects in bonds is typically interpreted as evidence that bonds are more substitutable than equities. While we do not dispute this interpretation, we offer an alternative explanation for the difficulties in documenting index effects in bonds: funds' portfolio sparsity. According to our theory, index effects should be present only in bonds that are likely to be included in sparse portfolios, and our empirical findings confirm this implication. Corporate bond markets are not the only markets in which we observe portfolio sparsity. We expect to see similar effects of benchmarked capital in other markets with high rebalancing costs.

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 nuance to this concern. We show that mutual fund holdings are highly sparse and concentrated in bonds included in fund portfolios. These bonds—being more widely held across mutual funds—are precisely those most exposed to coordinated sales pressures during periods of outflows. In contrast, bonds excluded from sparse portfolios, which are largely held by investors less exposed to fund outflows, are less likely to be subject to such fire-sale dynamics. A natural next step in this agenda is to examine whether firms account for potential financial fragility when making their debt issuance decisions (Mota and Siani, 2025).

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 of the methods to estimate price elasticities of demand has been to use changes in *passive* bond ownership around index rebalancing events as demand shifters. Our results demonstrate that 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-driven demand and thus contribute significantly to these aggregate effects. One future avenue for research is to revisit demand elasticity estimates in light of these findings. Furthermore, a growing strand of research in corporate finance shows that passive ownership can lower firms' cost of bond financing and influence real economic outcomes (e.g., Dathan and Davydenko, 2025). Our findings indicate that active bond mutual funds behave similarly to passive funds in that they closely follow their mandates. As a result, it may be challenging to fully disentangle the effects of passive ownership from those attributable to mutual funds and ETFs more broadly.

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.

Brennan, Michael, 1993, Agency and Asset Pricing, Working paper UCLA.

Bretscher, Lorenzo, Lukas Schmid, Ishita Sen, and Varun Sharma, 2023, Institutional Demand and Corporate Bond Prices, Working Paper.

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, 2024, ETF sampling and index arbitrage, Working paper.

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.

Chaudhary, Manav, 2024, Regulator beliefs, .

——— , Zhiyu Fu, and Jian Li, 2024, Corporate Bond Multipliers: Substitutes Matter, Working Paper.

Dannhauser, Caitlin D., 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, 2025, Debt issuance in the era of passive investment, *Management Science*.

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.

Gabaix, Xavier, 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.

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, 2023, Remeasuring scale in active management, *SSRN Electronic Journal*.

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, forthcoming, *Review of Financial Studies*.

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, Jessica S., Maureen O'Hara, Andreas C. Rapp, and Xing Alex Zhou, 2025, Bond market illiquidity: Is portfolio trading the solution?, Discussion paper, Working paper.

Li, Ziang, 2023, The long-term interest rate and corporate bond credit spreads, *SSRN Electronic Journal*.

Ma, Linlin, Yuehua Tang, and Juan-Pedro Gomez, 2019, Portfolio manager compensation in the u.s. mutual fund industry, *The 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, MIT Sloan School of Management.

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.

Shim, J. J., and K. 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.

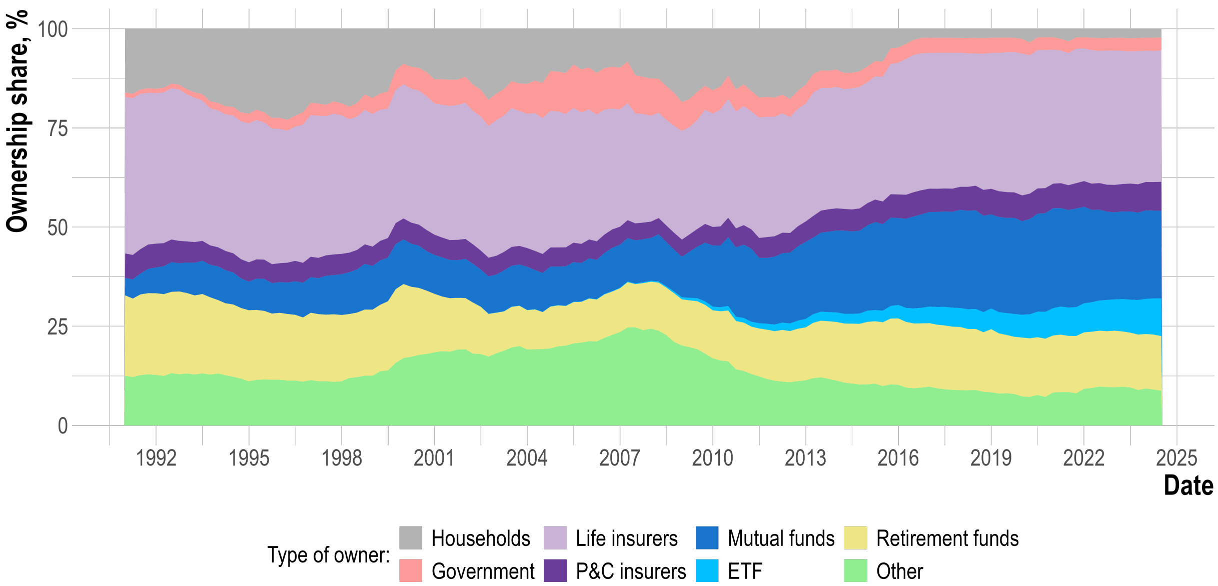
Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2010, Information Acquisition and Under-Diversification, *Review of Economic Studies* 77, 779–805.

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

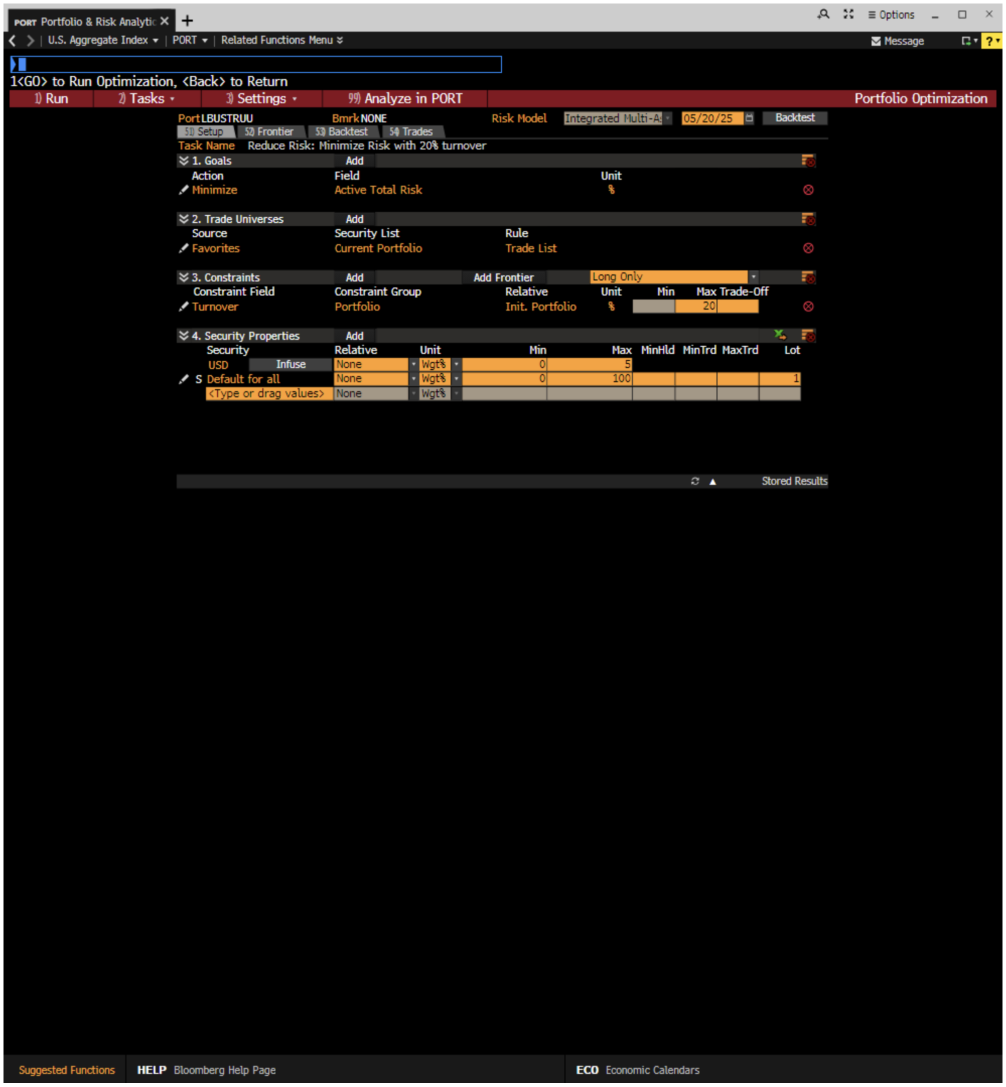
Figure A1: Ownership of U.S. Corporate Bonds by Domestic Investors



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 the screenshots of Bloomberg portfolio allocation tool. Subfigure (a) demonstrates the tradeoff between tracking error and turnover which the tool defaults to, and subfigure (b) highlights the saliency of Bloomberg industry classifications.

A.3 Fund samples summary statistics

Table A1: Descriptive statistics for our Canadian sample of funds

	Counts			Portfolio size (billion CAD)			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	514.0	542.9
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	382	52	36	2151.4	450.2	307.4	229.7	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	674.0	453.3	256.1	853.5	856.9
2019	449	73	55	3266.6	809.7	527.9	265.5	877.9	865.8
2020	465	77	61	3767.4	991.6	669.1	306.1	958.5	903.5
2021	459	79	63	3984.6	1088.8	722.5	306.9	972.3	931.7
2022	447	78	63	3237.6	956.2	626.8	301.1	934.9	892.1
2023	438	79	64	3519.5	1094.8	715.2	326.5	1032.2	987.1

TNA figures include matched separate accounts. 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 quantile	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 analysis for Canadian corporate bond monthly sample in 2004–2023. Yield spread and yield spread change are in basis point, and all other variables in panels A to D 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	577,932	128.168	74.918	-1015.200	48.560	114.080	224.420	424.310
Yield spread change	577,932	-0.457	22.254	-95.780	-19.180	-0.760	17.100	141.520
Total return	577,932	0.198	2.310	-9.520	-2.230	0.210	2.500	10.160
Panel B: BMI								
Total BMI	577,932	17.165	3.487	0.000	13.455	16.597	22.997	29.743
Active BMI	577,932	13.114	2.766	0.000	10.417	12.570	18.006	22.279
Passive BMI	577,932	4.051	1.522	0.000	2.250	3.743	6.482	7.484
Total BMI change	577,932	0.085	0.687	-1.124	-0.242	0.025	0.280	7.465
Active BMI change	577,932	0.048	0.592	-1.007	-0.225	-0.002	0.188	6.556
Passive BMI changes	577,932	0.038	0.130	-0.263	-0.031	0.028	0.095	1.434
Panel C: Ownership								
Total ownership	577,932	12.267	11.427	0.000	2.777	9.075	25.057	100.000
Active ownership	577,932	8.122	11.043	0.000	0.067	4.475	20.134	100.000
Passive ownership	577,932	4.152	2.368	0.000	1.316	3.835	7.343	43.934
Total ownership change	577,932	0.014	1.405	-10.049	-0.556	0.001	0.542	10.624
Active ownership change	577,932	-0.034	1.305	-8.948	-0.511	0.000	0.331	9.798
Passive ownership change	577,932	0.047	0.289	-3.507	-0.071	0.004	0.211	1.979
Panel D: Other bond characteristics								
log(Amount outstanding)	577,932	13.416	0.622	12.429	12.612	13.305	14.221	16.524
Rating (numeric)	577,932	7.403	1.929	1.000	5.000	8.000	10.000	10.000
Bid-ask spread quantile	577,932	2.481	1.125	1.000	1.000	2.000	4.000	4.000
Time to maturity (years)	577,932	10.858	8.957	1.510	2.460	7.120	26.580	30.000
Option Adjusted Duration	577,932	7.592	4.906	0.003	2.296	6.140	15.495	22.718

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

A.5 Fund tracking errors

Table A5: Tracking errors of bond funds

	Rolling 3-year TE			**Rolling 5-year TE**			
	Mean	**Median**	**St. dev.**	**Mean**	**Median**	**St. dev.**	**N**
Panel A: US							
Active Fund	2.070	1.377	2.087	2.064	1.445	1.845	57,393
Passive Fund	0.922	0.326	1.251	0.798	0.302	1.075	10,257
Panel B: Canada							
Active Fund	1.164	0.686	1.290	1.195	0.773	1.240	25,887
Passive Fund	0.590	0.175	0.930	0.452	0.153	0.678	4,221

This table reports tracking errors (TE) of funds in our US (panel A) and Canadian (panel B) samples, all in percent. TEs are computed from monthly fund return deviations 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.6 Aggregate sparsity

Table A6: 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	60	58
Short Corporate Bond Index	19	75	68
Short Overall Bond Index	68	45	68
Universe Corporate BBB Bond Index	5	92	313
Universe Corporate Bond Index	43	59	183
Universe Overall Bond Index	77	43	156
Weighted Average	73	45	141
Panel B: Passive funds			
Long Corporate Bond Index	56	44	36
Long Overall Bond Index	60	37	14
Mid Corporate Bond Index	74	28	18
Short Corporate Bond Index	71	30	92
Short Mid Corporate BBB Bond Index	58	44	-
Short Overall Bond Index	74	27	158
Universe Corporate Bond Index	75	28	24
Universe Overall Bond Index	78	42	89
Weighted Average	76	37	95

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 portfoliois 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 is based on the total AUM benchmarked to each index.

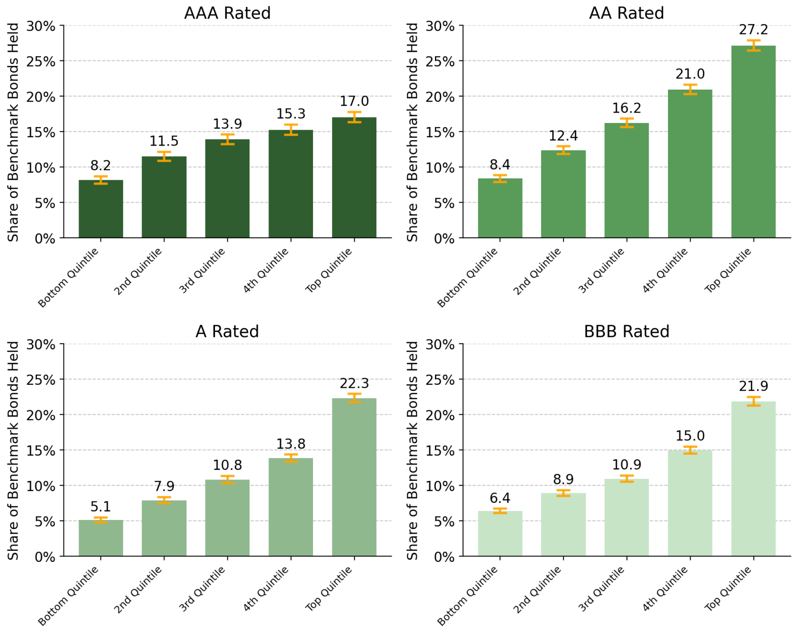
Table A7: Characteristics of the aggregate portfolios of U.S. funds with the same benchmark

	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	133	185
Aggregate 1-3 year	6	133	212
Aggregate 3-5 year	0	100	479
Aggregate	78	113	229
Corporate 1-10 year	2	97	119
Corporate 1-3 year	6	91	88
Corporate 1-5 year	7	92	94
Corporate 10+ year	24	77	207
Corporate	17	97	132
Credit 1-10 year	8	91	147
Credit 1-3 year	14	91	71
Credit 1-5 year	49	54	142
Credit 10+ year A+	39	52	-
Credit 10+ year	17	84	222
Credit 5-10 year	45	52	113
Credit	30	104	167
Govt/Credit 1-10 year A+	2	94	68
Govt/Credit 1-10 year	34	110	123
Govt/Credit 1-3 year Ex-BAA	21	69	59
Govt/Credit 1-3 year	60	117	142
Govt/Credit 1-5 year	22	92	200
Govt/Credit 10+ year	15	84	216
Govt/Credit	9	137	369
Weighted Average	66	108	210
Panel B: Passive funds			
Aggregate 1-3 year	3	96	184
Aggregate 1-5 year	13	83	132
Aggregate	93	64	108
Corporate 1-10 year	69	33	28
Corporate 1-3 year	83	24	7
Corporate 1-5 year	88	20	17
Corporate 10+ year	89	20	75
Corporate 5-10 year	90	22	43
Corporate	42	72	185
Credit 1-10 year	55	50	7
Credit 1-3 year	45	61	41
Credit 10+ year	88	19	66
Credit	41	58	59
Govt/Credit 1-10 year	51	44	5
Govt/Credit 1-3 year	24	76	112
Govt/Credit 1-5 year	82	42	51
Govt/Credit 10+ year	88	16	77
Govt/Credit 5-10 year	85	19	34
Govt/Credit	20	74	9
Weighted Average	90	56	91

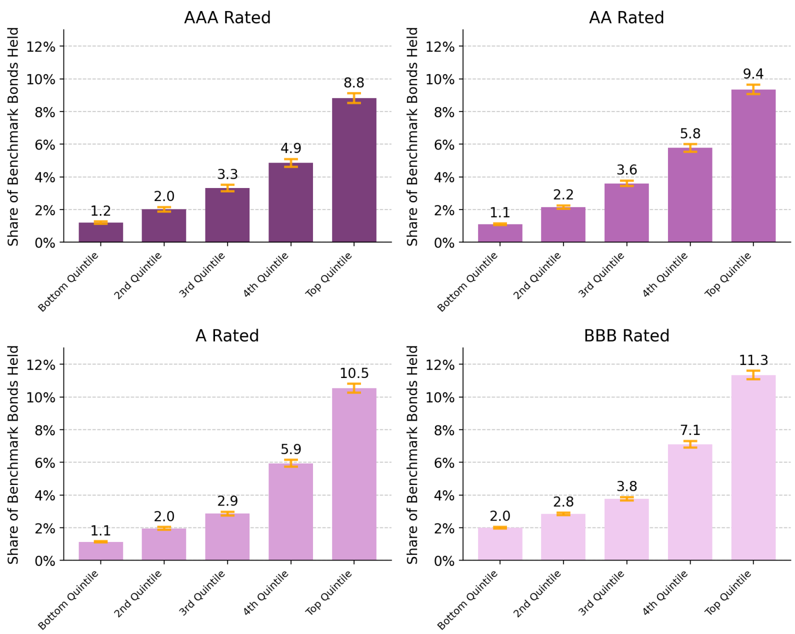
This tables shows characteristics of the aggregate portfolios of active mutual funds (Panel A) and passive mutual funds and ETFs (Panel B) with the same benchmark in the United States. Each portfoliois a value-weighted sum of funds benchmarked to the respective index. Tracking error is reported in basis points. Share of bonds held and Active Share are expressed in percent. The weighted average is based on the total AUM benchmarked to each index.

A.7 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

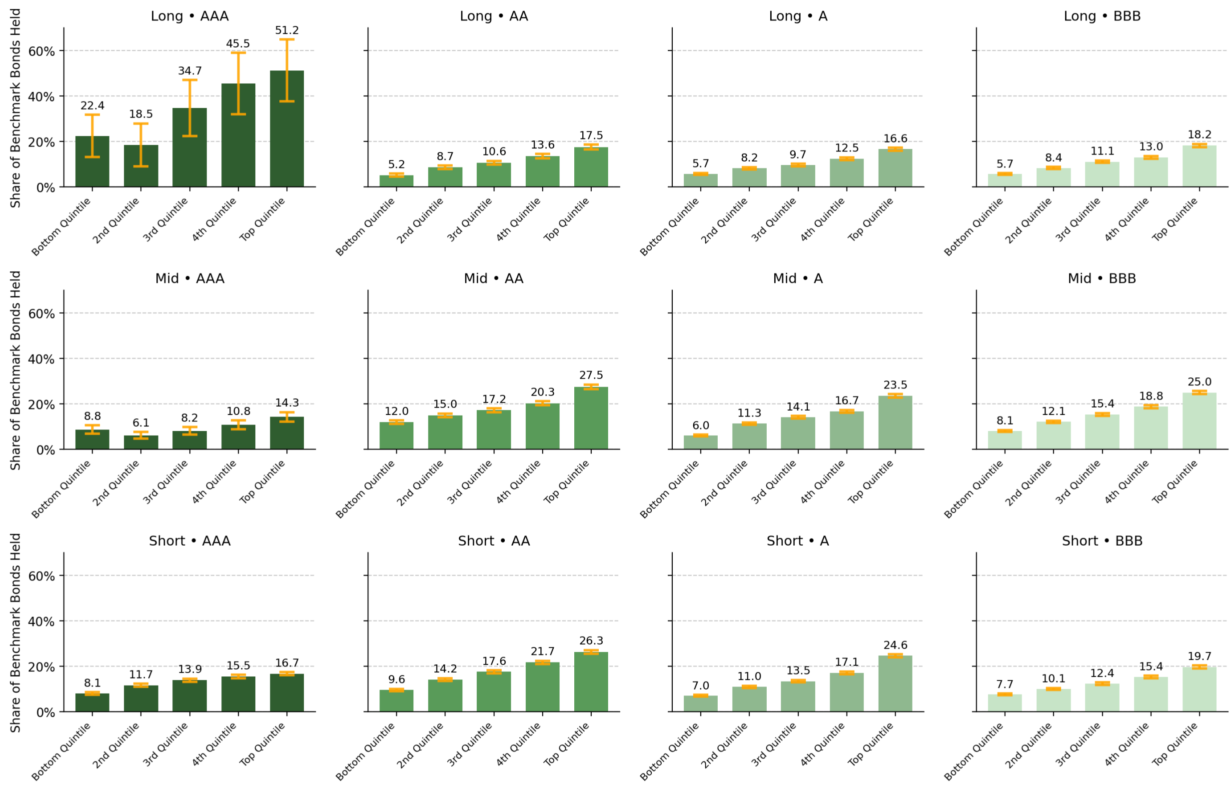


(b) US

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

A.8 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) US

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

A.9 Panel regressions of fund holdings on benchmark index membership and weights

Table A8: Regressions of active and passive fund holdings on benchmark index membership and weights for Canada and US

	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**		0.577***	0.435***
		(0.001)	(0.001)		(0.011)	(0.013)
Benchmark weight			0.961***			52.98***
			(0.054)			(3.212)
Observations	151,309	151,309	151,309	529,632	529,632	529,632
R^2	71.2%	71.2%	72.1%	42.7%	58.7%	59.4%
Panel A2: Canada – Active funds						
Dummy(within benchmark)		0.002***	0.0001		0.157***	0.085***
		(0.0003)	(0.005)		(0.008)	(0.005)
Benchmark weight			0.725***			33.38***
			(0.146)			(3.134)
Observations	285,342	285,342	285,342	3,052,538	3,052,538	3,052,538
R^2	75.3%	75.3%	75.4%	14.0%	15.5%	15.7%
Panel B1: US – Passive funds						
Dummy(within benchmark)		0.05***	-0.01		0.544***	0.487***
		(0.009)	(0.009)		(0.008)	(0.008)
Benchmark weight			0.781***			111.7***
			(0.032)			(9.211)
Observations	2,637,235	2,637,235	2,637,235	17,781,200	17,781,200	17,781,200
R^2	84.1%	84.1%	84.3%	20.0%	36.4%	36.6%
Panel B2: US – Active funds						
Dummy(within benchmark)		0.11***	0.04***		0.054***	0.024***
		(0.007)	(0.01)		(0.001)	(0.001)
Benchmark weight			0.656***			51.96***
			(0.091)			(1.544)
Observations	2,451,671	2,451,671	2,451,671	99,288,413	99,288,413	99,288,413
R^2	11.0%	11.0%	11.0%	6.8%	7.4%	7.6%

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 US. All regressions are saturated by including fund-by-date and bond-by-date fixed effects. Standard errors 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.10 Construction of historical benchmark data for US funds

We assemble a dataset of historical benchmarks of US mutual funds and ETFs from the database of historcial fund prospectus available on the website of the U.S. Securities and Exchange Commission (SEC)[11]. We further validate them using two addtional sources: (1) snapshot of benchmarks in Morningstar as of March 2025, and (2) SEC Mutual Fund Prospectus Risk/Return Summary data sets (MFRR)[12]. Benchmarks are mentioned in the annual return summary published in prospectuses.

A.10.1 Extract Benchmarks

For the crsp_cik_map table in WRDS (containing a total of 2586 company CIKs), all the 485APOS and 485BPOS files for 2504 company CIKs were downloaded using the sec_edgar_downloader Python package. By comparing several samples, it was confirmed that all required file types were successfully downloaded. Through a manual check, it was verified that the remaining company CIKs without downloads do not have data available on the EDGAR website.

From the 485APOS and 485BPOS files, extract 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 the table contains any of the following terms, retain 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*)

Use the large language model (Gemini API) to determine if the extracted table is an average annual return table. If it is, extract the corresponding benchmark for each fund and label the benchmark in sequence.

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

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

For tables where the closest preceding series name was not found in the previous step, identify 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 succeeds, use the Gemini API to analyze the text preceding each table and identify the string closest to the table that is most likely to be a fund name.

For each table, prioritize 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, extract the SEC mapping for the corresponding company CIK for year t-1, t, and t+1. If the table is from a year earlier than the earliest available year for the company CIK, retrieve the minimum available year and the next year's SEC mapping. For each extracted series name, identify the totally-matched series name and its corresponding series ID from the SEC-provided mapping.

A.10.2 Fund Categories

We classify US mutual funds and ETFs into active and passive following Pavlova and Sikorskaya (2023).

Active v.s Inactive

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 key words: ‘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 key words: ‘spdr’, ‘trackers’, ‘holdrs’, ‘powershares’, ‘street-tracks’, ‘etf’, ‘exchange traded’, ‘exchange-traded’, it is an ETF

Assets 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 doesn’t 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 the other categories are “Other”

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

A.10.3 Clean Benchmarks

We let Gemini API to create a mapping, manually check it, and then use 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.10.4 Correct Mismatched Funds and Benchmarks

If a SEC file contains 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 does not provide the seriesId for fund 2, or the name of fund 2 in the file does 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, I have adopted the following method to correct it: if a fund consistently uses benchmark A but suddenly shows benchmark B midway (like A A A A B A A A A), benchmark B is clearly a mismatched entry, and I replace B with A.

A.10.5 Validation

Validation Against MFRR We validate primary benchmarks extracted from SEC with MFRR time series benchmark data. Due to the incompleteness of MFRR data, 4.4 trillion are only in our sample, 0.3 trillion are only in MFRR, so the validation is for 2.1 trillion co-exist funds 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 only include the primary benchmarks for both SEC and MS benchmarks.

We deleted 205 seriesId (SEC fund id) which have more than one primary prospectus benchmarks in Morningstar. It's usually because the mapping file wrongly map different funds' secid (morningstar id) to the same seriesId.

Since MS benchmarks are the latest snapshot, for each seriesId (fund), I only compare the latest benchmark extracted from SEC with MS.

There are 1982 series (2.82 trillion) which only have benchmarks from SEC, and 248 series (0.28 trillion) which only have benchmarks from MS. We can only compare 1360 series (3.63 trillion) which have benchmarks both in SEC and MS.

86% (by the latest total net assets) have the same benchmark.

B Appendix: Proofs

B.1 Reduction to Mean–Variance Problems: Additional Detail

To evaluate the expectations in the direct investor's and manager's 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 $\alpha \in \mathbb{R}$,

$$E\Big[\exp^{\alpha x' Y}\Big] = e^{\alpha x' m + \frac{\alpha^2}{2} x' V x}. \tag{B1}$$

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 (B1) and dropping constants, one can see that the problem is equivalent to

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

which yields (4).

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 (B1) and dropping constants, one can see 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 (5).

B.2 Proof of Proposition 1

Proof of Proposition 1. The solution to the direct investor's problem (4) is well-known and yields the following demand function:

$$\theta_D = \frac{1}{\gamma}\Sigma^{-1}(\mu - p). \tag{B2}$$

Specializing this solution to the case of $\beta_{mi} = 0$, for which the matrix $\Sigma = \sigma_\epsilon^2 I_N$ is diagonal (with I_N denoting the $N \times N$ identity matrix), we arrive at (6).

The rest of the proof proceeds in steps.

Step 1 (Manager's optimal demand for a fixed set of assets). Fix any subset $\mathsf{I} \subset \{1, \ldots, N\}$ with the number of assets in I given by $|\mathsf{I}| = n$ and restrict the manager to positions $\theta_i = 0$ for all $i \notin \mathsf{I}$. The manager solves (5), which is a strictly concave quadratic program in θ_{I}, an n-dimensional vector of portfolio holdings. The first-order condition is

$$(a+b)(D_{\mathsf{I}} - p_{\mathsf{I}}) - \gamma(a+b)\Sigma_{\mathsf{I}}\Big((a+b)\theta_{\mathsf{I}} - b\omega_{\mathsf{I}}\Big) = 0,$$

hence

$$\theta_{\mathsf{I}} = \frac{1}{\gamma(a+b)}\Sigma_{\mathsf{I}}^{-1}(D_{\mathsf{I}} - p_{\mathsf{I}}) + \frac{b}{a+b}\omega_{\mathsf{I}}.$$

Under $\beta_{m,i} = 0$, $\Sigma = \sigma_\epsilon^2 I_N$, and thus, for $i \in \mathsf{I}$,

$$\theta_i = \frac{1}{\gamma\sigma_\epsilon^2(a+b)}(D_i - p_i) + \frac{b}{a+b}\omega_i, \qquad \theta_i = 0 \text{ for } i \notin \mathsf{I}. \tag{B3}$$

Step 2 (Value from holding a fixed set; marginal benefit). For the purposes of this proof, define $V(\mathsf{I})$ as the fund manager's maximized objective value in (5) when she is restricted to hold only assets in I. This value function is discussed in more detail in Section 3.2S. Using the first-order condition $(a+b)\theta_{\mathsf{I}}^* - b\omega_{\mathsf{I}} = \frac{1}{\gamma}\Sigma_{\mathsf{I}}^{-1}(D_{\mathsf{I}} - p_{\mathsf{I}})$ and substituting θ_{I}^* into (5), we obtain

$$V(\mathsf{I}) = \frac{1}{2\gamma}(D_{\mathsf{I}} - p_{\mathsf{I}})'\Sigma_{\mathsf{I}}^{-1}(D_{\mathsf{I}} - p_{\mathsf{I}}) + b\,\omega_{\mathsf{I}}'(D_{\mathsf{I}} - p_{\mathsf{I}}) - C|\mathsf{I}|. \tag{B4}$$

In the diagonal case $\Sigma = \sigma_\epsilon^2 I_N$, this becomes

$$V(\mathsf{I}) = \sum_{i\in\mathsf{I}}\left[\frac{1}{2\gamma\sigma_\epsilon^2}(D_i - p_i)^2 + b\omega_i(D_i - p_i)\right] - C|\mathsf{I}|. \tag{B5}$$

Under the assumption that an individual manager is atomistic and does not internalize price effects, adding asset $j \notin \mathsf{I}$ changes the value by

$$\Delta V_j \equiv V(\mathsf{I} \cup \{j\}) - V(\mathsf{I}) = \frac{1}{2\gamma\sigma_\epsilon^2}(D_j - p_j)^2 + b\omega_j(D_j - p_j) - C. \tag{B6}$$

Step 3 (Equilibrium prices for a conjectured cutoff n). Conjecture that the manager holds exactly the first n assets (ordered by decreasing benchmark weights), i.e., $\mathsf{I} = \{1, \ldots, n\}$. Then substitute the direct investors demand (6) into market clearing, $\lambda_D\theta_D + \lambda_F\theta = \overline{\theta}$. For $i \leq n$, use (B3) and solve for $(D_i - p_i)$:

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

so

$$D_i - p_i = \gamma\sigma_\epsilon^2\Big(\lambda_D + \tfrac{\lambda_F}{a+b}\Big)^{-1}\Big(\overline{\theta}_i - \tfrac{b}{a+b}\lambda_F\omega_i\Big),$$

which yields the first line of (8). For $i > n$, $\theta_i = 0$, and market clearing implies $\lambda_D\theta_{D,i} = \overline{\theta}_i$, hence $D_i - p_i = \gamma\sigma_\epsilon^2\,\theta_i/\lambda_D$, which gives the second line of (8).

Step 4 (Verification of the cutoff rule and determination of n^*). Define $V(n) \equiv V(\{1,\dots,n\})$. Using the equilibrium expressions for $D_i - p_i$ from Step 3, the incremental value of adding asset n simplifies to

$$V(n) - V(n-1) = \frac{\gamma}{2}\sigma_\epsilon^2\Big(\frac{\theta_n}{\lambda_D} + b\omega_n\Big)^2 - C. \tag{B7}$$

Since assets are ordered by decreasing benchmark weights, the right-hand side of (B7) is weakly decreasing in n. Therefore, the optimal portfolio size is characterized by a cutoff n^* satisfying

$$V(n^*) - V(n^*-1) \geq 0, \qquad V(n^*+1) - V(n^*) < 0,$$

which is equivalent to inequalities (9)–(10). This verifies the conjecture that the manager holds exactly the assets $\{1,\dots,n^*\}$.

Finally, substituting the equilibrium prices from Step 3 into (B3) delivers the manager demand (7), while the direct-investor demand is (6). Together with (8) and the cutoff characterization (9)–(10), this proves the proposition. □

B.3 Mixed equilibrium in Section 3.2

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 (7). For $i = n^*{+}1$, only a fraction φ of managers participates, while for $i > n^*{+}1$ no managers participate.

Market clearing is, 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 (6) and $\widetilde{\theta}_{M,i}$ is aggregate manager demand. Hence equilibrium prices coincide with (8) for all assets except $n^*{+}1$. In particular,

$$\mu_{n^*+1} - p_{n^*+1} = \gamma\sigma_\epsilon^2 \left(\lambda_D + \frac{\varphi\lambda_F}{a+b}\right)^{-1} \left(\overline{\theta}_{n^*+1} - \frac{b}{a+b}\varphi\lambda_F\omega_{n^*+1}\right). \tag{B8}$$

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-flows 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

$$\frac{1}{2\gamma\sigma_\epsilon^2}\Big(\mu_{n^*+1} - p_{n^*+1}\Big)^2 + b\,\omega_{n^*+1}\Big(\mu_{n^*+1} - p_{n^*+1}\Big) - C = 0. \tag{B9}$$

This condition is equivalent to

$$\frac{1}{2\gamma\sigma_\epsilon^2}\Big(\mu_{n^*+1} - p_{n^*+1}\Big)^2 + b\,\omega_{n^*+1}\Big(\mu_{n^*+1} - p_{n^*+1}\Big) - C = 0. \tag{B10}$$

Let $x \equiv \mu_{n^*+1} - p_{n^*+1}$. Equation (B10) implies

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

where we take the (economically relevant) root that yields $x > 0$. Define $y \equiv x/(\gamma\sigma_\epsilon^2)$. Equating this y to the market-clearing expression in (B8) yields the mixing probability in closed form:

$$\varphi = \frac{(a+b)\Big(\overline{\theta}_{n^*+1} - \lambda_D y\Big)}{\lambda_F\Big(y + b\,\omega_{n^*+1}\Big)}. \tag{B11}$$

Verification. By construction, managers are indifferent between portfolios of sizes n^* and n^*+1, so the proposed mixing is optimal. Assets $i \le n^*$ are held by all managers and are priced as in (8); assets $i > n^*+1$ are held by no managers and are therefore also priced as in the second line of (8). Finally, the no-profitable-deviation condition to adding n^*+2 (and beyond) holds because those assets are priced without manager participation and the marginal value of adding an asset is weakly decreasing in rank; in particular, the cutoff condition (10) rules out profitable expansion beyond n^*+1.

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 $\theta_2 = 0$. Ignoring constants and the (fixed) portfolio-management cost, the manager's objective in (5) can be written as

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

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 (15).

If the manager can hold both assets, the unconstrained optimum is the standard quadratic-program solution,

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

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

$$\mu_1 - p_1 = \gamma\Big(\sigma_1^2\theta_{D,1} + \rho_{12}\sigma_1\sigma_2\theta_{D,2}\Big) = \gamma\Big(\sigma_1^2\theta_{D,1} + \rho_{12}\sigma_1\sigma_2\frac{\overline{\theta}_2}{\lambda_D}\Big), \tag{B12}$$

$$\mu_2 - p_2 = \gamma\Big(\rho_{12}\sigma_1\sigma_2\theta_{D,1} + \sigma_2^2\theta_{D,2}\Big) = \gamma\Big(\rho_{12}\sigma_1\sigma_2\theta_{D,1} + \sigma_2^2\frac{\overline{\theta}_2}{\lambda_D}\Big). \tag{B13}$$

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 (B12) 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 is as defined in the lemma. Plugging this into (B12) yields the expression in (17) .

Next, substitute $\theta_{D,1}$ into (B13). 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 (18).

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 (17), 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. $\square$

Proof of Lemma 3.

Recall that for any admissible set $\mathsf{I} \subset \{1, 2\}$,

$$V(\mathsf{I}) = \max_{\theta:\, \theta_i = 0\ \forall i \notin \mathsf{I}} \Big\{ (a+b)\theta'(\mu - p) - \frac{\gamma}{2}\Big((a+b)\theta - b\omega\Big)'\Sigma\Big((a+b)\theta - b\omega\Big) - C|\mathsf{I}| \Big\}.$$

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

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

Substituting into the objective yields

$$V(\{1, 2\}) = \frac{1}{2\gamma}(\mu - p)'\Sigma^{-1}(\mu - p) - 2C.$$

Step 2 (Value function when $\mathsf{I} = \{1\}$). When the manager can trade only Asset 1, the optimal portfolio is given in Lemma 1. Substituting this portfolio into the objective function and using the equilibrium prices from Lemma 2, straightforward algebra yields

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

Step 3 (Difference). Under the assumption that an individual manager is atomistic and does not move asset prices when she decides to switch from trading only Asset 1 to trading both assets, we can simply subtract the expressions obtained in the previous two steps. This yields

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

which proves the lemma. $\square$

C Appendix: Empirical Analysis

C.1 Changes in BMI around maturity cutoffs

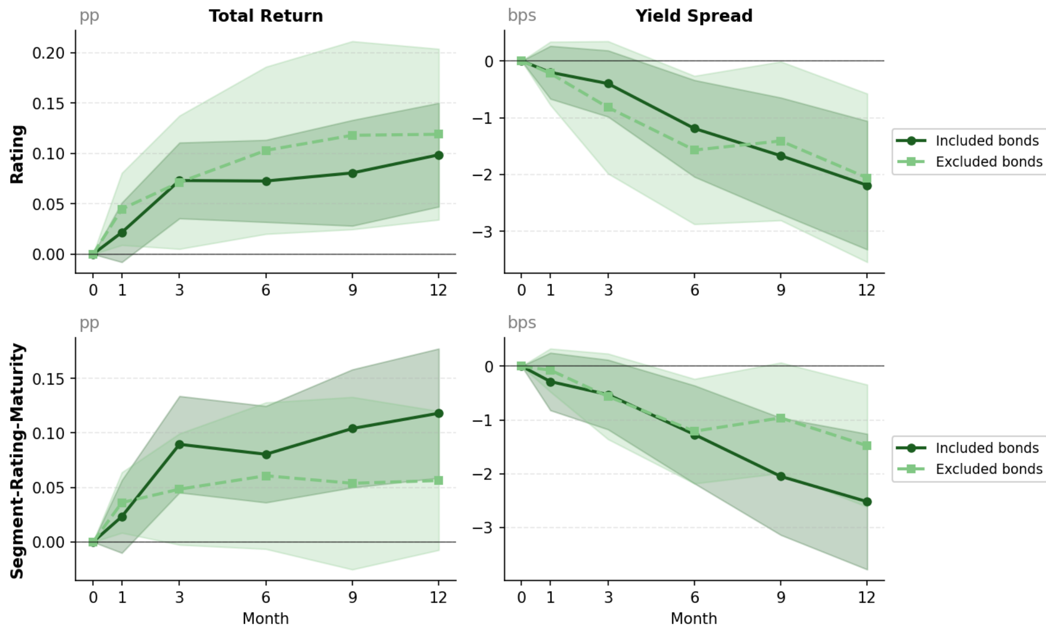
Table B1: Changes in BMI around maturity cutoffs

	All bonds	**Included**	**Excluded**
Panel A: United States			
Total BMI	2.273***	2.308***	2.229***
	(0.059)	(0.060)	(0.059)
Observations	86,185	48,624	37,329
R-squared	0.921	0.925	0.920
Panel B: Canada			
Total BMI	1.860***	1.848***	1.881***
	(0.038)	(0.041)	(0.042)
Observations	21,848	13,283	8,512
R-squared	0.987	0.987	0.987

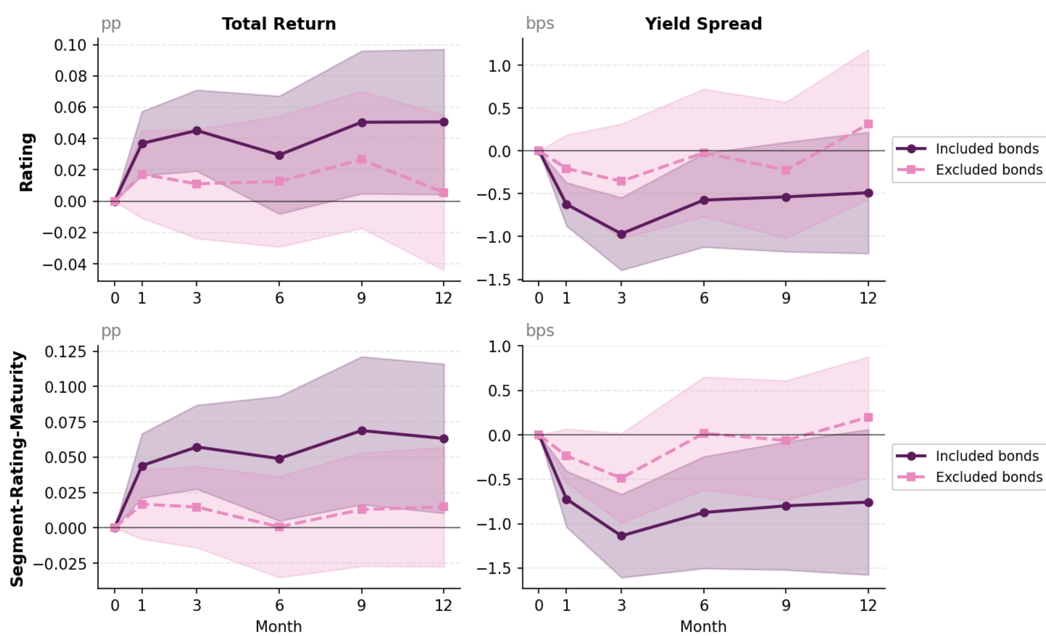
This table reports the estimates of ΔBMI regressed on the 5-year switch dummy in the U.S. (Panel A) and Canadian (Panel B) samples of bonds with time to maturity between 4 and 6 years. We consider full sample as well as the samples of included and excluded bonds. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes it belongs to and *included* otherwise. All regressions include baseline controls from specification (21), including the logarithm of the bond's par value, numeric credit rating, bid–ask spread quintile, time to maturity, age, duration, as well as bond and year–month fixed effects. Standard errors, clustered by issuer and year–month, are reported in parentheses. ${}^{*}p < 0.10$, ${}^{**}p < 0.05$, ${}^{***}p < 0.01$.

C.2 BMIs and bond prices with alternative sparsity adjustments

Figure B1: The estimated effect of BMI on bond prices with alternative substitute buckets



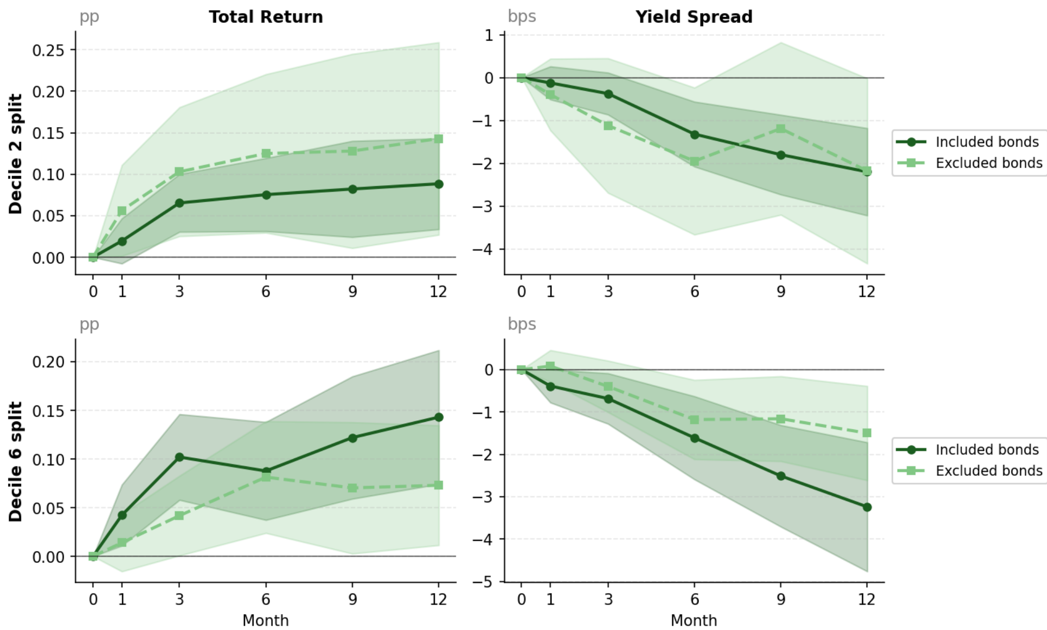
(a) Canada



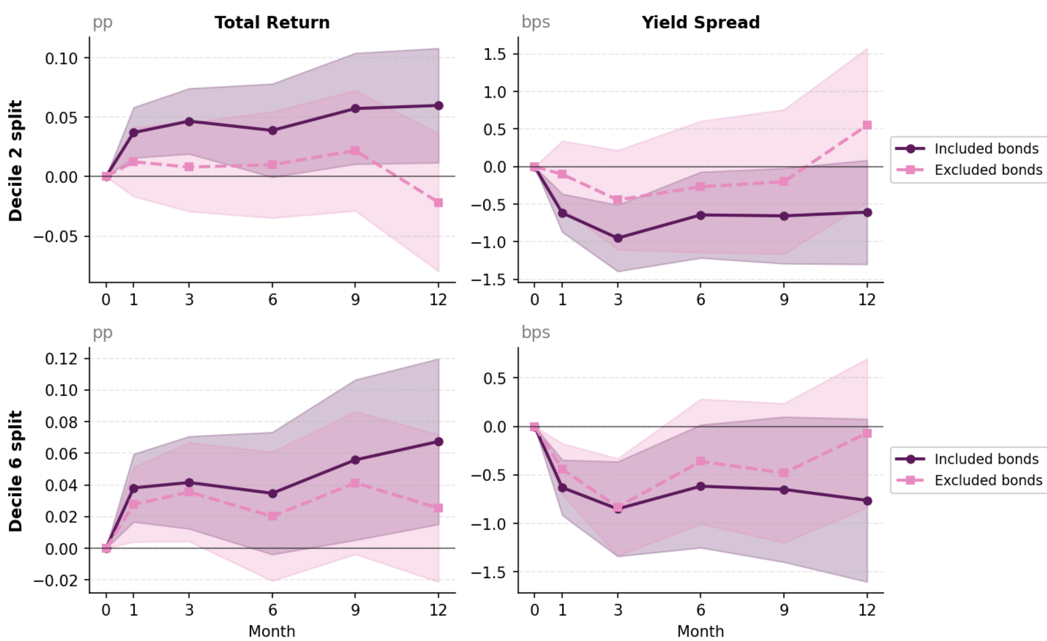
(b) United States

This figure plots the estimates and 90% confidence intervals based on regression (21) in (a) Canada and (b) the United States. In constrast with Figure 5, we use rating and segment-maturity-rating buckets to classify bonds into excluded and included ones, while keeping the size decile threshold unchanged.

Figure B2: The estimated effect of BMI on bond prices with alternative decile thresholds



(a) Canada

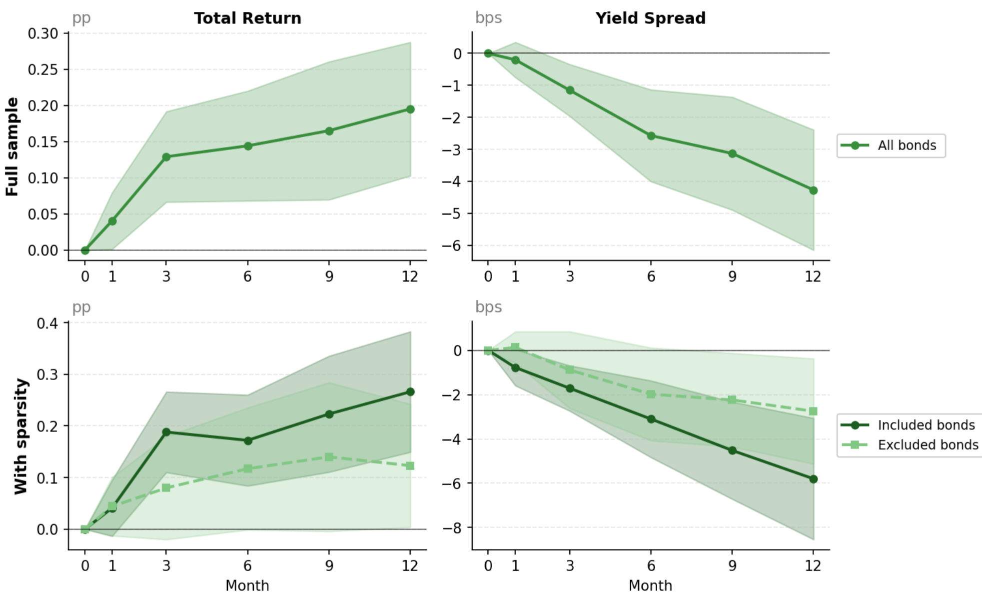


(b) United States

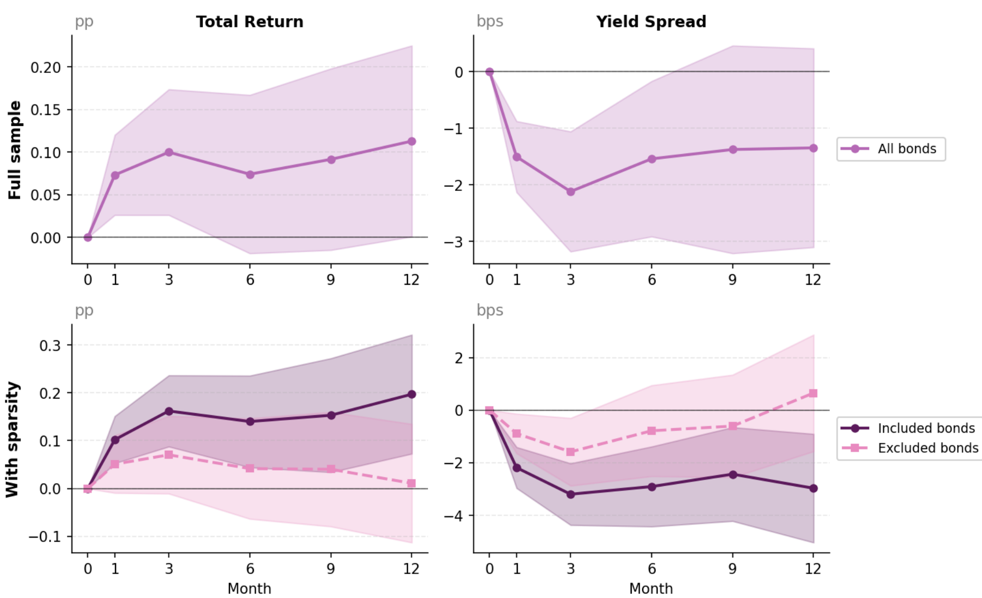
This figure plots the estimates and 90% confidence intervals based on regression (21) in (a) Canada and (b) the United States. In constrast with Figure 5, we use decile 1 and median of bond size disctribution within sector-maturity-rating bucket to classify bonds into excluded and included ones.

C.3 Estimation results for a switch dummy variable

Figure B3: The estimated effect of crossing the 5-year cutoff on corporate bond prices



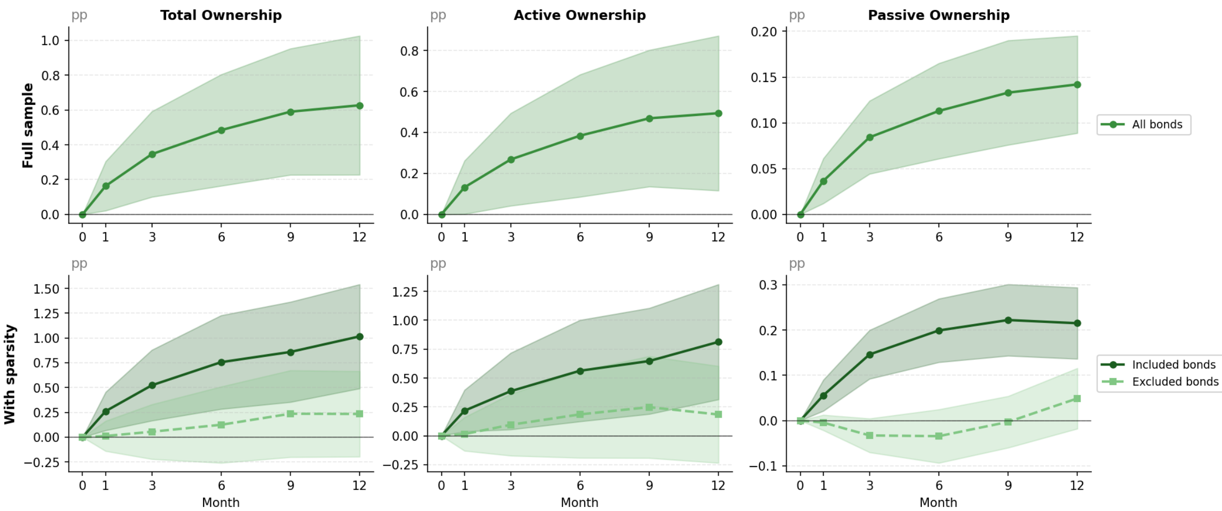
(a) Canada



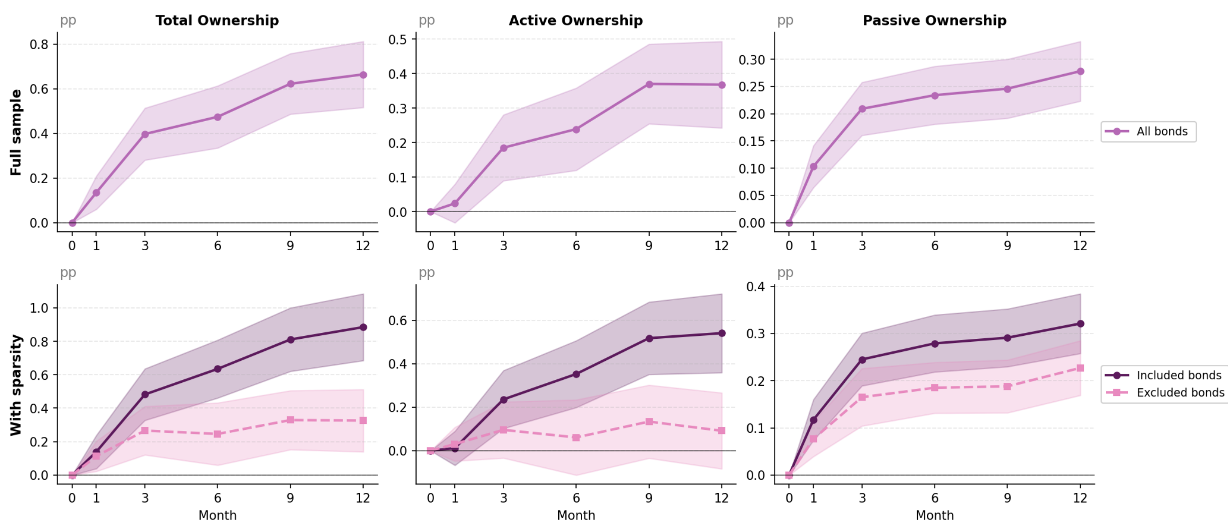
(b) United States

This figure plots the estimates and 90% confidence intervals from regression (21) using a switch dummy variable instead of BMI change in (a) Canada and (b) the United States. The switch dummy variable is 1 in the month when 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 it belongs to and *included* otherwise.

Figure B4: The estimated effect of crossing the 5-year cutoff on corporate bond fund ownership



(a) Canada



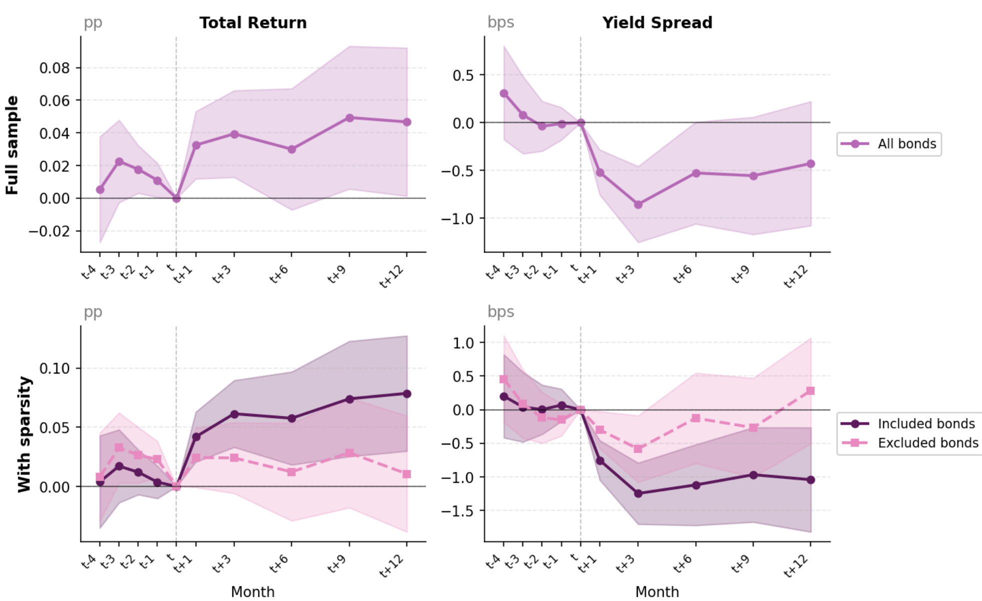
(b) United States

This figure plots the estimates and 90% confidence intervals from regression (21) using a switch dummy variable instead of BMI change in (a) Canada and (b) the United States. The switch dummy variable is 1 in the month when 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 it belongs to and *included* otherwise.

C.4 Bond prices and ownership before crossing the 5-year cutoff

Figure B5: The estimated effect of BMI on corporate bond prices (with lags)

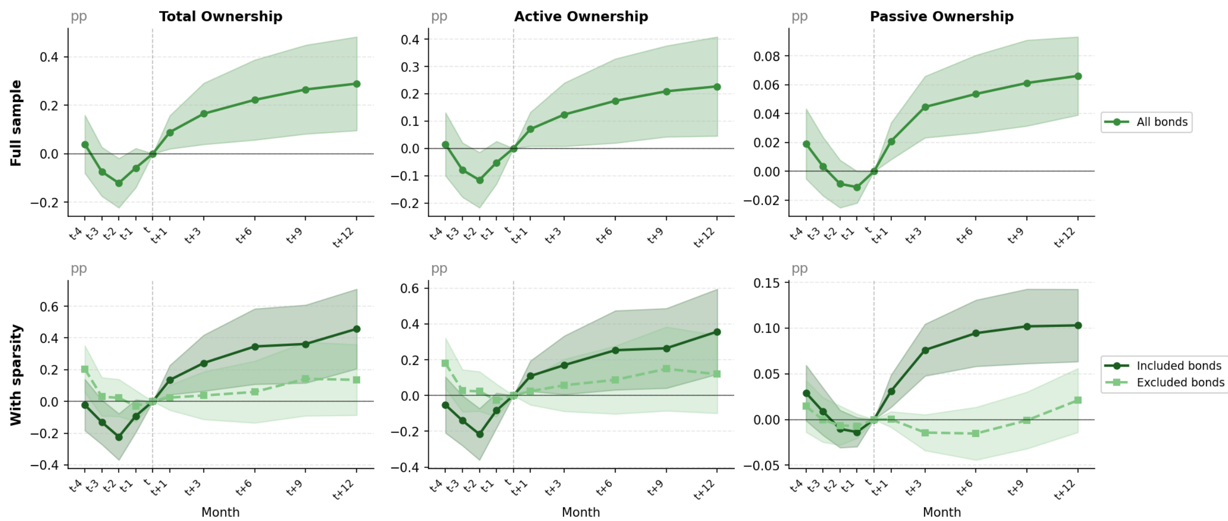
(a) Canada



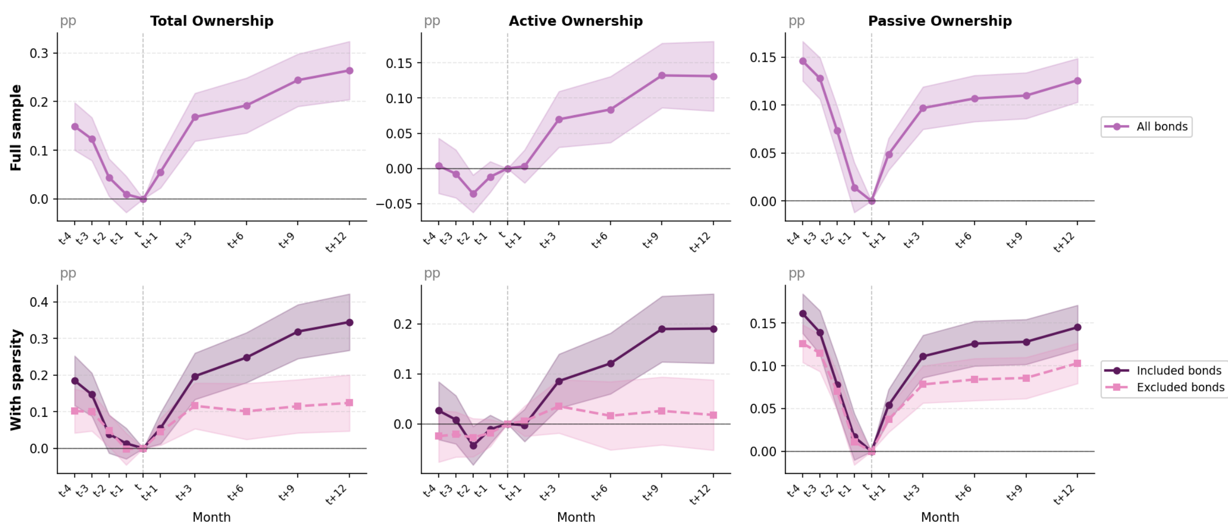
(b) United States

This figure plots the estimates and 90% confidence intervals from regression (21) with additional lags of the dependent variables in (a) Canada and (b) the United States. 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 it belongs to and *included* otherwise.

Figure B6: The estimated effect of BMI on corporate bond fund ownership (with lags)



(a) Canada

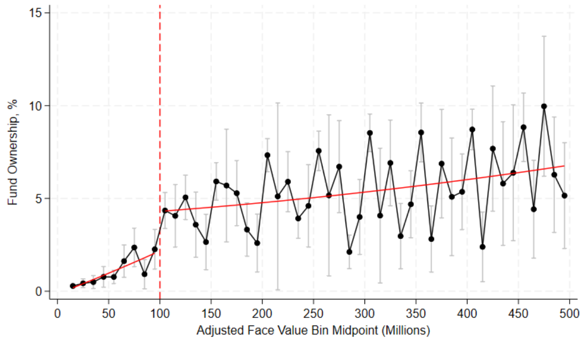


(b) United States

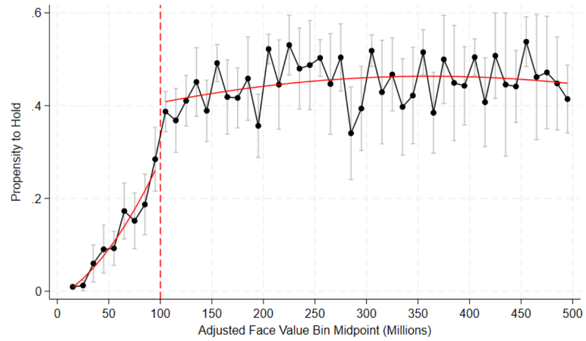
This figure plots the estimates and 90% confidence intervals from regression (21) with additional lags of the dependent variables in (a) Canada and (b) the United States. Control variables are lagged to match the lags of the dependent variables. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes it belongs to and *included* otherwise.

C.5 Benchmark Eligibility Size Cutoff and Fund Ownership

Figure B7: Ownership around the FTSE size cutoff in Canada



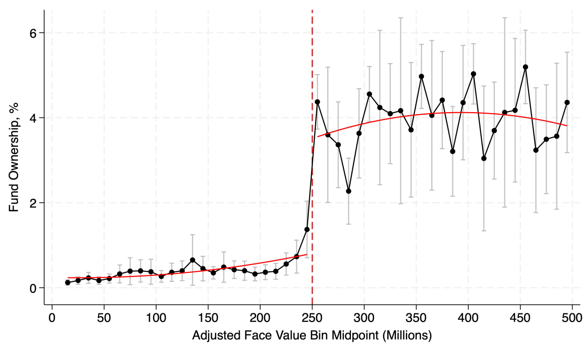
(a) Fund ownership



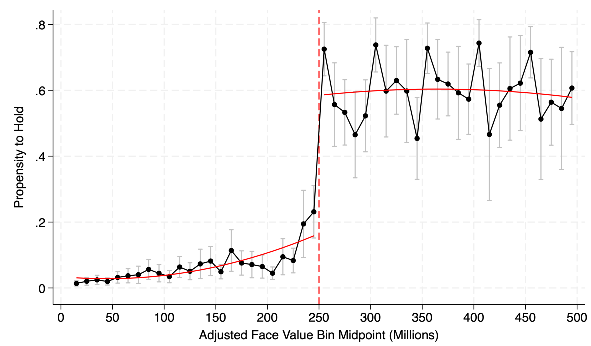
(b) Propensity to hold

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

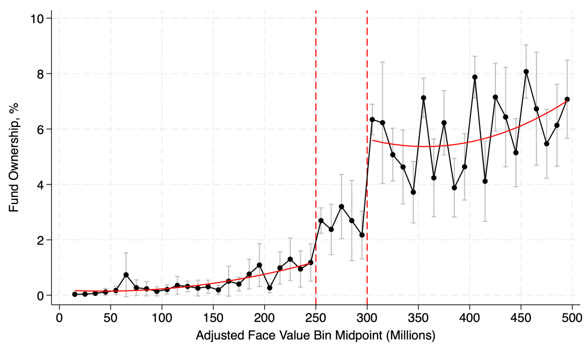
Figure B8: Ownership around the Bloomberg size cutoff in United States



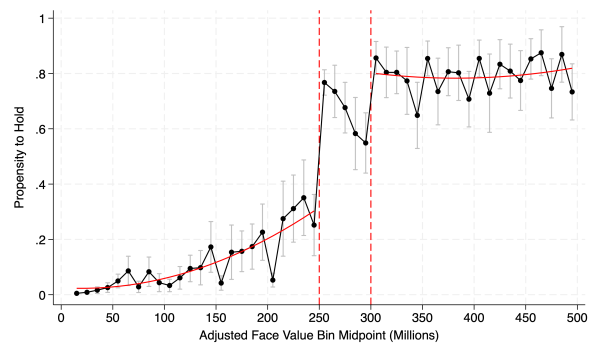
(a) Fund ownership before Q2-2017



(b) Propensity to hold before Q2-2017



(c) Fund ownership after Q2-2017



(d) Propensity to hold after Q2-2017

This figure illustrates fund ownership and propensity to hold corporate bonds around 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 sample includes a union of bonds in Bloomberg Aggregate bond index with bonds in WRDS sample.