Why Do Index Funds Have Market Power? Quantifying Frictions in the Index Fund Market*

Zach Y. Brown,† Mark Egan,‡ Jihye Jeon,§ Chuqing Jin,¶ and Alex A. Wu||

December 2025

Abstract

The number of index funds increased drastically from 2000 to 2020, driven in part by the emergence of exchange-traded funds. Yet, despite the proliferation of near-identical products, fee dispersion remains large and persistent, implying substantial market power. Motivated by evidence of strong inertia and other demand-side frictions, we develop a dynamic model of demand and supply for index funds that incorporates inertia, information frictions, and heterogeneous preferences. Using a new identification strategy that leverages fund-level new-sales data, we estimate that only 3–5% of households update their index holdings each month. However, counterfactual simulations show that while eliminating inertia alone reduces average fees by only 20%, eliminating both information frictions and inertia cuts them by 80%. This highlights that high information frictions—and their interaction with inertia—are the primary barriers to reallocation. Finally, we show that while ETF entry reduced expense ratios through both lower costs and additional product-market competition, demand-side frictions substantially dampened this competitive effect.

*We thank Malcolm Baker, Rob Clark, Dean Corbae, Ying Fan, JF Houde, Gaston Illanes, Alex MacKay, Sarit Weisburd, and Hanbin Yang and the seminar participants at American University, the Boston Conference on Markets and Competition, Boston University, Carnegie Mellon University, the International Industrial Organization Conference, London School of Economics, the Montreal Summer Conference on Industrial Organization, Northwestern, Ohio State University, Rice University, Singapore Management University, Stanford Institute for Theoretical Economics, Stockholm School of Economics, Toulouse School of Economics, UCI, UCLA, UCR, the University of Chicago, the University of Florida, the University of Georgia, and the University of Wisconsin IO/Finance Reading Group. We gratefully acknowledge funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

†University of Michigan & NBER. Email: zachb@umich.edu.

‡Harvard Business School & NBER. Email: megan@hbs.edu.

§Boston University. Email: jjeon@bu.edu.

¶Toulouse School of Economics. Email: chuqing.jin@tse-fr.eu.

||Harvard Business School. Email: alwu@hbs.edu.

1 Introduction

Index funds are widely viewed as a low-cost and transparent alternative to active investment management and have grown rapidly in popularity. In principle, competition among funds that track the same benchmark should drive fees toward marginal cost. In practice, fee dispersion across index funds is large and persistent, and many investors continue to hold high-fee funds even when nearly identical low-fee alternatives are available. This pattern is especially striking in light of the rapid expansion of low-cost exchange-traded funds (ETFs) over the past two decades, which many expected would trigger a "price war".[1] This paper asks why the entry of new funds and the growth of ETFs have not eliminated high-fee index products.

We document three demand-side frictions that allow index funds to sustain market power. First, investors rarely rebalance, so assets remain in high-fee funds for long periods. Second, even when investors make active choices, search and information frictions (as in Hortaçsu and Syverson, 2004) make it difficult to identify and select the cheapest funds. Third, even well-informed investors view funds as somewhat differentiated, and do not always choose the lowest-cost fund. We develop and estimate a framework that separately identifies inertia, information frictions, and horizontal differentiation. We then quantify how each friction and their interactions sustain fee dispersion and market power.

Disentangling these frictions is crucial for understanding market outcomes and for designing policy. If information frictions make it difficult for investors to identify low-cost funds, this would motivate transparency rules—such as the SEC's recent proposal to address misleading or deceptive practices—and tools such as the FINRA Fund Analyzer to facilitate fund comparisons.[2] In contrast, if investors remain in expensive funds due to inertia, then policies such as investor nudges or changes in the tax treatment of capital gains may have a greater impact.[3] Interactions among frictions are also important. For example, inertia can be much less costly for less sophisticated investors facing severe choice frictions, because they fail to optimize even when attentive. More broadly, the effectiveness of regulating product entry, price discrimination, and broker

[1]See, for instance, "Index Mutual Funds Face Price War With ETFs", Wall Street Journal, May 12, 2000.

[2]See https://www.sec.gov/news/press-release/2022-91 and https://tools.finra.org/fund_analyzer/.

[3]Capital gains from selling assets that are held for less than a year are taxed at a higher rate, which generates incentives for investors to hold assets for longer periods.

commissions depends on the specific frictions that investors face.[4]

We start by documenting that, despite the growth of low-cost ETFs, price dispersion remains substantial. Using detailed CRSP data and data on 401(k) plans, we document three key facts. First, the growth trajectories of new funds suggest that inertia plays an important role: it often takes roughly five years for low-cost entrants to reach scale. This pattern indicates that while many investors eventually shift to cheaper alternatives, they do so slowly. Second, information frictions matter: price dispersion in 401(k) plans, where information is standardized and prices are transparent, is about one-fifth of that in the broader market. The fact that dispersion does not disappear entirely in the 401(k) setting suggests that investors perceive index funds as differentiated products. Third, within identical portfolios, institutional funds charge nearly one percentage point less than retail funds, potentially due to institutional investors facing lower frictions.

We model index fund demand as a discrete choice problem in which an investor chooses an individual branded index fund within a specific investment class (e.g., Lipper Class), conditional on the investor's initial decision on how much to invest in that class. In this way, we abstract from the investor's overall portfolio allocation problem. Each period, with some probability, investors are either active (re-optimizing fund choice) or inactive (maintaining prior holdings). When active, investors still face information frictions which diminish their ability to identify the optimal index fund. We model information frictions and investor inactivity separately, reflecting their distinct underlying causes and policy implications in this market.[5] Investors also have heterogeneous preferences over index funds, so the product space is horizontally differentiated. We allow preferences, the level of frictions, and the set of available products to vary across two investor types: retail and institutional investors. ETFs are available to both groups, while institutional index funds are only purchased by institutions.

We estimate the model using fund-level data from CRSP from 2000 to 2020. We face two key empirical challenges. The first is separately identifying inertia from persistent unobserved preferences, since both give rise to the same empirical pattern of persistent market shares. We address this challenge by exploiting data on new sales

[4]Regulations of price discrimination and broker commissions are already subject to Securities and Exchange Commission rule-making. For instance, some forms of price discrimination for mutual funds are already barred under SEC Rule 22d of the Investment Company Act of 1940.

[5]Some models suggest that inertia is an endogenous response to information frictions. If investors believe that they will not be able to make optimal choices, they may decide to abstain from choosing at all (Steiner et al., 2017). In Section 4.1, we provide evidence that inertia and information frictions are linked.

in estimating inertia, relying on the fact that these data reflect active decisions. On average, new sales account for roughly 3% of assets under management (AUM), suggesting that demand is quite sticky. Our estimates indicate that roughly 95% (97%) of retail (institutional) investors are inactive each month, implying that about 46% of retail investors update their portfolios at least once a year. As a robustness check, we use an alternative identification strategy based on the persistence of exogenous demand shocks and obtain similar inertia estimates. Overall, our estimates of inertia align closely with survey evidence on investor rebalancing frequency.

The second empirical challenge is to disentangle preferences from information frictions, as both tend to make demand less elastic. To isolate information frictions, we exploit variation in the information environment when investors choose 401(k) plans, where investors choose from a simplified and transparent menu of funds determined by the plan sponsor (i.e., the participant's employer). By law, 401(k) investors receive standardized disclosures on the menu of options, expenses, and performance, making this a setting with minimal information frictions (Kronlund et al., 2021).[6] Experimental evidence from Choi et al. (2010) suggests that investors are more price-sensitive when fee information is transparent and salient, as in the 401(k) setting, though investors may still view index funds as horizontally differentiated products. We jointly estimate demand for index funds within and outside the 401(k) context using a discrete choice framework (Berry, 1994). We find that demand is more than twice as elastic in the 401(k) setting (3.3 vs. 1.6). The difference in demand elasticities between these two settings allows us to separately identify the contributions of information frictions and preference heterogeneity: for retail investors, information frictions are roughly 75% more important than preference heterogeneity.

We specify the supply side to understand why competition fails to eliminate market power. We develop a tractable model that incorporates dynamic incentives by examining the steady state of a differentiated Nash–Bertrand expense-ratio–setting game. Consistent with the data, index fund managers can operate multiple funds and price discriminate across institutional and retail investors by separately offering funds that are only available to institutional investors. When setting expense ratios, managers account for investor inertia, information frictions, and preference heterogeneity.

We show that these frictions have distinct implications for optimal pricing. Inertia has two potentially offsetting effects on steady-state pricing (e.g., Beggs and Klem-

[6] To account for investor inertia, we restrict our attention to new 401(k) plans in which all participants make an active decision.

perer, 1992). On the one hand, inertia increases the value of acquiring new customers, as those who switch to a fund are likely to remain, creating an incentive to set lower expense ratios to attract investors. On the other hand, inertia makes demand less elastic, giving managers an incentive to raise expense ratios and harvest existing clients. We show that with efficient capital markets (e.g., expected returns equal required returns), these opposing forces exactly offset, leaving equilibrium prices unchanged. However, in practice, when the growth-adjusted discount factor is smaller than 1, inertia can lead to higher steady-state expense ratios. An increase in either information frictions or preference heterogeneity unambiguously leads to higher prices.

We estimate the marginal cost of running an index fund by inverting each index fund manager's dynamic first order conditions. Given the presence of demand frictions, the estimates imply substantial market power. The estimated average (median) marginal cost is 35 (23) basis points, which implies an average (median) markup of 53 (32) basis points.

We use our model estimates to first quantify how demand-side frictions sustain market power and find that information frictions are the dominant force. Eliminating these frictions by, for instance, mandating additional disclosure, would lower average expense ratios for retail investors by 52%. Although 95% of retail investors are inactive each month, inertia alone has a more modest effect–a 22% decline. The limited demand-side response, which may be surprising, can be attributed to severe information frictions: investors cannot effectively optimize even if they are given more frequent opportunities to re-optimize. When both inertia and information frictions are eliminated, the average retail expense ratio falls by 81%, and the standard deviation falls by 71%, underscoring the strong interaction between these two forces.

Institutional investors are slightly less active than retail investors but face less severe information frictions and are more price-elastic. Overall, these differences give fund managers incentives to price discriminate across investor types. We find that eliminating price discrimination would lower the average expense ratio paid by retail investors by 11 basis points. However, if investors did not suffer from information frictions and inertia, then price discrimination would have a negligible impact. In an extension, we also analyze the role of agency frictions in the index fund market and find a small effect.[7]

[7] Using the framework in Robles-Garcia (2019), we find evidence that financial advisers distort demand; however, the conflicts of interest we estimate are smaller than have been found in other settings (Christoffersen et al., 2013; Hastings et al., 2017; Egan, 2019; Bhattacharya et al., 2019; Robles-Garcia, 2019). This is intuitive given that the index fund market is more transparent than

Lastly, we analyze how the introduction of ETFs impacted market competition and interacted with inertia and information frictions. ETFs differ from mutual funds in two critical dimensions: they have lower marginal operating costs (Jiang et al., 2023) and are available to all investors, precluding institutional-retail price discrimination. Our counterfactual results indicate that the introduction of ETFs lowered the average retail index fund expenses by 39%; 59% of the effect comes from increased competition, and 41% comes from the cost advantage of ETF. However, the effect would have been much larger in the absence of information frictions and inertia since these frictions limit the uptake of low-cost ETFs, highlighting how consumer frictions can dampen the competitive gains from product innovation.

Related Literature

It is well documented that even when financial products are similar, substantial price dispersion persists and consumers often fail to choose the lowest-price option (see Campbell (2016) and Clark et al. (2021) for an overview).[8] One strand of the literature focuses on the role of search costs and other information frictions in explaining these facts. In their seminal paper, Hortaçsu and Syverson (2004) document price dispersion in relatively homogeneous S&P 500 index funds and show that modest search costs can rationalize the observed price dispersion. Roussanov et al. (2021) extend the search model in Hortaçsu and Syverson (2004) to the market for active funds to study misallocation in the industry.[9] Similarly, Janssen and Thiel (2024) study how search frictions and preferences contribute to persistent investment in active funds.

However, these models often conflate search costs with other factors such as individual preference heterogeneity and switching costs, which cannot be separately identified in their setting, as noted by Hortaçsu and Syverson (2004). While information frictions, including search costs, are one explanation for why individuals choose expensive financial products, several studies argue that they are unlikely to fully explain choice behavior in certain settings (Woodward and Hall, 2012; Grubb, 2015).

the other markets in which agency frictions have been studied. Furthermore, we find that removing conflicts of interest has a limited impact on the expense ratios. This is because, especially in recent times, the incentives financial advisers face are relatively low-powered. For example, no-load mutual funds without 12b-1 fees accounted for 89% of mutual fund sales in 2021. See https://www.ici.org/system/files/2022-03/per28-02_2.pdf.

[8] Grubb (2015) notes that the failure to choose the lowest price is often observed more generally when price is complicated and consumers have limited experience in the market.

[9] Honka et al. (2017) and Roussanov et al. (2021) show that search costs may also be affected by marketing.

A number of studies have shown that consumer inertia and switching costs play an important role in settings including retirement fund choice (Madrian and Shea, 2001; Illanes, 2016; Luco, 2019), portfolio choice (Gabaix et al., 2024), mortgage choice (Allen and Li, 2020; Andersen et al., 2020; Zhang, 2022), and banking choice (Kiser, 2002).[10]

Our empirical setting also suggests that both inertia and information frictions may be responsible for the slow adoption of new low-cost funds and the persistence of high-fee funds. Our paper, therefore, contributes to these bodies of literature by developing a tractable model that jointly incorporates these key frictions and by proposing a new identification approach. This allows us to estimate inertia and information frictions, each separately from preferences, and ultimately answer why market power persists even with active entry in this market.

Finally, our paper connects to other studies on the ETF and mutual fund markets, and more broadly to the growing literature at the intersection of industrial organization and finance. An et al. (2021) develop a structural model of the ETF market that incorporates two-tiered competition between index providers and ETF managers. The authors document that the index providers (e.g., S&P Dow Jones) that create and license indices for ETFs to track also have substantial market power in addition to the ETF managers.[11] Gavazza (2011) shows how demand for product varieties and demand spillovers affect the market structure and the level of fees for mutual funds. While we focus on index fund choice, rather than the more general problem of portfolio choice, our framework relates to the literature using a demand-system approach to asset pricing (Koijen and Yogo, 2019a) and to other financial contexts more generally.[12] We show how to extend these frameworks to quantify the role of various frictions in a dynamic environment characterized by investor inertia, which is an important feature of many household financial markets.

The remainder of the paper is structured as follows. In Section 2 we describe our data and present motivating evidence for the frictions incorporated in our model. In

[10]Inertia has been shown to be important for competition in many other settings including electricity (Hortaçsu et al., 2017), retail gasoline (MacKay and Remer, 2022), cloud computing (Jin et al., 2022) and health insurance markets (e.g. Handel, 2013; Ho et al., 2017).

[11]Baker et al. (2022) and Egan et al. (2022) use similar demand-side approaches to recover investor expectations in the index fund market.

[12]Other asset examples include Koijen and Yogo (2019b); Bretscher et al. (2020); Benetton and Compiani (2021) and Haddad et al. (2021). These IO-type demand systems (e.g., Berry (1994), Berry et al. (1995), etc.) have been used in other financial settings such as demand for banks (Dick, 2008; Egan et al., 2017; Xiao, 2020), mortgages (Allen et al., 2014; Benetton, 2021; Robles-Garcia, 2019) and insurance (Koijen and Yogo, 2016, 2022).

Section 3 we develop our structural model, and we present the corresponding estimates in Section 4. We present our counterfactual analysis in Section 5. Lastly, Section 6 concludes.

2 Data and Motivating Evidence

2.1 Data

Our base index fund data set comes from CRSP Mutual Fund Database for 2000 to 2020. We restrict our attention to funds classified in CRSP as index funds, including both mutual funds and ETFs.[13] Index funds in the data are defined at the share class level, which means that multiple share classes may represent the same underlying portfolio. While some of the mutual fund literature aggregates share classes to the fund level, we keep the unit of observation at the share-class level to capture how retail and institutional share classes contribute to the observed price dispersion. We observe monthly data on total net assets and returns and quarterly information on other fund characteristics such as expense ratios and Lipper classification.

Table 1 displays the summary statistics corresponding to our base data set. We have roughly 500,000 month-by-fund observations, which covers 5,266 index funds across 150 Lipper Classes. On average, we have roughly 8 years of monthly AUM data for each fund in our sample and 35 funds per Lipper Class. Consistent with the previous literature, we find a large degree of price dispersion. The average expense ratio is 77 basis points with a standard deviation of 65.[14]

An important dimension of heterogeneity in our analysis is investor type. For index funds classified as mutual funds, CRSP indicates whether the fund is an institutional or retail mutual fund. In contrast, ETFs may be purchased by either institutional or retail investors. We use quarterly institutional holdings data (13F) to determine the share of

[13] We restrict our data set to all funds defined in CRSP as index funds (i.e., $index\_fund\_flag$ is equal to "B", "D" or "E"). We focus on index funds given that these products are relatively homogeneous and form an important part of the market. Because CRSP only started reporting whether a fund is an index fund in 2003, we define a fund as an index fund if it is ever classified by CRSP as an index fund in the data. We find quantitatively similar results if we restrict our attention to those funds classified as "pure" index funds as per CRSP (i.e., $index\_fund\_flag$ is equal to "D").

[14] Following Hortaçsu and Syverson (2004), we construct load-adjusted expense ratios by adding one-third of all loads to the reported expense ratio, reflecting that investors update their portfolios roughly once every three years. This adjustment has only a minor effect because front and rear loads are rare, especially in recent years; for example, 90% of index funds in our sample had no loads in 2020. Our results are quantitatively similar if we omit this adjustment or exclude all load-charging funds.

Table 1: Summary Statistics

	Count	Mean	Std. Dev.	Median
Total Net Assets ($mm)	564,218	1,372.07	7,887.25	61.60
Expense Ratio (bp)	564,218	96.27	91.72	63.00
Exp Ratio (Unadj. for Loads; bp)	564,218	76.53	64.63	60.00
Annual Returns (%)	507,091	5.54	23.05	6.13
Retail Mutual Fund	564,218	0.35	0.48	0.00
Institutional Mutual Fund	564,218	0.26	0.44	0.00
ETF	564,218	0.38	0.49	0.00
ln(# of Funds in Same Mgmt. Company)	564,218	4.04	1.41	4.34
12b-1 Fees (bp)	564,218	13.74	28.94	0.00
Has Front Load	564,218	0.07	0.26	0.00
Has Rear Load	564,218	0.13	0.34	0.00
Std. of Daily Returns (pp, annualized)	559,562	18.56	13.79	15.06
Number of Index Funds	5,266			
Number of Lipper Classes	150			

Note: Table 1 displays summary statistics corresponding to our main sample, all CRSP mutual funds and ETFs flagged as index funds for 2000—2020. Observations are at the fund-by-month level. The variables *Retail Mutual Fund*, *Institutional Mutual Fund*, and *ETF* are all indicator variables. Expense ratios are load-adjusted as described in Section 2.1.

ETF assets held by institutional versus retail investors. Roughly 35% of the funds in our sample are retail mutual funds, 26% are institutional mutual funds, and the remaining 38% are ETFs.

We supplement the CRSP data with information on new sales for mutual funds from their N-SAR and N-PORT filings, compiled by Morningstar. Mutual funds report new sales in their N-SAR filings before 2018 and N-PORT filings after 2018. For each mutual fund, we observe the number of new shares purchased (excluding reinvestments of dividends and distributions) at the monthly level. The new sales reflect new capital entering the mutual fund market as well as assets moving across funds. We therefore leverage the new sales data as a key input to identifying investor inertia, as discussed in Section 4.1.1. The new-sales data are available for 38% of our fund-month observations for mutual funds, and 55% of market-month pairs have at least one new-sales observation.

Employer-sponsored investment accounts offer a simplified menu of funds due to disclosure requirements, providing a setting with minimal information frictions. We supplement our analysis with data on the menu and allocation of funds within 401(k) plans from 2009 to 2019 from BrightScope Beacon. The data cover 85 percent of employer-sponsored investment accounts subject to ERISA. Additional detail on the data can be found in Egan et al. (2021).[15]

2.2 Motivating Evidence

2.2.1 Price Dispersion

We start by documenting that there is substantial price dispersion in index funds even after the advent of ETFs. Because these products are relatively homogeneous, the presence of price dispersion provides initial evidence of market power. Figure 1 Panel (a) shows that the number of index mutual funds increased by almost threefold between 2002 and 2020, and ETFs, while almost non-existent initially, eventually outnumbered mutual funds. Panel (b) illustrates that the market size, measured by total AUM, also increased substantially over the period.

Figure 1: Growth of Index Funds

(a) Number of Funds

(b) Fund Assets Under Management

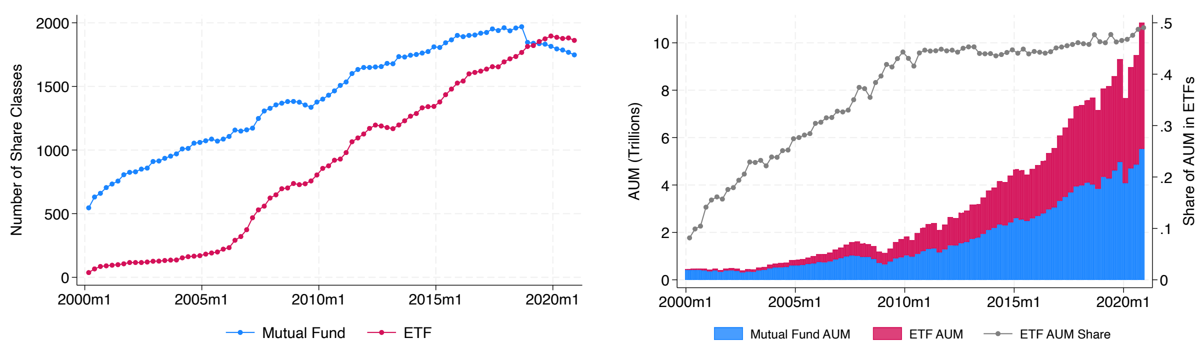


Figure 1 displays the growth of index funds over time. Panel (a) displays the number of index mutual funds and ETFs over time. Panel (b) displays the AUM of index mutual funds and ETFs, along with the ETF share of total index fund AUM.

One might expect such rapid entry would effectively erode market power, yet surprisingly, price dispersion has remained relatively constant over time. Figure 2 displays the distribution of fund expense ratios over the period with Panels (a) and (b) showing the equal-weighted and asset-weighted distributions, respectively. Panel (a) indicates

[15] Bhattacharya and Illanes (2022) use these data to study the design of defined contribution plans.

that average index fund expense ratios fell from 90 basis points in 2000 to roughly 65 basis points in 2020. The 10th percentile and 90th percentile of expense ratios have experienced similar declines, which indicates that the decline in average expense ratios has been driven by a general level shift in the distribution of fund expense ratios. The interdecile range has also remained relatively constant at 150-160 basis points over the bulk of our sample.

Figure 2: Distribution of Fund Expense Ratios over Time

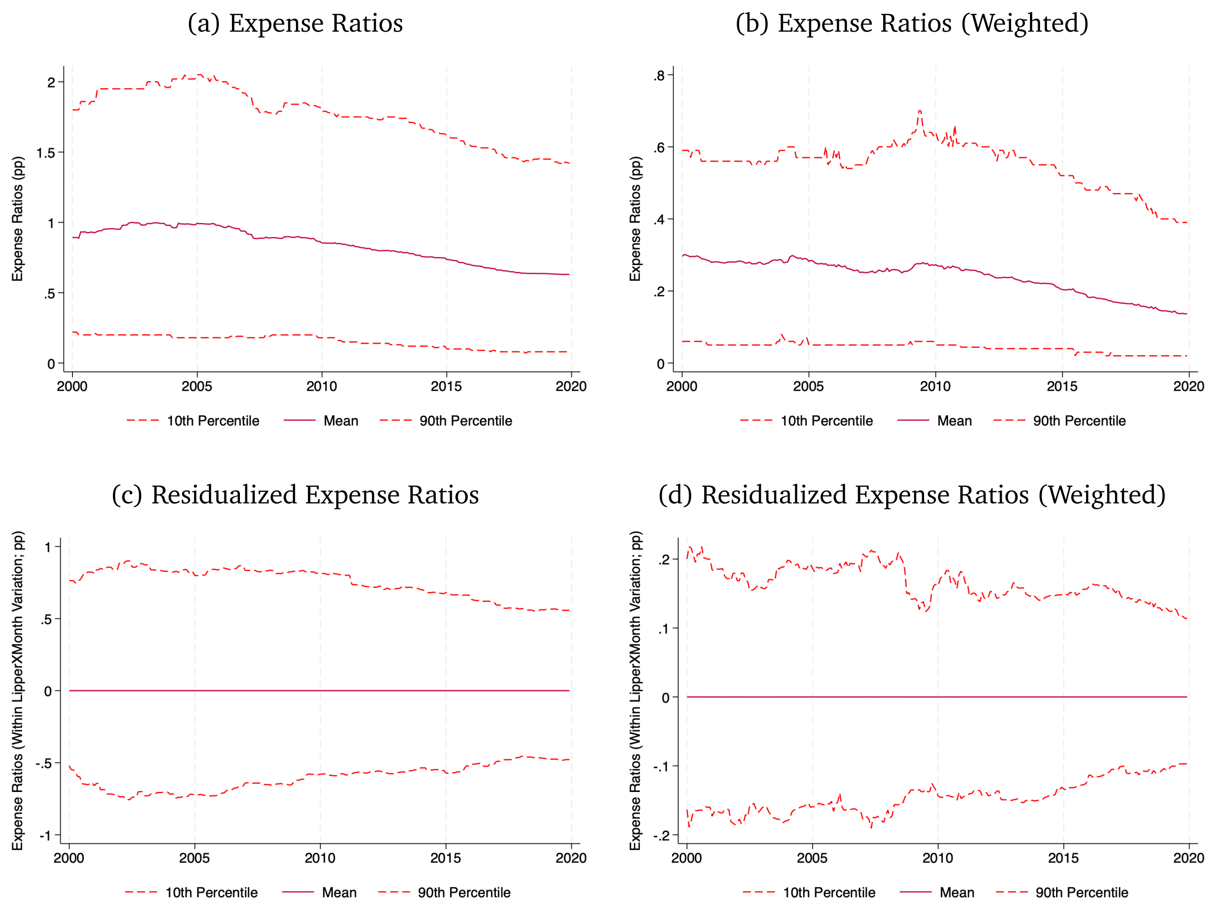


Figure 2 displays the distribution of index fund expense ratios over time. Panels (a) and (b) display the equal weighted and asset-weighted distribution of expense ratios. Panels (c) and (d) display the equal weighted and asset-weighted distribution of residualized expense ratios, where we residualize expense ratios by regressing them on Lipper Class $\times$ Month fixed effects. Panels (c) and (d) therefore display the within Lipper Class $\times$ Month variation in expense ratios.

Comparing the equal-weighted with the asset-weighted distributions provides prima facie evidence about investors' elasticity of demand. The asset-weighted distribution is shifted downwards relative to the equal-weighted distribution, which suggests investors are price sensitive. However, there is still substantial dispersion in expense ratios even when we weight by assets, which suggests that a large fraction of investors still purchase high-fee index funds.

Some of the observed dispersion in expense ratios could be driven by differences across index funds or asset classes. For example, index funds classified in Lipper as commodity-based metals funds are generally more expensive than funds classified as S&P 500 index funds. To account for these differences, we residualize expense ratios by regressing them on Lipper Class-by-month fixed effects and plot the residualized expense ratios in Figure 2 Panels (c) and (d). These fixed effects explain 35% of the variation in fund expense ratios and 79% of the variation in fund returns, suggesting that much of the fee dispersion is within Lipper Classes. Even after accounting for differences across Lipper Classes, funds in the 90th percentile are on average 1 percentage point more expensive than funds in the 10th percentile. Overall, the results show there has been substantial price dispersion for seemingly homogeneous products that has persisted over the 20-year period.

2.2.2 Potential Drivers of Price Dispersion

We wish to provide insight into the drivers of index fund market power and observed dispersion in prices. Here, we provide initial motivating evidence for three mechanisms that appear to be important: inertia, information frictions, and price discrimination.

Investor Inertia: It is well documented that investors exhibit inertia. Recent survey evidence indicates that roughly 12-18% of defined contribution plan investors update their portfolio each year.[16] Given the introduction of new low-fee ETFs, inertia could be quite costly for the 82–88% of investors who do not update their portfolios each year and could help explain the persistent dispersion in expense ratios.

We provide initial evidence on the role of investor inertia in index funds in two ways. First, we plot the ratio of new sales to total AUM at the market level in Figure 3. For the median market, new sales account for only 4% (3%) of AUM for retail (institutional) investors, suggesting that market shares are highly persistent. Second, we examine how fund flows respond to the introduction of new low-cost funds, separately using total sales in the overall sample and using the new-sales data alone. Since inertia

[16] See `https://www.ici.org/system/files/2021-09/21_rpt_recsurveyq2.pdf`. ICI reports rebalancing activity for the first half of the year, which we annualize by multiplying by two.

Figure 3: Mutual Fund Sales

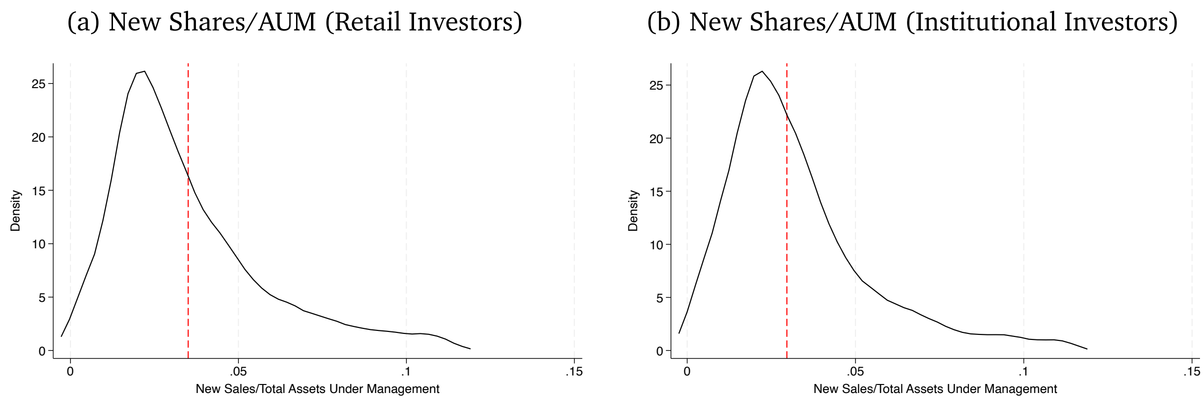


Figure 3a and Figure 3b display the distribution of the total net asset value of new shares purchased relative to total AUM calculated at the Lipper-class-by-month level for retail and institutional investors. To account for outliers we restrict the data set to those observations with positive sales, and we censor the distribution at the 99% level (weighted by assets). The red dashed lined in each figure corresponds to the median observation.

prevents investors from switching to cheaper new funds, the sluggishness of inflows into these funds provides a preliminary test of whether inertia plays an important role in this market. Moreover, since new sales represent choices by active investors, we expect the market share of new low-cost funds to be higher in the presence of inertia.

Figure 4 shows how the average market share of a newly launched low-cost fund within a Lipper Class evolves over time. We identify low-cost funds as those in the bottom quartile of the expense ratio distribution within their Lipper class at the time of launch. Then, we plot the average of these funds' market shares in each month since the fund was introduced. We plot market share in terms of total assets and new sales separately. The share of total assets invested in inexpensive funds is initially low and then rises over time, consistent with the fact that it takes multiple years for investors to switch. After five years, the average market share of a fund in the bottom quartile of the price distribution is still only 5%. We find that the share is much higher for new sales, which is also consistent with a large degree of inertia. The fact that demand for low-cost funds is not higher in the long run—even after most investors have made an active choice—may reflect information frictions or preference heterogeneity.

Figure 4: Market Share of a Newly Launched Low Cost Fund

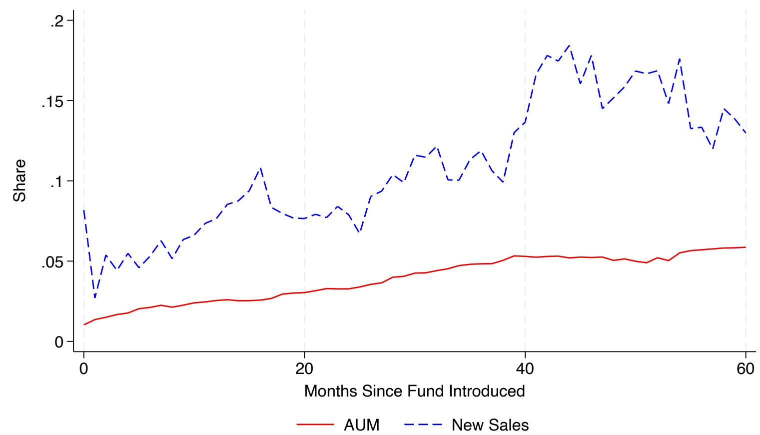


Figure 4 displays the average market share of a newly launched low-cost fund which survives at least 5 years by month since introduction. Low-cost funds are defined as those in the bottom quartile of the price distribution in their lipper class at the time of launch.

Information Frictions: Previous work highlights the importance of search and information frictions in the index fund market (Hortaçsu and Syverson, 2004). The 401(k) environment provides a useful contrast: participants are presented with a limited, transparent menu of 10-30 funds with standardized fee disclosure. Thus, we expect investors to be more price sensitive in this setting.

We examine price dispersion for 401(k) plans in Figure A1. While the interdecile range of asset-weighted expense ratios is about 40 basis points in the broad market for index funds, the range is only 8 basis points at the end of the sample for 401(k) plans. Given that preference heterogeneity is likely similar in these two settings, lower price dispersion in 401(k) plans provides initial evidence that information frictions play an important role in the index fund market.

Price Discrimination: Index fund managers often issue multiple funds and ETFs that track the same index or underlying portfolio. In particular, mutual funds often have a share-class structure, which allows intermediaries to explicitly price discriminate across investor types. Institutional share classes typically charge lower fees than retail ones, reflecting higher price sensitivity among institutional investors. To explore this channel, for a given underlying portfolio (identified in the data as crsp_portno) and year, we calculate the difference between the average expense ratio of retail mutual funds and that of institutional mutual funds. Figure 5 displays the distribution of this difference for those portfolios that are held by at least one retail and one institutional fund. On average, an institutional fund charges 94 basis points less than the retail fund within the same portfolio. These results suggest that part of the observed dispersion in expense ratios is driven by the ability of index managers to segment the market and further exercise their market power.

Figure 5: Within Portfolio Variation in Expense Ratios

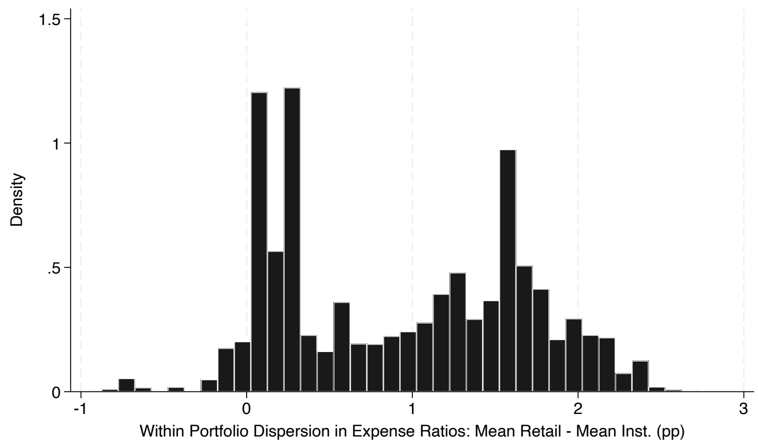


Figure 5 displays the within portfolio dispersion in expense ratios across retail and institutional funds. We focus on differences between retail and institutional funds. For a given underlying portfolio (identified in the data as $crsp\_portno$) and moment in time, we calculate the difference in the average expense ratio of retail funds and that of institutional funds for those portfolios that are held by at least one retail and one institutional fund. Observations are at the fund portfolio-by-year level.

3 Framework

We develop a dynamic model of supply and demand for index funds featuring inertia, information frictions, and preference heterogeneity. Our objective is to provide new insight into the mechanisms driving the persistent price dispersion we observe in the data. The baseline version of the model includes two types of agents: heterogeneous investors with demand for index funds and fund managers in charge of pricing index funds. We also consider an extension of the model in Appendix E where we introduce financial advisers who potentially distort the investment decisions of investors due to conflicts of interest.

3.1 Investors: Demand for Index Funds

We model index fund choice as a discrete choice among individual branded index funds within a given investment category or asset class (e.g., small-cap value, mid-cap growth) as classified by Lipper, taking the investor's overall asset allocation across categories as given. For example, we model whether an investor chooses a Vanguard or a BlackRock S&P 500 index fund, but not the investor's initial decision of whether, or how much, to invest in S&P 500 index funds rather than another asset class. In our counterfactuals, we therefore assume that changes in expense ratios reallocate demand across funds within a category/asset class without altering investors' allocations across categories.

There are two types of investors: retail and institutional. We denote investor type by (T) such that $T \in \{R, I\}$ where R denotes retail investors and I denotes institutional investors. Investor-types differ with respect to their preference parameters, frictions (e.g., inertia and information frictions), and product availability. That is, the set of index funds available to retail investors is potentially different from the set of index funds available to institutional investors.

3.1.1 Investor Preferences

Investor i's indirect utility from choosing fund j at time t is given by:

$$u_{i,j,t} = \underbrace{-p_{j,t} + X'_{j,t}\theta_{T(i)} + \xi_{T(i),j,t}}_{\bar{u}_{T(i),j,t}} + \sigma_{\epsilon,T(i)}\epsilon_{i,j,t}.$$

The term $-p_{j,t}$ reflects the dis-utility investors get from paying expense ratio $p_{j,t}$ where, without any loss in generality, we normalize the coefficient to one. The term $X'_{j,t}\theta_{T(i)}$ measures the utility generated by other fund characteristics $X_{j,t}$ where $\theta_{T(i)}$ captures investor preferences over those characteristics.

The indirect utility function includes two latent terms. The term $\xi_{T(i),j,t}$ measures unobserved product characteristics that are commonly valued among investors of type T. The term $\epsilon_{i,j,t}$ captures preference heterogeneity which varies across investors. This implies that index funds are horizontally differentiated such that any two investors may disagree on which index fund is the best. The degree of product differentiation also varies across type T which is captured by the term $\sigma_{\epsilon,T(i)}$. The horizontal differentiation is also captured by $\theta_{T(i)}$ as different types of investors disagree on the relative importance of the expense ratio and other fund characteristics.[17]

3.1.2 Information Frictions

Investors may not research all funds or fully understand fund characteristics such as the expense ratio. We assume that each investor faces information frictions such that the investor's perceived utility when selecting a fund may differ from the realized utility from owning a fund. Investors choose index funds based on their perceived utility $\tilde{u}_{i,j,t}$, which is a noisy signal of their indirect utility function:

$$\begin{aligned}\tilde{u}_{i,j,t} &= u_{i,j,t} + \nu_{i,j,t} \\ &= \bar{u}_{T(i),j,t} + \sigma_{\epsilon,T(i)}\epsilon_{i,j,t} + \nu_{i,j,t}.\end{aligned} \tag{1}$$

The term $\nu_{i,j,t}$ reflects idiosyncratic choice/information frictions that cause individuals to not always choose their preferred index fund. Let $Var[\nu_{i,j,t}] = \frac{\pi^2}{6}\sigma^2_{\nu,T(i)}$, where $\sigma_{\nu,T(i)}$ reflects the degree of information frictions. An increase in information frictions (larger $\sigma_{\nu,T(i)}$) makes individuals less likely to choose the lowest cost fund. In the second line of Eq. (1), we write perceived utility in terms of the common component of utility, $\bar{u}_{T(i),j,t}$.

This reduced-form representation of information frictions is consistent with rational inattention models that have been increasingly used to explain demand for complex financial products (e.g., Kacperczyk et al., 2016; Brown and Jeon, 2024). As shown in Appendix B, under common priors across options and homogeneous information costs within type, the rational inattention model yields expected utility equivalent to Eq. (1). The variance of $\nu_{i,j,t}$ can then be interpreted as proportional to the unit cost of information.

Following the literature we assume that $\nu_{i,j,t}$ is distributed according to Cardell (1997); i.e., $\nu_{i,j,t} \sim C(\frac{\sigma_{\epsilon,T(i)}}{\sigma_{\eta,T(i)}}, \sigma_{\epsilon,T(i)})$ such that the composite error term $\eta_{i,j,t} = \sigma_{\epsilon,T(i)}\epsilon_{i,j,t} + \nu_{i,j,t}$ is distributed Type 1 Extreme Value with scale parameter $\sigma_{\eta,T(i)} = \left(\sigma^2_{\nu,T(i)} + \sigma^2_{\epsilon,T(i)}\right)^{1/2}$. We can then write investor i's perceived utility as:

$$\tilde{u}_{i,j,t} = \bar{u}_{T(i),j,t} + \sigma_{\eta,T(i)}\eta_{i,j,t}. \tag{2}$$

[17] It is straightforward to incorporate additional taste differences by allowing random coefficients on the expense ratio, for example.

3.1.3 Fund Choice

When individuals make an active choice (and are not subject to inertia as described below), they maximize perceived utility given by Eq. (2). The market share of fund j among active type T investors at time t is given by:

$$s_{T,j,t} = \frac{\exp\left(\frac{-p_{j,t}+X'_{j,t}\theta_T+\xi_{j,T,t}}{\sigma_{\eta,T}}\right)}{\sum_{l\in\mathcal{J}_{T,m(j),t}}\exp\left(\frac{-p_{l,t}+X'_{l,t}\theta_T+\xi_{l,T,t}}{\sigma_{\eta,T}}\right)}. \tag{3}$$

The set $\mathcal{J}_{T,m(j),t}$ denotes the investor's consideration set: the set of index funds available to a type T investor in market $m(j)$ at time t. Recall that our model is a model of index fund choice, conditional on an investor's choice to buy a fund in a given market. The above equation is a core part of our estimation strategy below, where we separately identify investors' preferences (θ_T) as well as decompose the error term into two components: one due to information frictions ($\sigma_{\nu,T}$) and the other due to product differentiation ($\sigma_{\epsilon,T}$).

3.1.4 Inertia

In each period, with some probability an investor is either active or inactive, similar to the setup in Beggs and Klemperer (1992). Inactive investors simply maintain their investments from the previous period, while active investors update their portfolios to maximize their objective function. We assume that the probability an investor is inactive varies across investor types, markets, and time, but is constant across investors and funds within a given market and period. The probability that a type T investor in market m is inactive in a given period t is denoted by $\phi_{T,m,t}$ and the probability she is active is $1-\phi_{T,m,t}$. This model of inertia is consistent with the idea that investors may only check their portfolio at specific intervals, such as when they file taxes or receive annual reports (e.g., Benartzi and Thaler, 1995).

When active, investors choose the fund that maximizes their perceived utility such that choice probabilities and market shares are given by Eq. (3). Thus, investors are either myopic or they assume that their preferences and the product space will be constant over time. Given that $\phi_{T,m(j),t}$ of investors of type T are inactive each period, the total assets under management of fund j held by type T investors at time t can be written as:

$$AUM_{T,j,t} = \underbrace{\phi_{T,m(j),t} AUM_{T,j,t-1}(1 + r_{m(j),t-1})}_{AUM^{Inactive}_{T,j,t}} + \underbrace{(1 - \phi_{T,m(j),t}) M_{T,m(j),t} s_{T,j,t}}_{AUM^{Active}_{T,j,t}}. \tag{4}$$

The term $AUM^{Inactive}_{T,j,t} \equiv \phi_{T,m(j),t} AUM_{T,j,t-1}(1 + r_{m(j),t-1})$ captures demand from inactive investors who simply maintain their holdings from the previous period, which grow based on the return of fund j over the period $t-1$ to t, denoted $r_{m(j),t-1}$. We assume, in part for ease of exposition, that fund returns are constant across all index funds in a given market such that $r_{j,t} = r_{m(j),t}$. The term $AUM^{Active}_{T,j,t} \equiv (1 - \phi_{T,m,t}) M_{T,m(j),t} s_{T,j,t}$ measures demand from active investors, where $M_{T,m(j),t}$ denotes the total assets invested in market $m(j)$ held by investors of type T at time t.

3.2 Index Fund Managers: Supply of Index Funds

Index funds are created and managed by a set of differentiated index fund managers k. Index fund managers create three different types of products, retail mutual funds, institutional mutual funds, and ETFs. The products are functionally equivalent except that retail mutual funds are purchased only by retail investors, institutional mutual funds, only by institutional investors, and ETFs by both types.

Index fund managers' per-period profits in a market m are given by

$$\Pi_{k,m,t} = \sum_{j \in \mathcal{K}_{k,m}} (AUM_{R,j,t} + AUM_{I,j,t})(p_{j,t} - c_j),$$

where $\mathcal{K}_{k,m}$ denotes the set of index funds sold by index fund manager k in market m. The terms $AUM_{R,j,t}$ and $AUM_{I,j,t}$ denote demand for fund j from retail and institutional and retail investors, and c_j is the marginal cost of operating fund j.

We assume that index fund managers play a differentiated, multi-product, dynamic, Nash-Bertrand, expense-ratio-setting game. Ex ante, they know the distribution of the inertia parameter, but do not know the realization of the parameter. Let $\mathbf{p}_{k,t}$ be the vector of prices for funds managed by k in period t. An index fund manager's problem is to set the sequence of prices $\mathbf{p}_{k,t}$, $\mathbf{p}_{k,t+1}, \ldots$ to maximize the expected present discounted value of future profits.

$$\max_{\mathbf{p}_{k,t}, \mathbf{p}_{k,t+1} \cdots | \mathbf{p}_{-k,t}, \mathbf{p}_{-k,t+1} \cdots} \sum_{\tau=t}^{\infty} \beta^{\tau - t} \sum_{j \in \mathcal{K}_{k,m}} \mathbb{E}\left[(AUM_{R,j,\tau} + AUM_{I,j,\tau})(p_{j,\tau} - c_j)\right] \tag{5}$$

where the expectation is over inertia ($\phi_{T,m,t}$) and fund returns ($r_{m,t}$). We assume that returns and inertia are distributed i.i.d. over time, constant within a market, and are uncorrelated such that $\mathbb{E}[\phi_{T,m,t}] = \bar{\phi}_{T,m}$ and $\mathbb{E}[r_{m,t}] = \bar{r}_m$.

For tractability, we assume that fund managers observe and condition on the full sequence of competitors' prices when setting their own prices. This assumption simplifies the suppliers' problem because it rules out strategic pricing interactions where a firm may change its price today to influence the future prices of its competitors. We believe this assumption is reasonable in the index fund setting for two reasons. First, most funds have extremely small market shares (approximately 70% have market shares below 1%), so dynamic strategic interactions are likely second-order. Second, prices appear quite sticky. As of 2020, only 3.5% funds (weighted by assets) charged more than 10 basis points lower than the price in 2015. The sluggish price adjustment implies that a manager's current pricing decision is unlikely to trigger meaningful price responses from competitors. It is important to note, however, that this assumption does not eliminate the dynamic considerations in the pricing problem: firms still internalize that current pricing decisions impact future demand through investor inertia, and thus future profitability.

To develop intuition for how firms set prices with consumer inertia, we first consider the simple case where an index fund manager operates a single retail mutual fund. We then extend our model to the multi-product and multi-investor-type setting.

3.2.1 Single Product Retail Mutual Fund Manager

Consider a fund manager's profit maximization problem for retail mutual fund j. The corresponding first order condition for price $p_{j,t}$ is:

$$0 = \frac{\partial AUM_{R,j,t}}{\partial p_{j,t}} \left\{ \underbrace{p_{j,t} - c_j}_{\text{Static Profits}} + \underbrace{\sum_{\tau=t+1}^{\infty} \left[\beta(1+\bar{r}_{m(j)})\bar{\phi}_{R,m(j)}\right]^{\tau-t} (p_{j,\tau} - c_j)}_{\text{Present Value of Future Profits}} \right\} + AUM_{R,j,t}.$$

The first order condition is standard except for the term $\sum_{\tau=t+1}^{\infty}[\beta(1+\bar{r}_{m(j)})\ \bar{\phi}_{R,m(j)}]^{\tau-t}(p_{j,\tau} - c_j)$, which captures the effects of inertia. For every investor the fund attracts today (period t), there is a $\bar{\phi}_{R,m(j)}$ chance she remains in period $t+1$, a $\bar{\phi}^2_{R,m(j)}$ chance she remains for at least two periods, and so on. Furthermore, inactive investors' assets are expected to grow based on fund expected returns and market size growth rate, $\bar{r}_{m(j)}$.

In the static problem (e.g., with no inertia), a firm's assets today do not affect its assets tomorrow such that $\frac{\partial AUM_{R,j,\tau}}{\partial AUM_{R,j,t}} = 0, \forall \tau \neq t$. As a result of investor inertia, $\frac{\partial AUM_{R,j,\tau}}{\partial AUM_{R,j,t}} = [(1+\bar{r}_{m(j)})\phi_{R,m(j)}]^{\tau-t}$. Thus, when setting prices, firms account for how changing prices impact both current and future demand. Inertia puts downward pressure on prices because firms have stronger incentives to attract new investors who are more likely to remain with the fund in future periods. This is often referred to as the incentive to "invest" in new customers. However, inertia also makes demand less elastic. To see this, note that:

$$\frac{\partial AUM_{R,j,t}}{\partial p_{j,t}} = (1-\phi_{R,m(j),t})M_{R,m(j),t}\frac{\partial s_{R,j,t}}{\partial p_{j,t}}.$$

Consequently, all else equal, higher inertia will cause firms to want to set higher prices. This is often referred to as the incentive to "harvest" current consumers.

We study a steady-state equilibrium where firms' market shares are constant over time such that $p_{j,t} = p_j$ and $s_{j,t} = s_j \; \forall j$ and $M_{R,m(j),t} = M_{R,m(j),t-1}(1+r_{m(j),t-1})$. Thus, dropping the t subscripts and noting that $p_{j,t} - c_j + \sum_{\tau=t+1}^{\infty}[\beta(1+\bar{r}_{m(j)})\bar{\phi}_{R,m(j)}]^{\tau-t}(p_{j,\tau} - c_j) = \frac{p_j - c_j}{1-\beta\bar{\phi}_{R,m(j)}(1+\bar{r}_{m(j)})}$, the manager's first order condition simplifies to:

$$\frac{p_j - c_j}{p_j} = \frac{1-\beta(1+\bar{r}_{m(j)})\bar{\phi}_{R,m(j)}}{1-\bar{\phi}_{R,m(j)}} \times \frac{1}{-\varepsilon_j^D}, \tag{6}$$

where ε_j^D denotes the elasticity of demand of product j for active investors given prices for all other funds.

A couple of features of this first order condition are worth noting. First, if inertia is equal to zero in expectation (i.e., $\bar{\phi}_{R,m(j)} = 0$), the first order condition simplifies to a standard static first order condition. Second, given a growth-adjusted discount factor of 1 (i.e., $\beta(1+\bar{r}_{m(j)}) = 1$ such that the growth-adjusted discount rate is zero), the dynamic pricing condition simplifies to the standard static first order condition even if inertia is greater than zero, and the share of inactive investors does not affect steady-state prices. Note that this result would hold for any generic demand system given how inertia works in our model. While the incentive to invest in new consumers can lower prices and the incentive to harvest existing customers can raise prices, these forces would perfectly offset when the growth-adjusted discount factor is equal to 1.[18]

[18] The CAPM model would imply that we would expect the growth-adjusted discount factor to be close to 1 because, if expected returns are equal to required returns, it should be the case that $\beta_{m(j)}$ varies at the market level such that $\beta_m = \frac{1}{1+\bar{r}_m}$.

In practice, we would expect the growth-adjusted discount factor to be slightly less than one, for example, because fee-adjusted growth is lower than the required return. In this case, inertia increases markups more than in the static model without inertia, with managers placing more value on harvesting existing investors than on investing in future ones as the growth-adjusted discount factor decreases.

3.2.2 Multi-product Managers

In the data, index fund managers often issue multiple retail funds, institutional funds, and ETFs in a single market. Consider the profit maximization problem of manager k who issues set of index funds $\mathcal{K}_{k,T,m}$ available to type T investors in market m. The corresponding first order condition with respect to $p_{j,t}$in steady state is given by

$$\begin{aligned} 0 =& \mathbf{1}(j \in \mathcal{K}_{k,R,m}) \frac{M_{R,m(j)}}{M_{I,m(j)}} s_{R,j} \left[1 - \frac{\frac{1}{\sigma_{\eta,R}}(1-\bar{\phi}_{R,m(j)})}{1-\beta(1+\bar{r}_{m(j)})\bar{\phi}_{R,m(j)}} \left(p_j - c_j - \sum_{l \in \mathcal{K}_{k,R,m(j)}} s_{R,l}\left(p_l - c_l\right) \right) \right] \\ &+ \mathbf{1}(j \in \mathcal{K}_{k,I,m}) s_{I,j} \left[1 - \frac{\frac{1}{\sigma_{\eta,I}}(1-\bar{\phi}_{I,m(j)})}{1-\beta(1+\bar{r}_{m(j)})\bar{\phi}_{I,m(j)}} \left(p_j - c_j - \sum_{l \in \mathcal{K}_{k,I,m(j)}} s_{I,l}\left(p_l - c_l\right) \right) \right], \end{aligned} \tag{7}$$

where $\frac{M_{I,m(j)}}{M_{R,m(j)}}$ denotes the relative size of the institutional and retail markets which are constant in steady state. As before, if either $\bar{\phi}_{R,m(j)} = \bar{\phi}_{I,m(j)} = 0$ or $\beta(1+\bar{r}_{m(j)}) = 1$, then the firm's first-order condition for setting prices in the dynamic model simplifies to the standard first-order condition in the static model.

4 Estimation

We estimate our structural model of demand and supply for index funds using the data sets described in Section 2.1 in three steps. First, using data on new sales, we recover inertia. Second, once inertia is pinned down, we jointly estimate preferences and information frictions, using additional data on choices in 401(k) plans. Finally, with the demand parameters in hand, we estimate the index fund managers' marginal costs of operating index funds based on the pricing conditions implied by dynamic Nash-Bertrand competition among multi-product managers.

4.1 Investor Demand

4.1.1 Inertia

It is not possible to directly calculate inertia using equation (4) since active assets under management are not observed. Our key insight is that new sales provide a proxy for active inflows to the fund.

Any new sales of a fund must come from active investors who did not choose this fund in the previous period. We can thus write new sales by investor type T for fund j in time t as:

$$NewSales_{T,j,t} = s_{T,j,t}(1 - \phi_{T,m(j),t}) \sum_{l \neq j} AUM_{T,l,t-1}(1 + r_{m(l),t-1}). \tag{8}$$

New sales represent active investors who switched to fund j whereas active assets under management for fund j include both those active investors who switched to fund j and those active investors who remained invested in it.[19] Active AUM in fund j at time t is given by

$$AUM^{Active}_{T,j,t} = s_{T,j,t}(1 - \phi_{T,m(j),t}) \sum_{l \in \mathcal{J}_{T,m,t}} AUM_{T,l,t-1}(1 + r_{m(l),t-1}). \tag{9}$$

Then, dividing equation (9) by equation (8), we can now express active AUM as a function of new sales and past AUM:

$$AUM^{Active}_{T,j,t} = NewSales_{T,j,t} \frac{\sum_{l \in \mathcal{J}_{T,m,t}} AUM_{T,l,t-1}}{\sum_{l \neq j} AUM_{T,l,t-1}}. \tag{10}$$

Substituting equation (10) into equation (4), we can estimate the market-by-time level inertia parameter for each investor type as:

$$\hat{\phi}_{T,m,t} = \frac{1}{\|\mathcal{J}_{T,m,t}\|} \sum_{j \in \mathcal{J}_{T,m,t}} \frac{AUM_{T,j,t} - NewSales_{T,j,t} \frac{\sum_{l \in \mathcal{J}_{T,m,t}} AUM_{T,j,t-1}}{\sum_{l \neq j} AUM_{T,j,t-1}}}{AUM_{T,j,t-1}(1 + r_{j,t-1})}, \tag{11}$$

where $\|\mathcal{J}_{T,m,t}\|$ represents the number of funds available for type T investors in market m and time t. We make two assumptions when implementing this estimation step. First, as discussed in Section 2.1, new sales data are available only for mutual funds.

[19] Note that both expressions in equations (9) and (8) are adjusted by $(1 + r_{m(j),t-1})$, reflecting our assumption about the growth due to market-level returns.

Figure 6: Investor Inertia

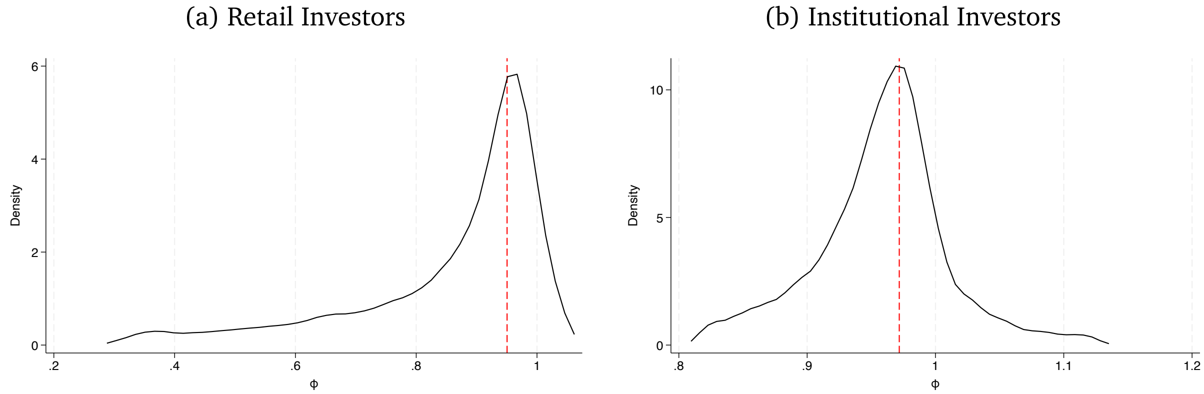


Figure 6a and Figure 6b display the distribution of inertia at the market-by-month level over the period 2000-2020 for retail and institutional investors. To account for outliers, we censor the distribution at the AUM-weighted 5th and 95th percentiles. The red dashed lined in each figure corresponds to the AUM-weighted median observation.

Therefore, when computing the inertia parameter based on equation (11), we take the average only across mutual funds. We then assign the same inertia level to ETFs in the corresponding market. Second, to account for growth in total market size, we work with growth-adjusted versions of current new sales and AUM in equations (4), (8), and (9).[20] This adjustment ensures that growth in assets that is unrelated to investor switching does not mechanically bias the calculation of active demand. Alternatively, we develop models that explicitly specify how assets are newly added to the market and show in Appendix C that our estimates of inertia are robust to these specifications.

We report the estimated distribution of inertia in Figure 6, for retail investors in Panel (a) and for institutional investors in Panel (b). The red dashed lines in each figure correspond to the AUM-weighted medians. In the median Lipper class and month, 95.0% of retail investors are inactive. Put differently, 46% ($= 1 - 0.950^{12}$) of retail investors update their portfolios at least once per year. We find similar rates of inertia among institutional index fund investors: in the median Lipper class and month, 97.2% of institutional investors are inactive. Retail investors also display greater heterogeneity in inertia.

One caveat with our approach is that, consistent with the structure of the model

[20]Specifically, we measure the growth rate as $g_{T,m,t} = \frac{\sum_{j \in \mathcal{J}_{T,m,t}} AUM_{T,j,t}}{\sum_{j \in \mathcal{J}_{T,m,t}} AUM_{T,j,t-1}(1+r_{j,t-1})}$. Then, we divide $AUM_{T,j,t}$, $NewSales_{T,j,t}$, and $AUM^{Active}_{T,j,t}$ in equations (4), (8), and (9) by $g_{T,m,t}$.

and as implied by equation (10), the probability that an active investor selects fund j at time t is assumed to be the same across all investors, regardless of which fund they held in the previous period. Thus, while we allow for persistent unobserved common product quality, we assume that idiosyncratic investor utility shocks are independent and identically distributed over time. Therefore, a potential concern is that some of the persistence we observe in the data could reflect investor-specific persistent tastes rather than investor inactivity. We address this in two ways.

First, new product entry provides untargeted moments that help us distinguish between these two mechanisms. If the observed persistence in demand were driven by inertia, a new product entrant would be expected to have a low initial market share that grows rapidly over the first few periods before stabilizing at its steady state. In contrast, if the persistence were instead driven by investor-specific utility shocks, new entrants would reach their steady-state market shares immediately upon entry. Figure 7 compares the model-implied evolution of a low-cost new fund's market share (dashed line) with the corresponding patterns observed in the data (solid line). Specifically, we simulate the market share of a fund with a steady-state share of 6%, assuming that 95.0% of investors are inactive each month. The two align closely—with low initial market shares that grow quickly before leveling off—implying that inertia is the main mechanism.

Second, we employ an alternative estimation approach in which inertia is identified through the persistence of exogenous demand shocks orthogonal to product quality. This approach yields quantitatively similar estimates of inertia. We discuss the method and estimates under this approach in further detail in Section 4.3 and present the results in Appendix Table A1.

Figure 7: Market Share of a Newly Launched Low Cost Fund: Model vs. Data

Figure 7 displays the average market share of a newly launched low-cost fund in the data versus the model. The dark red line shows the market share of a newly launched low-cost which survives at least 5 years by month since introduction. Low-cost funds are defined as those in the bottom quartile of the price distribution in their lipper class at the time of launch. The dashed red line displays the market share of a newly launched low cost fund as implied by our estimates of inertia. We simulate the market share assuming that, consistent with our estimates for retail investors, 95.0% of investors are inactive each month. We also keep all funds and characteristics fixed and assume that the steady-state market share for the new entrant is 6%.

4.1.2 Investor Preferences and Information Frictions

Given the inertia estimates from Section 4.1.1, we can estimate demand when investors make an active choice. We jointly estimate investor preferences and information frictions, taking the inertia estimates as given. A well-known empirical challenge is that information frictions and preference heterogeneity both appear as reduced price sensitivity in the data, making them difficult to distinguish using choice data alone. To address this, we draw on investor choices in the 401(k) setting, where information frictions are likely minimal. In 401(k) plans, investors choose from a simplified and transparent menu of funds, and regulations require standardized disclosures of expense ratios and performance. Within each broad asset class (e.g. large-cap US equities), the average plan offers only about two index funds.[21]

When investors face market frictions, the logged market share of fund j at time t

[21]Broad asset classes are defined as per Brightscope and include: allocation funds, alternatives, bonds, cash, international equities, large cap equities, mid cap equities, and small cap equities.

among type T investors relative to the outside good is given by:

$$\ln s_{T,j,t} - \ln s_{T,0,t} = \frac{1}{\sigma_{\eta,T}} \left(-p_{j,t} + X'_{j,t}\theta_T + \xi_{T,j,t}\right). \quad (12)$$

When investors do not face market frictions (as in the 401(k) setting), the corresponding logged market share—denoted in capitals as $S_{T,j,t}$—relative to the outside good is:

$$\ln S_{T,j,t} - \ln S_{T,0,t} = \frac{1}{\sigma_{\epsilon,T}} \left(-p_{j,t} + X'_{j,t}\theta_T + \xi_{T,j,t}\right). \quad (13)$$

Given the observed active market shares of funds with and without information frictions, along with product characteristics $p_{j,t}$ and $X_{j,t}$, we jointly estimate investor preferences (θ_T), and two components that introduce noise into choices: preference heterogeneity ($\sigma_{\epsilon,T}$) and information frictions ($\sigma_{\nu,T}$).

We jointly estimate equations (12) and (13) using the Generalized Method of Moments (GMM) and the standard OLS moment conditions. We calculate market shares when investors face information frictions using the full dataset, defining market shares at the Lipper-class-by-month level. We also calculate market shares in the absence of information frictions using investor choices in the 401(k) setting at the plan-by-year-by-asset-class level.

There are two empirical challenges associated with directly estimating equations (12) and (13). First, fund expense ratios ($p_{j,t}$) are potentially endogenous. For example, if an index fund manager anticipates high latent demand for their fund (i.e., high $\xi_{T,j,t}$), they may optimally choose to charge a higher expense ratio. This behavior would bias OLS estimates of price sensitivity, α_T, downward—making investors appear less price-sensitive than they truly are. To address this potential endogeneity, we instrument for expense ratios using Hausman (1996)-type instruments. Specifically, the expense ratio that an index fund manager k charges on its fund j at time t is instrumented with the average expense ratio that manager k charges on all of its other funds in markets other than $m(j)$ at time t. For example, we instrument for the fee that BlackRock charges on its large-cap value funds using the average fees it charges on its non–large-cap value funds, such as BlackRock's high-yield bond funds. The instrument is relevant because BlackRock's costs of managing its large-cap equity funds are correlated with its costs of managing its high-yield bond funds. The instrument is exogenous under the assumption that the fee BlackRock charges on its high-yield bond funds is uncorrelated with demand shocks for its large-cap value funds.

The second challenge in directly estimating equations (12) and (13) is that the left-hand-side variables in both equations are active market shares, which are not always directly observable. In the 401(k) setting, we restrict attention to plans in their first year of introduction, such that all investors are active in the initial period.

In the non-401(k) setting, we construct active market shares using our estimates of $\phi_{T,m,t}$. While the realized active market shares depend on the specific realization of $\phi_{T,m,t}$, our demand system in equation (12) predicts expected choice probabilities among active investors, as the right-hand side of the equation does not depend on $\phi_{T,m,t}$. That is, ϕ affects only the scale of active AUM, not the relative choices made by active investors. Moreover, our estimates of $\phi_{T,m,t}$ are potentially subject to measurement error from new sales data and sampling noise. Therefore, for each market m, we integrate over the empirical distribution of feasible values of $\phi_{T,m,t}$ recovered in Section 4.1.1 to compute the expected active market shares $s_{T,j,t}$.[22]

We stack the OLS moment conditions from the retail sample and from the 401(k) sample and minimize the corresponding GMM objective function to jointly estimate the retail investors' preferences and information frictions. Estimation follows a two-step procedure: the first step uses the inverse of the inner product of the instrument matrix as the weighting matrix; and the second step uses the inverse of the variance matrix from the first step as the optimal weighting matrix. Because the full set of fixed effects $(\mu, \tilde{\mu})$ includes more than 40,000 parameters, we use a nested algorithm. In the outer loop, we optimize over the nonlinear parameter $\sigma_{\eta,R}$ and σ_ϵ using a BFGS quasi-Newton algorithm. In the inner loop, given values of the nonlinear parameters, we optimize over the linear parameters θ_R, μ and $\tilde{\mu}$ using a linear regression with fixed effects.

Our empirical specification controls for: 1-, 3-, 6-, and 12-month cumulative returns; the log number of funds offered by the management company; the standard deviation of daily fund returns over the past 12 months; and indicator variables for whether the fund is an ETF, has a front load, or has a rear load. We also include market-by-time fixed effects, defined at the Lipper-class-by-month level for the full dataset and

[22]In practice, we compute expected active AUM as

$$\mathbb{E}\left[AUM^{Active}_{T,j,t}\right] = \int \left[AUM_{T,j,t} - \phi AUM_{T,j,t-1}(1+r_{j,t})\right] \times dF(\phi \,|\, AUM_{T,j,t} \geq \phi AUM_{T,j,t-1}(1+r_{j,t}))\,.$$

Expected active market shares for fund j are then constructed by dividing expected active AUM for fund j by the sum of expected active AUM across all funds in the market.

Table 2: Investor Preferences and Information Frictions

	Retail	Institutional	401k
Expense Ratio	-3.143^{***}	-4.929^{***}	-6.506^{***}
	(0.036)	(0.093)	(0.506)
Observations	111,152	109,758	32,777
Mkt-Time FE	X	X	X
IV	X	X	X
Elasticity of Demand	1.6	2.5	3.3

Note: Table 2 displays the baseline demand estimates. Observations are at the index fund-by-month-by-investor type (i.e., retail vs. institutional) level in columns (1) and (2). Observations are at the index fund-by-401(k) plan-by-asset-class-by-year level in column (3). In all specifications we control for: the log number of funds offered by the management company; the standard deviation of daily fund returns over the past 12 months; 1-, 3-, 6-, and 12-month cumulative returns; and whether the fund is an ETF, has a front load, or has a rear load. Standard errors in parenthesis are clustered at portfolio-by-month level in column (1) and (2). GMM asymptotic standard errors are in parenthesis in column (1) and (3). *** p<0.01, ** p<0.05, * p<0.10.

at the plan-by-year-by-asset-class level for the 401(k) setting.

Table 2 displays our baseline demand estimates. We report the estimated price coefficient for retail investors ($-\frac{1}{\sigma_{\eta,R}}$) in column (1), for institutional investors ($-\frac{1}{\sigma_{\eta,I}}$) in column (2), and for the retail investors in the 401(k) setting ($-\frac{1}{\sigma_{\epsilon}}$) in column (3). In each specification, as expected, we estimate a negative and significant relationship between expense ratios and demand. As shown in the bottom row of the table, the implied elasticity of demand under information frictions is 1.6 for retail investors and 2.5 for institutional investors. Consistent with intuition, institutional investor demand is substantially more elastic than retail investor demand (by 56%). In the 401(k) setting, retail investors are more than twice as elastic as in the main sample, implying severe information frictions. We estimate the preference heterogeneity parameter σ_ϵ to be 0.16 ($= \frac{1}{6.506}$), so the information friction parameter for retail investors $\sigma_{\nu,R}$ is 0.28 ($= \sqrt{\frac{1}{3.143^2} - \frac{1}{6.506^2}}$). This indicates that information frictions are roughly 75% ($= (0.28 - 0.16)/0.16$) more important than preference heterogeneity in the index fund market.

4.2 Index Fund Managers: Supply

We estimate the supply side of the model by inverting the index fund manager's first order condition to solve for the marginal cost that rationalizes the manager's chosen expense ratio. Given our demand specifications, we rewrite the first order condition in Eq. (7) in matrix form as

$$M_{R,t}\mathbf{s}_{R,t} + M_{I,t}\mathbf{s}_{I,t} = (M_{R,t}\Omega_{R,t} + M_{I,t}\Omega_{I,t}) \times (\mathbf{p}_t - \mathbf{c}_t)$$

where elements of matrix $\Omega_{T,t}(\mathbf{p})$ are given by

$$\Omega_{(j,l)}(\mathbf{p}) = \begin{cases} -\frac{1-\bar{\phi}_{T,m(j)}}{1-\beta(1+\bar{r}_{m(j)})\bar{\phi}_{T,m(j)}} \frac{\partial s_j}{\partial p_l}(\mathbf{p}_t) & \text{if } j,l \in \mathcal{K}_{k(j),m(j)} \\ 0 & \text{else} \end{cases}$$

In the data we directly observe the scalars $M_{R,t}$ and $M_{I,t}$ and the vectors $\mathbf{s}_R$, $\mathbf{s}_I$, and $\mathbf{p}$. Given $\beta \times (1 + \bar{r})$, we can then use our inertia and demand estimates to compute the matrices $\Omega_{R,t}$ and $\Omega_{I,t}$. We assume managers' annualized growth-adjusted discount rate is 1%, which implies that, on a monthly basis, $\beta \times (1 + \bar{r}) = 0.999$. For each period t, we then directly solve for implied costs as:

$$\mathbf{c}_t = \mathbf{p}_t - (M_{R,t}\hat{\Omega}_{R,t} + M_{I,t}\hat{\Omega}_{I,t})^{-1}(M_{R,t}\mathbf{s}_{R,t} + M_{I,t}\mathbf{s}_{I,t}).$$

We report the estimated distribution of marginal costs and markups for mutual funds and ETFs in Appendix Figure A3.[23] To account for outliers, we winsorize marginal costs at the 5% level.[24] The mean (median) marginal cost is 35 (23) basis points, and the mean (median) markup is 53 (32) basis points. Mutual funds and ETFs have similar markup distributions. The marginal costs of mutual funds are, on average, 47 basis points higher than those of ETFs and are also more dispersed. Notably, many mutual funds have similar costs to ETFs, but their distribution has a heavier right tail. These differences in marginal costs contribute to the impact of ETF introduction, which we

[23] For computational ease, we restrict our attention to those index funds with an active market share such that $s_{R,j,t} + s_{I,j,t} \geq 1e-6$.

[24] Our estimates imply that some funds have negative marginal costs. One explanation for this is that mutual funds generate revenue by lending the shares that they own for a fee, which offsets the costs of running the fund. For example, State Street estimates that securities lending increases the yield of SPY by 7.5 bps per year (https://www.ssga.com/us/en/individual/etfs/insights/unlocking-the-securities-lending-potential-of-spy). The large fund families return these fees to investors: see for example https://www.vanguard.ca/documents/securities-lending-considerations.pdf.

explore in Section 5.[25]

4.3 Robustness

We evaluate the robustness of our estimates by examining how the key results behave under alternative specifications.

Estimating Inertia: We explore an alternative method of estimating inertia that does not rely on the new sales data. Without new sales, identification must rely on the serial correlation in AUM, but lagged AUM is endogenous due to persistent investor preferences. To address this, we use past monthly fund returns as an exogenous source of persistence in holdings. The idea is that if a fund experienced strong returns two months ago, for example, it mechanically increases lagged AUM. How these past return shocks carry over into current AUM reveals the degree of investor inertia.

A potential concern is that if investors chase returns, then past monthly returns could impact current demand for a fund. To account for this, we control for 1, 3, 6, and 12-month cumulative returns and year-to-date returns, with the idea that investors chase returns according to these horizons since they are the horizons reported in fund marketing documents. We implement this by estimating variants of the following empirical analog of equation 4:

$$AUM_{T,j,t} = \phi_T AUM_{T,j,t-1}(1 + r_{j,t}) + X'_{j,t-1}\Gamma + \iota_{T,j,t}, \tag{14}$$

where observations are at the fund-by-month-by-investor type level. We instrument for lagged assets using the past twelve monthly returns while controlling for the reported returns described above. The resulting inertia estimates closely match our baseline values: 95.9% of retail investors and 97.3% of institutional investors are inert (vs. 95.0% and 97.2% in the baseline). Appendix Table A1 reports the full set of results.

Alternative Measures of Active AUM: We also consider an alternative way of constructing active AUM, which serves as an input for estimating preferences and information frictions. Instead of integrating over the distribution of $\phi_{T,m,t}$ in market m that implies positive active AUM, we use the estimated market-by-time level inertia $\phi_{T,m,t}$

[25] We decompose the estimated marginal costs by regressing them on fund fixed effects and year-month fixed effects. We find that much of the variation (39%) is unexplained by these fixed effects.

directly and drop observations that produce non-positive active AUM. As shown in Appendix Table A2, our demand estimates remain similar.

Alternative Instruments: We next show that our demand estimates are robust to including cost shifters as instruments for expense ratios. In particular, we include asset-weighted average trading cost as measured by the bid-ask spreads of securities held by each index fund as an additional instrument. We also construct an additional cost shifter by proxying fund managers' markup with the sum of adviser fees and distribution fees, and then using the difference between expense ratio and the markup proxy as an instrument, following Janssen and Thiel (2024). Appendix Table A3 and A4 summarize the demand estimates obtained with these instruments.

Preference Heterogeneity Across 401(k) and Non-401(k) Investors: Lastly, we estimate preference parameters for the 401(k) and main samples separately, allowing preferences to differ across the two samples. As shown in Appendix Table A5, the estimates are similar to our baseline estimates: the implied elasticity of demand is 1.6 for retail investors and 5 in the 401(k) setting (vs. 1.6 and 3.3 in the baseline). The implied preference heterogeneity here is 0.10 ($= 1/9.956$), which is lower than our baseline estimate (0.16); and the implied information friction here is 0.31 ($= \sqrt{\frac{1}{3.094^2} - \frac{1}{9.956^2}}$), which is higher than our baseline estimate (0.28). This is because allowing utility parameters in the 401(k) sample to be different from the retail sample explains part of the unobserved preference heterogeneity. Nevertheless, given the small differences in estimates, we maintain the baseline specification for ease of interpretation.

5 Counterfactuals

We conduct counterfactual simulations based on our model estimates with three main goals. First, we assess how frictions faced by investors, namely inertia and information frictions, impact market outcomes. Understanding these frictions individually and how they interact with each other can inform policy design. Second, we examine the effects of price discrimination and how it interacts with these frictions. Third, we examine how the introduction of ETFs impacted the index fund market. Given the dramatic increase in the number of ETFs since their first introduction in the 1990s and the cost of advantage of ETFs, the persistent price dispersion in the market remains a puzzle.

For each counterfactual, we first consider a partial-equilibrium analysis in which we keep fund expense ratios fixed. We then consider a general equilibrium analysis where we allow fund managers to optimally update their expense ratios, and we solve for a new equilibrium. Separating the demand- and supply-side response is useful for understanding the full implications of each friction. Appendix D provides full implementation details.

5.1 Quantifying Frictions

To quantify the impact of frictions, we simulate counterfactual distributions of expense ratios when these frictions are eliminated individually and simultaneously. Table 3 summarizes results and Figure 8 shows the full distribution of expense ratios under each counterfactual scenario.

First, we consider eliminating inertia by setting the share of inactive investors to zero for both types: $\phi_R = \phi_I = 0$. Given that we find only about 5% of retail investors update their portfolio in a median market-month, one might expect this to have a large effect on expense ratios.

In the partial equilibrium where expense ratios are fixed, removing inertia reduces average expense ratios from 27 basis points to 22 basis points for retail investors as shown in the first column of Table 3. The fact that eliminating inertia only reduces expense ratios by 5 basis points is surprising. As discussed further below, this modest effect is in part because retail investors are not very good at selecting optimal funds in the first place. In other words, due to large information frictions, allowing them to select funds more frequently (i.e., removing inertia) has limited effect on their actual fund choices.

When we allow for supply-side response, eliminating inertia reduces average expense ratio by roughly 6 basis points, from 27 to 21 basis points. This is shown in the third column of Table 3. Recall from Section 3, that if index fund managers use a growth-adjusted discount factor of 1 (i.e., $\beta(1 + \bar{r}_{m(j)}) = 1$), inertia would have no impact on the price setting behavior of managers. However, given a growth-adjusted annual discount factor of 0.99, the optimal price index fund managers charge is increasing in investor inertia.

Removing inertia not only reduces the average price, but also the price dispersion. The second and fourth columns of Table 3 show that eliminating inertia reduces price dispersion, as measured by the standard deviation of expense ratios, by 12 basis points from 48 basis points to 36 basis points.

Table 3: Quantifying Frictions: Mean and Standard Deviation of Expense Ratios

Panel A: Retail Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.27	0.48		
Counterfactuals	Without Supply Response		With Supply Response	
No Inertia	0.22	0.36	0.21	0.36
No Info Frictions	0.15	0.21	0.13	0.41
No Inertia or Info Frictions	0.12	0.15	0.05	0.14
No Px Discrimination			0.19	0.38
No Inertia, Info Frictions, or Px Discrimination			0.05	0.16

Panel B: Institutional Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.20	0.25		
Counterfactuals	Without Supply Response		With Supply Response	
No Inertia	0.17	0.22	0.16	0.22
No Info Frictions	0.16	0.20	0.13	0.31
No Inertia or Info Frictions	0.13	0.17	0.07	0.15
No Px Discrimination			0.16	0.25
No Inertia, Info Frictions, or Px Discrimination			0.07	0.16

Note: Table 3 displays the mean and standard deviation of asset-weighted expense ratios investors pay in each counterfactual. Expense ratios are expressed in decimals; 0.27 corresponds to 0.27% (27 basis points).

Overall, our results imply that removing investor inertia lowers the average expense ratio by 22% and price dispersion by 25%, accounting for the supply-side response.

Second, we consider the counterfactual where we eliminate information frictions. We implement this counterfactual by setting α_T, investor price sensitivity, to the value we recover from our estimates using the 401(k) data for both retail and institutional investors (Table 2).[26] Eliminating information frictions reduces the average expense ratio retail investors pay by 44% from 27 basis points to 15 basis points in the partial equilibrium; and further to 13 basis points when accounting for the supply-side re-

[26] Alternatively, we can eliminate information frictions by re-scaling the unobserved component of the utility such that its variance is equal to our estimates from the 401(k) setting, i.e., $\sigma_{\nu,T} = 0$, and we get qualitatively similar results. To the extent that information frictions are still present in the 401(k) setting, this counterfactual can be interpreted as the effect of making the index fund market as transparent as the 401(k) setting. In this case, counterfactual estimates would be a lower bound of the effect of completely removing information frictions.

sponse. Thus, removing information frictions has a larger effect than removing inertia despite the high level of inertia.[27]

Third, we consider the counterfactual distributions of expenses that investors would pay if both inertia and information frictions are simultaneously eliminated. Here, eliminating inertia has a much larger effect on the expense ratios retail investors pay after information frictions are eliminated. For example, in partial equilibrium, if information frictions remain, eliminating inertia reduces the average expense ratio retail investors only modestly from 27 to 22 basis points. In contrast, if we eliminate inertia and investors face no information frictions, the average expense ratio retail investors pay falls to 12 basis points. This result is intuitive: removing inertia and allowing investors to shop for index funds more frequently is more valuable when investors are better informed and can identify low-cost funds more effectively. This is also reflected in the standard deviation of prices, which falls from 48 basis points to 15 basis points. Furthermore, if we allow firms to adjust their prices, then eliminating both inertia and information frictions lowers the expense ratios paid by retail investors to 5 basis points. Collectively, these results suggest that the high level of inertia may not be as costly as it appears at first glance, because investors also face relatively severe information frictions.[28]

While we focus on retail investors, we also report the corresponding findings for institutional investors in Figure A4 and Panel B of Table 3. We find qualitatively similar results: eliminating inertia would lower the average expense ratios institutional investors pay by 20%, from 20 basis points to 16 basis points; eliminating information frictions would lower their average ratios by 35%, to 13 basis points; eliminating both would lower their average expense ratios by 65%, to 7 basis points. Quantitatively, these effects are somewhat smaller than those we find for retail investors. The smaller effect of removing information frictions is expected, given that institutional investors face less such frictions. More surprisingly, eliminating inertia also has a smaller impact on institutions, despite their marginally higher inertia. This likely reflects the fact

[27]We also experiment with reducing information frictions by half, e.g., setting the price coefficient to the average of the estimates from the baseline setting and the 401(k) setting (see Figure A8 and Table A7). Compared to eliminating information frictions entirely, reducing information frictions by half leads to three-quarters of the decrease in expense ratios in the partial equilibrium, and nearly all of the effect in the general equilibrium.

[28]For illustrative purposes, we separately change information frictions and inertia in our decomposition; however, our estimates show that the two frictions are correlated. It would be relatively straightforward to extend the model to endogenize the level of inertia. Estimating such a model would require richer variation in information frictions, however.

that institutional investors face narrower price dispersions in the baseline, so there are fewer expensive funds for them to get attached to in the first place.

5.2 Price Discrimination

The frictions we document have important implications for the supply side and price discrimination. Retail investors face larger information frictions but have slightly lower inertia. Given that information frictions are dominant, there are larger markups for retail investors. As documented in Section 2.2, managers' ability to charge higher prices to its inelastic retail investors and lower prices to its elastic institutional investors contributes meaningfully to the dispersion in fund expense ratios.

To study the role of price discrimination and how it interacts with demand-side frictions, we consider a counterfactual policy in which price discrimination is banned in the index fund market. We require index fund managers to charge the same expense ratio for all funds that share the same underlying portfolio (as defined by the CRSP portfolio identifier), and then solving for a new equilibrium.[29] In this environment, institutional and retail investors buying the same portfolio pay the same fee.

As reported in Table 3, eliminating price discrimination would lower the expenses retail investors pay by 30% (8 basis points) (see Panels (d) of Figures 8 for the full expense-ratio distributions). Because institutional investors are more discerning and price-sensitive, the market for institutional funds is more competitive. Retail investors therefore benefit from pooling with institutional investors and purchasing the same funds. This finding is also consistent with our motivating evidence that fund managers charge higher prices to retail investors than institutional investors for most portfolios (Figure 5). The final row of Table 3 shows that price discrimination would have virtually no effect in the absence of information frictions and inertia. In other words, if investors are well informed and never inert, there is no scope for price discrimination.

In Appendix E, we explore other supply-side frictions and extend the model to study the role of financial advisers and the potential conflicts of interest that arise from their involvement. We find modest conflicts of interest, which is consistent with the relatively high transparency, and thus less scope for agency issues in the index fund market compared to other settings analyzed in the literature.

[29]We implement this counterfactual by also assuming that index fund managers operate funds with the same portfolio at the minimum marginal cost in this portfolio. This is equivalent to requiring fund managers to always offer the lowest-cost fund for any portfolio, all else equal.

Figure 8: Counterfactuals: Retail Investors

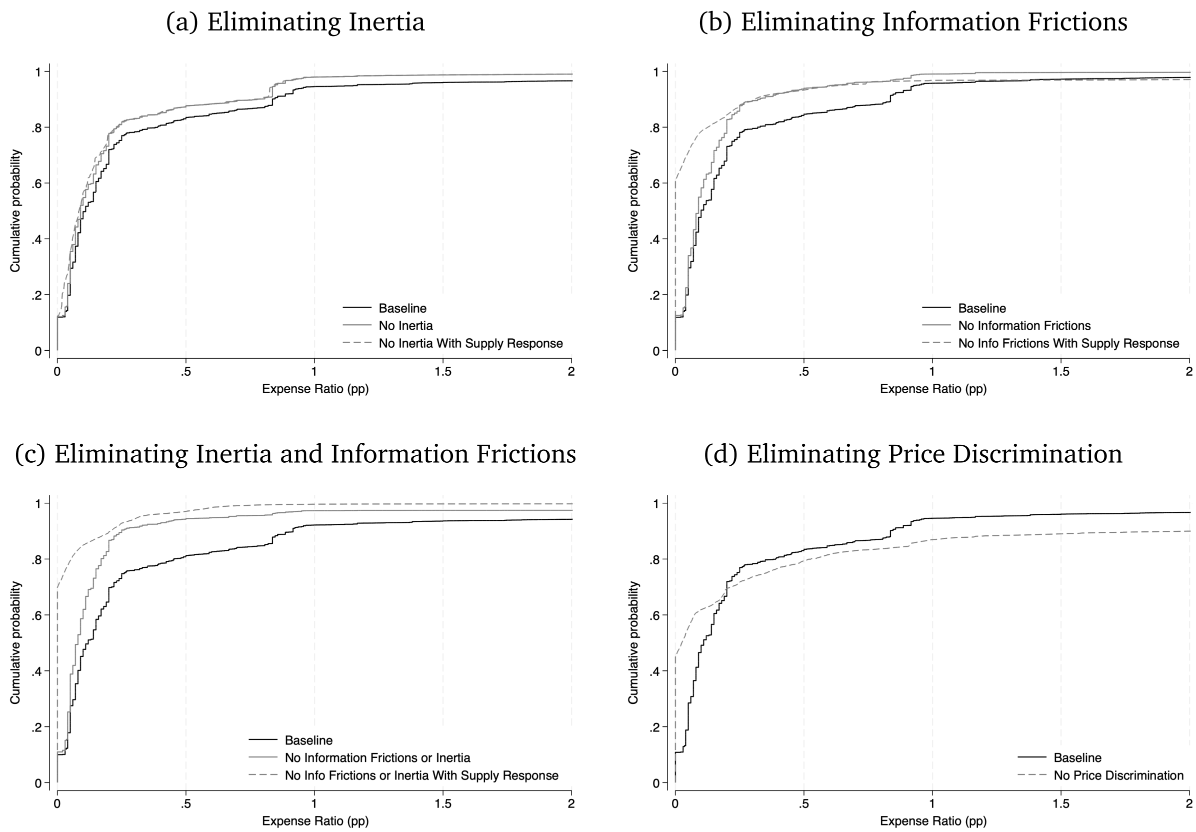


Figure 8 displays the estimated distribution of asset-weighted expense ratios in the baseline and counterfactual simulations where we eliminate inertia, information frictions and price discrimination.

5.3 The Introduction of ETFs

Lastly, we examine how the introduction of ETFs has impacted market competition and how frictions faced by investors shape this effect. ETFs are distinct from mutual funds along two key dimensions. First, ETFs generally have lower marginal costs. Second, ETFs are inherently available to both institutional and retail investors, which precludes prices discrimination.

We first simulate the counterfactual distributions of expenses that retail and institutional investors would pay on mutual funds if ETFs had not entered the market. We plot the distributions in Figure A7 and report means and standard deviations in Table 4. The results imply that retail index mutual fund expense ratios decrease by 39% (from 44 to 27 basis points) with the introduction of ETFs. Part of the effect is driven by the lower cost of ETFs and the other part comes from the competitive effect of ETFs.

To separate these channels, we consider a counterfactual in which ETFs are present but are forced to have the same (asset-weighted) average marginal cost as mutual funds in the same Lipper class and month. In that case, the average retail expense ratio would be 10 basis points lower than in the baseline case without ETFs. This suggests that the competitive effect of ETFs accounts for 59% of the overall effect for retail funds, while the cost effect accounts for the rest.

Table 4: The Introduction of ETFs: Mean and Standard Deviation of Expense Ratios

Panel A: Retail Investors

	Mean	Std. Dev.	Mean	Std. Dev.
	Without ETF		With ETF	
Baseline	0.44	0.58	0.27	0.48
No Inertia	0.36	0.49	0.21	0.36
No Info Frictions	0.17	0.39	0.13	0.41
No Inertia or Info Frictions	0.11	0.22	0.05	0.14
No ETF Cost Advantage			0.34	0.52

Panel B: Institutional Investors

	Mean	Std. Dev.	Mean	Std. Dev.
	Without ETF		With ETF	
Baseline	0.28	0.31	0.20	0.25
No Inertia	0.24	0.29	0.16	0.22
No Info Frictions	0.19	0.31	0.13	0.31
No Inertia or Info Frictions	0.15	0.22	0.07	0.15
No ETF Cost Advantage			0.24	0.31

Note: Table 4 displays the mean and standard deviation of asset-weighted mutual fund expense ratios investors pay in each counterfactual.

Many expected the introduction of ETFs to have larger competitive effects. One explanation is that frictions faced by investors allowed fund managers to maintain market power. To examine whether frictions dampened the effect of ETF entry, we simulate ETF introduction while eliminating frictions, as reported in Table 4. Introducing ETFs while also removing inertia decreases retail mutual fund expense ratios by 52% (from 44 to 21 basis points). Introducing ETFs while eliminating both inertia and information frictions magnifies the effect, as retail expense ratios decline to 5 basis points. The results highlight how frictions dampen the competitive effects of new products.

6 Conclusion

We quantify the key frictions in the index fund market and show how they support an equilibrium with substantial market power, even with the entry of low cost funds. We develop a tractable model featuring investor inertia, information frictions, and heterogeneous preferences, with forward-looking fund managers setting prices in response to these frictions. Using fund-level choices, new-sales data, and 401(k) menus, we separately identify inertia, information frictions, and investor preferences.

Our estimates imply that both inertia and information frictions give firms significant market power, but information frictions are the dominant force behind high and dispersed fees. Even though most investors are inactive in a given month, eliminating information frictions alone reduces expense ratios much more than eliminating inertia. This is because inertia matters most when investors are well informed and better able to identify low-fee funds. These results suggest that policy should prioritize reducing information frictions. Disclosure policies, rule-making that reduces misleading practices, or more effective comparison tools such as FINRA's Fund Analyzer, could lead to a meaningful reduction in market power and increase welfare.

We also show that mutual fund providers take advantage of these frictions by price discriminating between institutional and retail investors. In the presence of information frictions and inertia, price discrimination is quite costly for retail investors; however, its effect is negligible without these other frictions.

Finally, ETF entry affected the market by introducing a lower-cost technology and intensifying competition with similar mutual funds. However, demand-side frictions substantially muted these effects. Without inertia and information frictions, ETF entry would have generated larger reductions in equilibrium fees.

Taken together, our results help explain why expensive index funds maintain some market share despite the availability of cheap alternatives and why new entry has not eliminated fee dispersion. Because many financial markets share similar features—complex products, limited attention, and segmented demand—these mechanisms are likely to apply more broadly and illustrate the importance of understanding how distinct frictions interact with firms' pricing incentives when diagnosing the sources of market power.

References

Allen, J., R. Clark, and J.-F. Houde (2014). The effect of mergers in search markets: Evidence from the canadian mortgage industry. *American Economic Review 104*(10), 3365–96.

Allen, J. and S. Li (2020). Dynamic competition in negotiated price markets. *Working Paper*.

An, Y., M. Benetton, and Y. Song (2021). Index providers: Whales behind the scenes of etfs. *Working Paper*.

Andersen, S., J. Y. Campbell, K. M. Nielsen, and T. Ramadorai (2020). Sources of inaction in household finance: Evidence from the danish mortgage market. *American Economic Review 110*(10), 3184–3230.

Baker, M., M. L. Egan, and S. K. Sarkar (2022). How do investors value esg? *Working Paper*.

Beggs, A. and P. Klemperer (1992). Multi-period competition with switching costs. *Econometrica*, 651–666.

Benartzi, S. and R. H. Thaler (1995). Myopic loss aversion and the equity premium puzzle. *The quarterly journal of Economics 110*(1), 73–92.

Benetton, M. (2021). Leverage regulation and market structure: A structural model of the uk mortgage market. *The Journal of Finance 76*(6), 2997–3053.

Benetton, M. and G. Compiani (2021). Investors beliefs and cryptocurrency prices.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica*, 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Bhattacharya, V. and G. Illanes (2022). The design of defined contribution plans. *Working Paper*.

Bhattacharya, V., G. Illanes, and M. Padi (2019). Fiduciary duty and the market for financial advice. Technical report, National Bureau of Economic Research.

Bretscher, L., L. Schmid, I. Sen, and V. Sharma (2020). Swiss finance institute working paper. *Working Paper*.

Brown, Z. Y. and J. Jeon (2024). Endogenous information and simplifying insurance choice. *Econometrica 92*(3), 881–911.

Campbell, J. Y. (2016). Restoring rational choice: The challenge of consumer financial regulation. *American Economic Review 106*(5), 1–30.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory 13*(2), 185–213.

Choi, J. J., D. Laibson, and B. C. Madrian (2010). Why does the law of one price fail? an experiment on index mutual funds. *The Review of Financial Studies 23*(4), 1405–1432.

Christoffersen, S. E., R. Evans, and D. K. Musto (2013). What do consumers fund flows maximize? evidence from their brokers incentives. *The Journal of Finance 68*(1), 201–235.

Clark, R., J.-F. Houde, and J. Kastl (2021). The industrial organization of financial markets. In *Handbook of Industrial Organization*, Volume 5, pp. 427–520. Elsevier.

Dick, A. A. (2008). Demand estimation and consumer welfare in the banking industry. *Journal of Banking & Finance 32*(8), 1661–1676.

Egan, M. (2019). Brokers versus retail investors: Conflicting interests and dominated products. *The Journal of Finance 74*(3), 1217–1260.

Egan, M., S. Ge, and J. Tang (2022). Conflicting interests and the effect of fiduciary duty: Evidence from variable annuities. *The Review of Financial Studies 35*(12), 5334–5386.

Egan, M., A. Hortaçsu, and G. Matvos (2017). Deposit competition and financial fragility: Evidence from the us banking sector. *American Economic Review 107*(1), 169–216.

Egan, M. L., A. MacKay, and H. Yang (2021). What drives variation in investor portfolios? estimating the roles of beliefs and risk preferences. *Working paper*.

Egan, M. L., A. MacKay, and H. Yang (2022). Recovering investor expectations from demand for index funds. *Review of Economic Studies*.

Fosgerau, M., E. Melo, A. De Palma, and M. Shum (2020). Discrete choice and rational inattention: A general equivalence result. *International economic review 61*(4), 1569–1589.

Gabaix, X., R. S. Koijen, F. Mainardi, S. Oh, and M. Yogo (2024). Asset demand of us households. *Working Paper*.

Gavazza, A. (2011). Demand spillovers and market outcomes in the mutual fund industry. *The RAND Journal of Economics 42*(4), 776–804.

Grubb, M. D. (2015). Failing to choose the best price: Theory, evidence, and policy. *Review of Industrial Organization 47*, 303–340.

Haddad, V., P. Huebner, and E. Loualiche (2021). How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing. *Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing (April 7, 2021)*.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review 103*(7), 2643–82.

Hastings, J., A. Hortaçsu, and C. Syverson (2017). Sales force and competition in financial product markets: the case of mexico's social security privatization. *Econometrica 85*(6), 1723–1761.

Hausman, J. A. (1996). Volume title: The economics of new goods volume author/editor: Timothy f. bresnahan and robert j. gordon, editors volume publisher: University of chicago press volume isbn: 0-226-07415-3 volume url: http://www. nber. org/books/bres96-1.

Ho, K., J. Hogan, and F. Scott Morton (2017). The impact of consumer inattention on insurer pricing in the medicare part d program. *The RAND Journal of Economics 48*(4), 877–905.

Honka, E., A. Hortaçsu, and M. A. Vitorino (2017). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *The RAND Journal of Economics 48*(3), 611–646.

Hortaçsu, A., S. A. Madanizadeh, and S. L. Puller (2017). Power to choose? an analysis of consumer inertia in the residential electricity market. *American Economic Journal: Economic Policy 9*(4), 192–226.

Hortaçsu, A. and C. Syverson (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *The Quarterly Journal of Economics 119*(2), 403–456.

Illanes, G. (2016). Switching costs in pension plan choice. *Unpublished manuscript*.

Janssen, A. and J. Thiel (2024). Do search costs explain persistent investment in active mutual funds? *Available at SSRN 4759775*.

Jiang, G. J., T. Yao, and G. R. Zaynutdinova (2023). The effect of investor service costs on mutual fund performance. *Financial Review 58*(1), 91–115.

Jin, C., S. Peng, and P. Wang (2022). Sticky consumers and cloud welfare. *Working paper*.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A rational theory of mutual funds' attention allocation. *Econometrica 84*(2), 571–626.

Kiser, E. K. (2002). Predicting household switching behavior and switching costs at depository institutions. *Review of Industrial Organization 20*(4), 349–365.

Koijen, R. S. and M. Yogo (2016). Shadow insurance. *Econometrica 84*(3), 1265–1287.

Koijen, R. S. and M. Yogo (2019a). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

Koijen, R. S. and M. Yogo (2019b). Exchange rates and asset prices in a global demand system. Working paper, SSRN 3383677.

Koijen, R. S. and M. Yogo (2022). The fragility of market risk insurance. *The Journal of Finance 77*(2), 815–862.

Kronlund, M., V. K. Pool, C. Sialm, and I. Stefanescu (2021). Out of sight no more? the effect of fee disclosures on 401 (k) investment allocations. *Journal of Financial Economics 141*(2), 644–668.

Luco, F. (2019). Switching costs and competition in retirement investment. *American economic journal: Microeconomics 11*(2), 26–54.

MacKay, A. and M. Remer (2022). Consumer inertia and market power. *Available at SSRN 3380390*.

Madrian, B. C. and D. F. Shea (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly journal of economics 116*(4), 1149–1187.

Matějka, F. and A. McKay (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review 105*(1), 272–298.

Robles-Garcia, C. (2019). Competition and incentives in mortgage markets: The role of brokers. *Unpublished working paper*.

Roussanov, N., H. Ruan, and Y. Wei (2021). Marketing mutual funds. *The Review of Financial Studies 34*(6), 3045–3094.

Steiner, J., C. Stewart, and F. Matějka (2017). Rational inattention dynamics: Inertia and delay in decision-making. *Econometrica 85*(2), 521–553.

Woodward, S. E. and R. E. Hall (2012). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *American Economic Review 102*(7), 3249–3276.

Xiao, K. (2020). Monetary transmission through shadow banks. *The Review of Financial Studies 33*(6), 2379–2420.

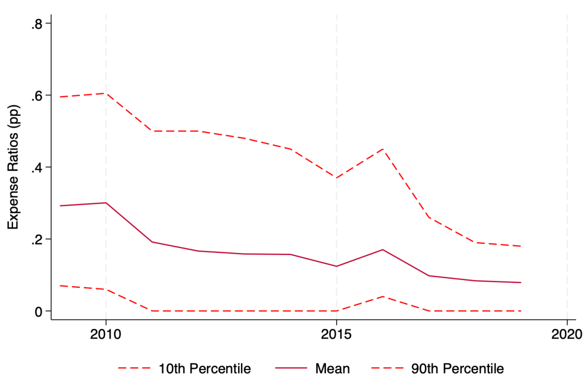
Zhang, D. (2022). Closing costs, refinancing, and inefficiencies in the mortgage market. *Working Paper*.

Appendix

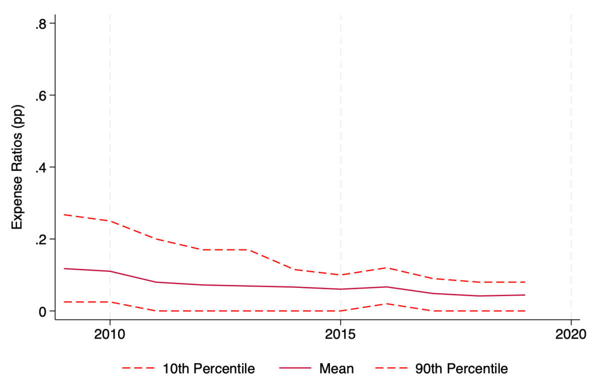
A Additional Figures and Tables

Figure A1: Distribution of Expense Ratios for 401(k) Plans over Time

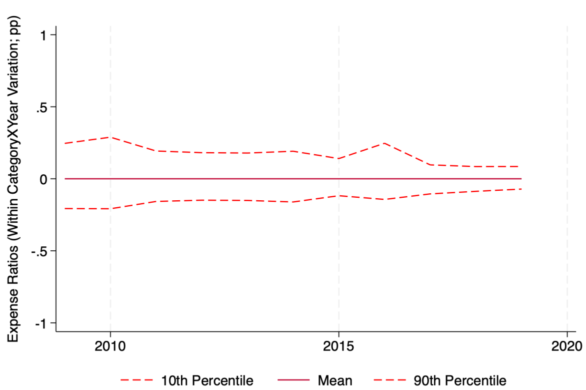
(a) Expense Ratios



(b) Expense Ratios (Weighted)



(c) Residualized Expense Ratios



(d) Residualized Expense Ratios (Weighted)

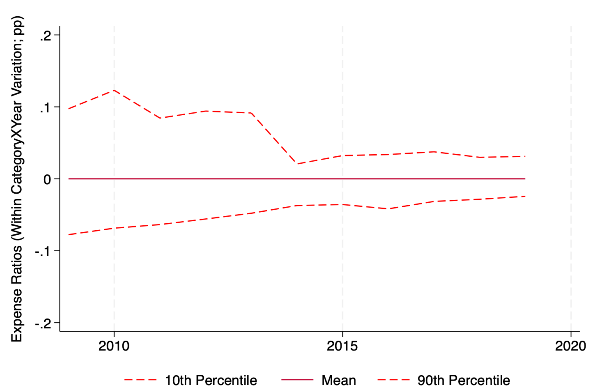


Figure A1 displays the distribution of index fund expense ratios in 401(k) plans over time. Panels (a) and (b) display the equal weighted and asset-weighted distribution of expense ratios. Panels (c) and (d) display the equal weighted and asset-weighted distribution of residualized expense ratios, where we residualize expense ratios by regressing them on Category $\times$ Year fixed effects. Panels (c) and (d) therefore display the within Category $\times$ Year variation in expense ratios.

Table A1: Robustness of Investor Inertia Using Persistence of Exogenous Demand Shocks

(a) Retail Investors

VARIABLES	(1)	(2)	(3)
Lag AUM	0.971***	0.960***	0.959***
	(0.003)	(0.004)	(0.004)
Observations	371,660	357,129	355,111
R-squared	0.996	0.996	0.996
IV		X	X
Year-Month FE	X	X	
Year-Month-Mkt FE			X

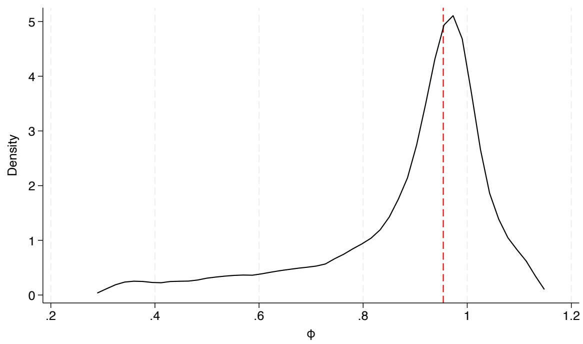
(b) Institutional Investors

VARIABLES	(1)	(2)	(3)
Lag AUM	0.986***	0.973***	0.973***
	(0.002)	(0.003)	(0.003)
Observations	324,099	313,415	311,598
R-squared	0.996	0.996	0.996
IV		X	X
Year-Month FE	X	X	
Year-Month-Mkt FE			X

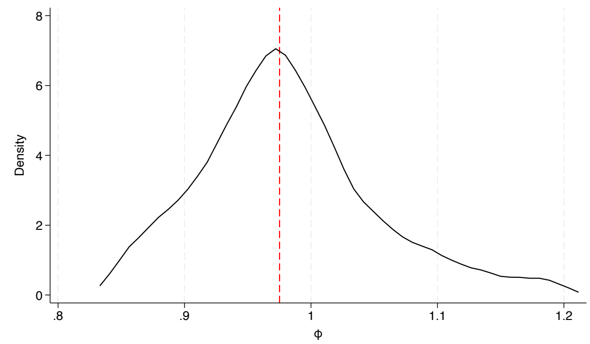
Note: Table A1 displays the estimates corresponding to a linear regression model (Eq. 14). Observations are at the index fund-by-month level. The dependent variable is assets under management. The independent variable of interest is $AUM_{j,T,t-1}(1+r_{j,t})$, where $r_{j,t}$ reflects the monthly return of the fund. In Panel (a) we restrict our attention to retail investors/AUM and in Panel (b) we restrict our attention to institutional investors/AUM. We address the endogeneity of *Lag AUM* using an instrumental variables approach in columns (2)-(3) using the past 12 monthly dollar returns of the fund. In all specifications we control for the log number of funds offered by the management company, the standard deviation of daily fund returns over the past 12 months, and whether the fund is an ETF, has a front load, or has a rear load. In columns (1)-(2) we control for 1-, 3-, 6-, 12-month, and year-to-date cumulative returns. In columns (3), where we include year-by-month-by-market fixed effects, we control for 1-month and year-to-date cumulative returns because the year-by-month-by-market fixed effects capture much of the variation in returns. Robust standard errors are in parenthesis. *** p<0.01, ** p<0.05, * p<0.10.

Figure A2: Robustness of Investor Inertia Under Alternative Assumptions About Market Growth

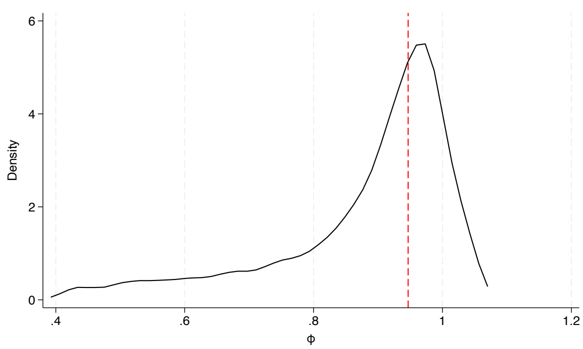
(a) Automatic Contributions (Retail)



(b) Automatic Contributions (Institutional)



(c) New Active Investors (Retail)



(d) New Active Investors (Institutional)

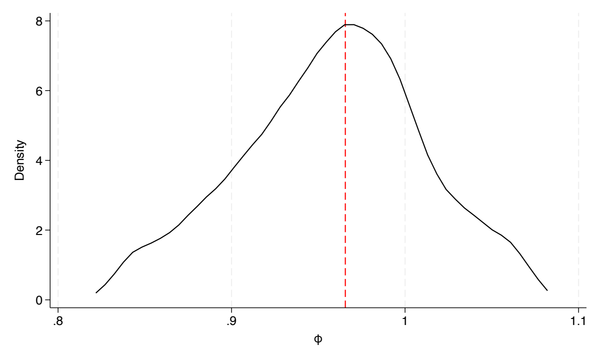


Figure A2 displays the distribution of inertia at the market-by-month level over the period 2000-2020 for retail and institutional investors under different assumptions about market growth. Figure A2a and Figure A2b assume that newly-added assets are investors' automatic contributions to their existing holdings. Figure A2d and Figure A2d assume that newly-added assets are from new active investors. To account for outliers, we censor the distribution at the AUM-weighted 5th and 95th percentiles. The red dashed line in each figure corresponds to the AUM-weighted median observation.

Table A2: Robustness of Investor Preferences Using Only Observations With Positive Active AUM

	Retail	Institutional	401k
Expense Ratio	-2.792^{***}	-4.505^{***}	-5.983^{***}
	(0.041)	(0.134)	(0.484)
Observations	76,799	56,530	32,777
Mkt-Time FE	X	X	X
IV	X	X	X
Elasticity of Demand	1.7	2.7	3.6

Note: Table A2 displays the demand estimates using only observations with positive revealed AUM from active investors. Observations are at the index fund-by-month-by-investor type (i.e., retail vs. institutional) level in columns (1) and (2). Observations are at the index fund-by-401(k) plan-by-market-by-time level in column (3). In all specifications we control for: the log number of funds offered by the management company; the standard deviation of daily fund returns over the past 12 months; 1-, 3-, 6-, and 12-month cumulative returns; and whether the fund is an ETF, has a front load, or has a rear load. Standard errors in parenthesis are clustered at portfolio-by-month level in column (1) and (2). GMM asymptotic standard errors are in parenthesis in column (1) and (3). *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A3: Robustness of Investor Preferences Including Spread Instrument

	Retail	Institutional	401k
Expense Ratio	-3.497^{***}	-5.393^{***}	-12.671^{***}
	(0.062)	(0.148)	(2.699)
Observations	37,885	51,232	15,111
Mkt-Time FE	X	X	X
IV	X	X	X
Elasticity of Demand	1.4	2.1	5

Note: Table A3 displays the demand estimates using asset-weighted average trading cost (bid-ask spreads) of the securities held by the fund as an instrument for expense ratios in addition to the standard Hausman instrument. Observations are at the index fund-by-month-by-investor type (i.e., retail vs. institutional) level. In all specifications we control for: the log number of funds offered by the management company; the standard deviation of daily fund returns over the past 12 months; 1-, 3-, 6-, and 12-month cumulative returns; and whether the fund is an ETF, has a front load, or has a rear load. Standard errors in parenthesis are clustered at portfolio-by-month level in column (1) and (2). GMM asymptotic standard errors are in parenthesis in column (1) and (3). *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A4: Robustness of Investor Preferences Using Markup Instrument

	Retail	Institutional	401k
Expense Ratio	-2.032^{***}	-1.368^{***}	-6.921
	(0.398)	(0.055)	(31.745)
Observations	191,016	166,369	21,563
Mkt-Time FE	X	X	X
IV	X	X	X
Elasticity of Demand	1	0.68	3.4

Note: Table A4 displays the demand estimates where we construct an instrument for expense ratios by proxying fund managers' markups. We proxy fund managers' markups using adviser fees and distribution fees, and then use the difference between expense ratios and the sum of adviser and distribution fees as an instrument. Observations are at the index fund-by-month-by-investor type (i.e., retail vs. institutional) level in columns (1) and (2). Observations are at the index fund-by-401(k) plan-by-market-by-time level in column (3). In all specifications we control for: the log number of funds offered by the management company; the standard deviation of daily fund returns over the past 12 months; 1-, 3-, 6-, and 12-month cumulative returns; and whether the fund is an ETF, has a front load, or has a rear load. Standard errors in parenthesis are clustered at portfolio-by-month level in column (1) and (2). GMM asymptotic standard errors are in parenthesis in column (1) and (3). *** p<0.01, ** p<0.05, * p<0.10.

Table A5: Robustness of Investor Preferences Estimating Retail and 401(k) Demand Separately

	Retail	Institutional	401k
Expense Ratio	-3.094^{***}	-4.929^{***}	-9.956^{***}
	(0.039)	(0.093)	(2.315)
Observations	105,499	109,758	32,777
Mkt-Time FE	X	X	X
IV	X	X	X
Elasticity of Demand	1.6	2.5	5

Note: Table A5 displays the demand estimates corresponding to an instrumental variable regression model (Eq. 12 and Eq. 13). Observations are at the index fund-by-month-by-investor type (i.e., retail vs. institutional) level in columns (1) and (2). Observations are at the index fund-by-401(k) plan-by-market-by-time level in column (3). In all specifications we control for: the log number of funds offered by the management company; the standard deviation of daily fund returns over the past 12 months; 1-, 3-, 6-, and 12-month cumulative returns; and whether the fund is an ETF, has a front load, or has a rear load. Standard errors in parenthesis are clustered at portfolio-by-month level in column (1) and (2). Robust standard errors are in parenthesis in column (3). *** p<0.01, ** p<0.05, * p<0.10.

Figure A3: Estimated Marginal Costs and Markups

(a) Distribution of Marginal Costs

(b) Distribution of Markups

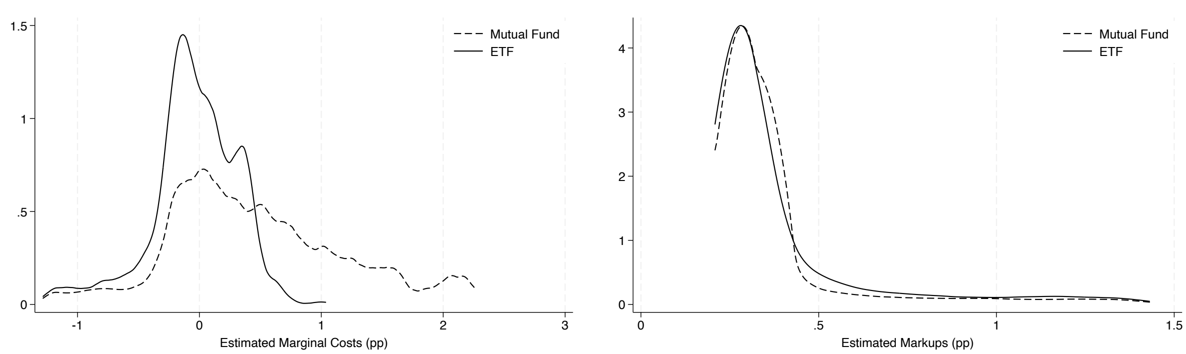
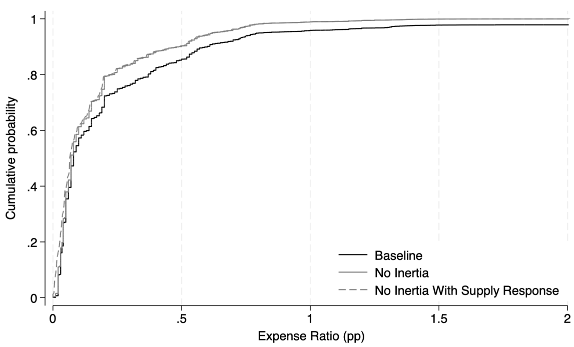


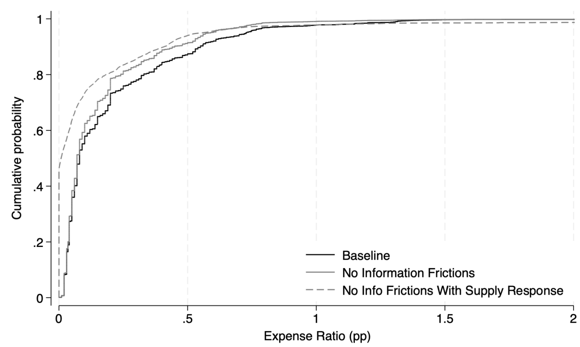
Figure A3 displays the estimated equal-weighted distributions of marginal costs and markups for mutual funds and ETFs. The solid line represents ETFs and the dashed line represents mutual funds. To account for outliers, both distributions are censored at the 5% and 95% level. Panel (a) displays the density of marginal costs, and panel (b) displays the density of markups.

Figure A4: Counterfactuals: Institutional Investors

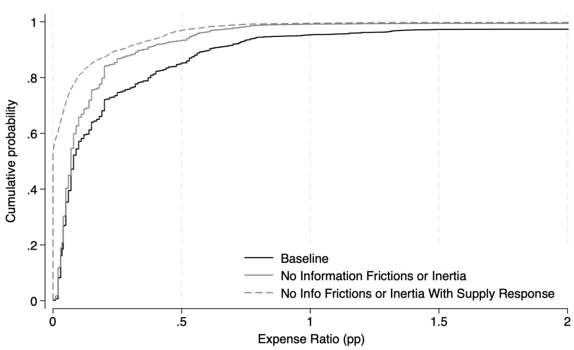
(a) Eliminating Inertia



(b) Eliminating Information Frictions



(c) Eliminating Inertia and Information Frictions



(d) Eliminating Price Discrimination

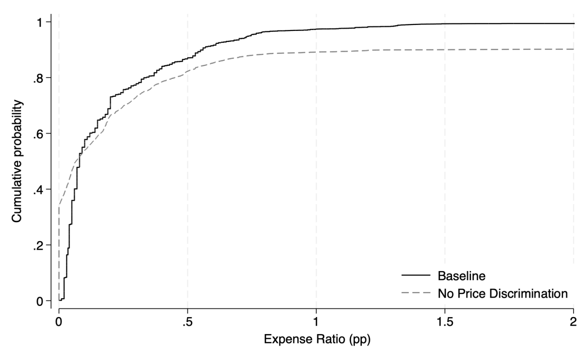


Figure A4 displays the estimated distribution of asset-weighted expense ratios in counterfactual analysis where we eliminate inertia, information frictions and price discrimination.

Figure A5: Counterfactuals: Remove All Frictions

(a) Retail Investors (b) Institutional Investors

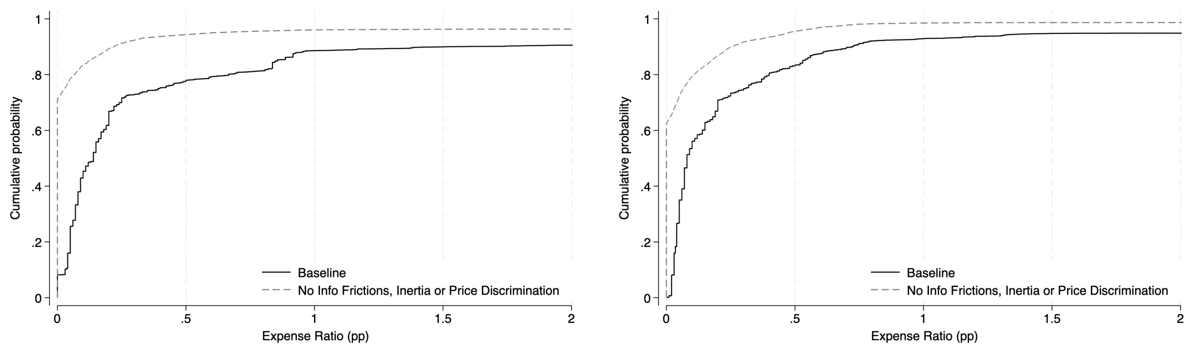


Figure A5 displays the estimated distribution of asset-weighted expense ratios in counterfactual analysis where we eliminate inertia, information frictions and price discrimination.

Figure A6: Sequential Decomposition by Mechanism

(a) Retail Investors (b) Institutional Investors

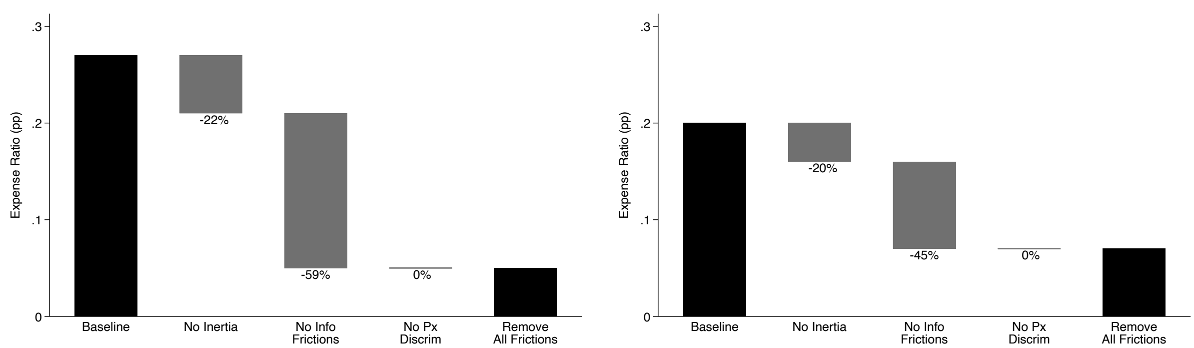
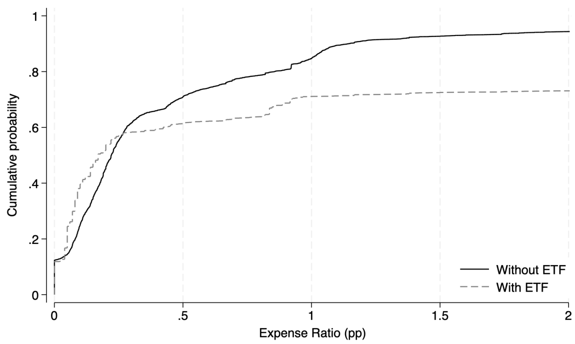


Figure A6 displays the mean asset-weighted expense ratios investors pay after sequentially removing inertia, information frictions, and price discrimination. All counterfactuals account for supply response.

Figure A7: Counterfactuals: The Introduction of ETFs

(a) Retail Investors



(b) Institutional Investors

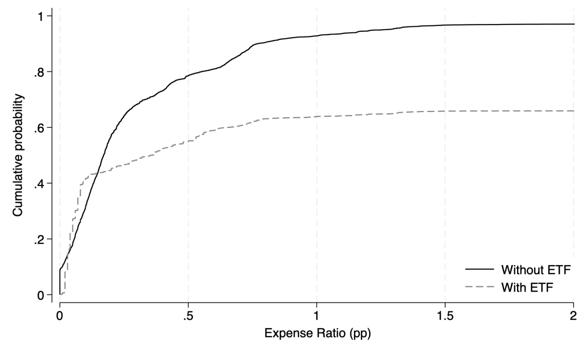
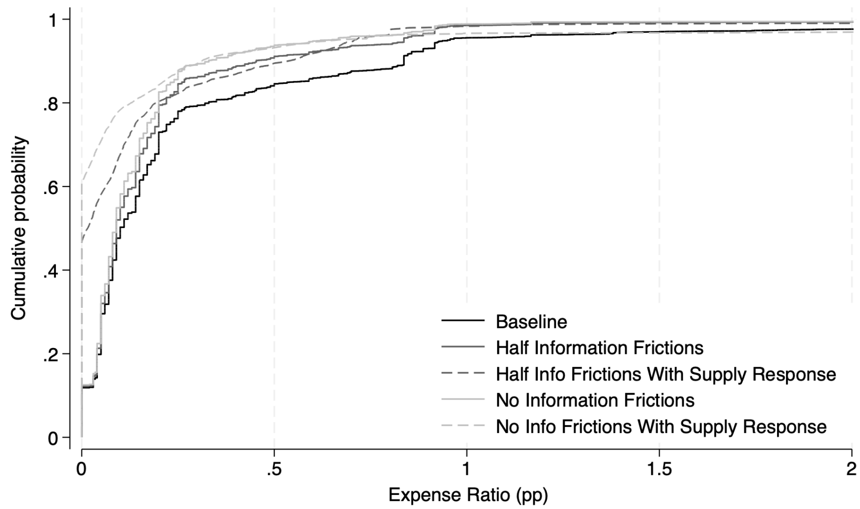


Figure 8 displays the estimated distribution of mutual fund expense ratios in counterfactual analysis where we eliminate ETFs.

Figure A8: Counterfactuals: Reducing Information Frictions

(a) Retail Investors



(b) Institutional Investors

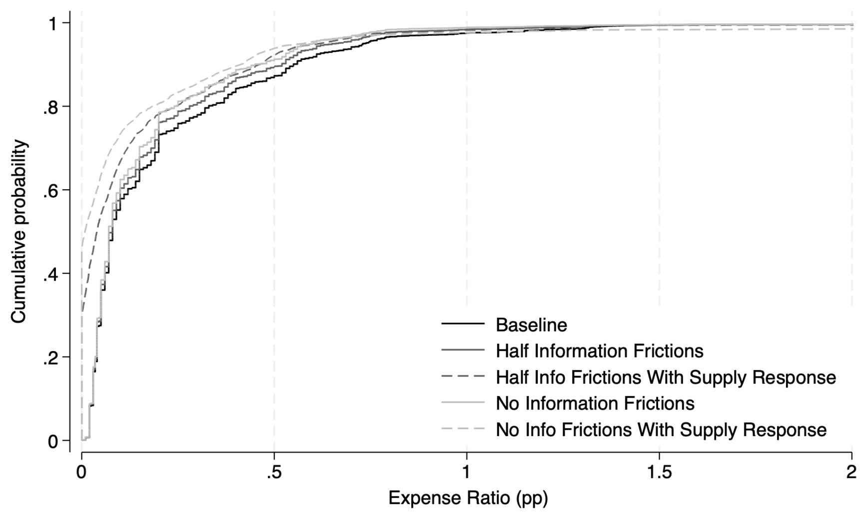
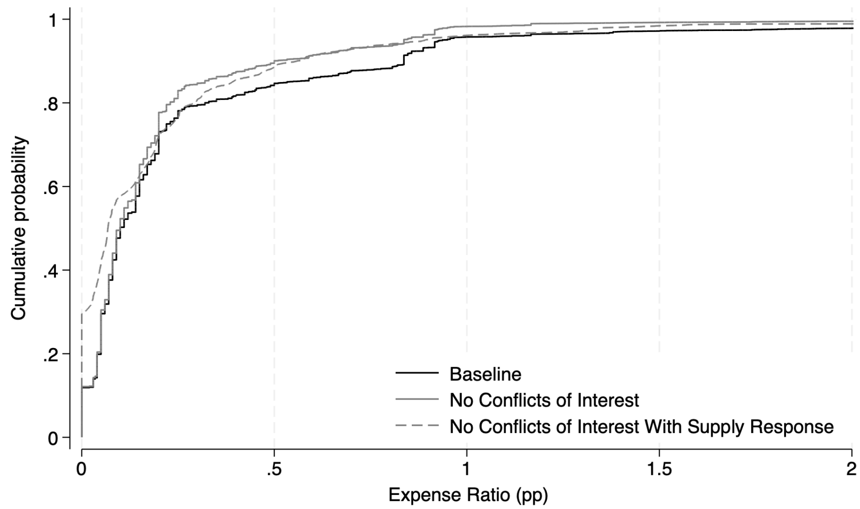


Figure A8 displays the estimated distribution of expense ratios in counterfactual analysis where we reduce information frictions first by half and then entirely.

Figure A9: Counterfactuals: Eliminating Conflicts of Interest

(a) Retail Investors



(b) Institutional Investors

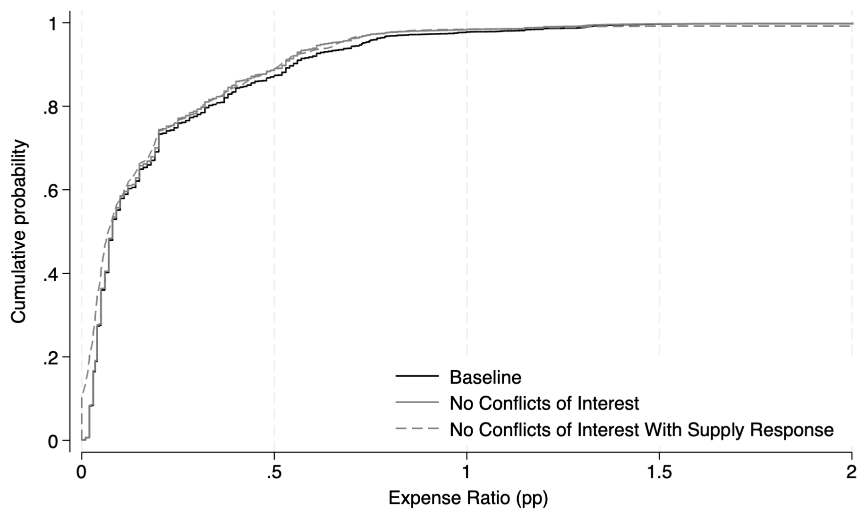


Figure A9 displays the estimated distribution of expense ratios in counterfactual analysis where we eliminate conflicts of interest.

Table A6: Active Demand for Index Funds: Accounting for Brokers

	(1)	(2)	(3)	(4)
Expense Ratio	-3.905^{***}	-3.593^{***}	-5.241^{***}	-5.167^{***}
	(0.064)	(0.067)	(0.101)	(0.104)
12b-1 Fee	2.394^{***}	1.792^{***}	3.477^{***}	2.368^{***}
	(0.075)	(0.087)	(0.162)	(0.197)
Observations	105,499	103,049	109,758	107,911
R-squared	0.493	0.508	0.363	0.367
Year-Month-Mkt FE	X	X	X	X
Exp Ratio IV	X	X	X	X
12b-1 IV		X		X
Retail Sample	X	X		
Inst. Sample			X	X
Elasticity of Demand	2.0	1.8	2.7	2.6
ω	0.38	0.33	0.40	0.31

Note: Table A6 displays the estimates corresponding to a linear regression model (Eq. 21). Observations are at the index fund-by-month-by-investor type (i.e., retail vs. institutional) level. In all specifications we control for: the log number of funds offered by the management company; the standard deviation of daily fund returns over the past 12 months; 1-, 3-, 6-, 12-month, and year-to-date cumulative returns; and whether the fund is an ETF, has a front load, or has a rear load. Robust standard errors are in parenthesis. *** p<0.01, ** p<0.05, * p<0.10.

Table A7: Reducing Information Frictions: Mean and Standard Deviation of Expense Ratios

Panel A: Retail Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.27	0.48		
Counterfactuals	Without Supply Response		With Supply Response	
Half Info Frictions	0.18	0.26	0.13	0.25
No Info Frictions	0.15	0.21	0.13	0.41

Panel B: Institutional Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.20	0.25		
Counterfactuals	Without Supply Response		With Supply Response	
Half Info Frictions	0.17	0.22	0.14	0.23
No Info Frictions	0.16	0.20	0.13	0.31

Note: Table A7 displays the mean and standard deviation of asset-weighted expense ratios investors pay in each counterfactual.

Table A8: Eliminating Conflicts of Interest: Mean and Standard Deviation of Expense Ratios

Panel A: Retail Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.27	0.48		
Counterfactuals	Without Supply Response		With Supply Response	
No Conflicts of Interest	0.19	0.28	0.18	0.28

Panel B: Institutional Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.20	0.25		
Counterfactuals	Without Supply Response		With Supply Response	
No Conflicts of Interest	0.18	0.23	0.17	0.22

Note: Table A8 displays the mean and standard deviation of asset-weighted expense ratios investors pay in each counterfactual.

B Information Frictions and Rational Inattention

This section highlights the connection between our model of information frictions and the rational inattention model. Starting from the general discrete choice problem with costly information acquisition presented in Matějka and McKay (2015), one can derive conditions under which our model is equivalent.

As in Section 3.1.1, consider investor i's utility from choosing fund j at time t

$$u_{i,j,t} = \bar{u}_{i,j,t} + \sigma_{\epsilon,T(i)}\epsilon_{i,j,t}$$

where $\bar{u}_{i,j,t}$ is the observable part of utility and $\sigma_{\epsilon,T(i)}\epsilon_{i,j}$ captures preference heterogeneity.

Following the model in Matějka and McKay (2015), individuals have a prior about the utility of each fund. There is unit cost of information $\lambda_{T(i)}$ which is the cost of reducing uncertainty, measured in units of entropy. The unit cost of information may vary by investor type, $T(i)$. Given this unit cost, individuals rationally choose how much to research each fund to maximize the expected payoff inclusive of the cost to acquire information. This leads to choice probabilities that take the form

$$p_{i,j,t} = \frac{p^0_{i,j,t}\exp((\bar{u}_{i,j,t} + \sigma_{\epsilon,T(i)}\epsilon_{i,j,t})/\lambda_{T(i)})}{\sum_{l\in\mathcal{J}_{T(i),m(j),t}} p^0_{i,l,t}\exp((\bar{u}_{i,l,t} + \sigma_{\epsilon,T(i)}\epsilon_{i,l,t})/\lambda_{T(i)})}$$

where $p^0_{i,j,t}$ is a function of the prior.

Given an assumption that the prior does not differ across options, Matějka and McKay (2015) show that choice probabilities simplify to

$$P_{i,j,t} = \frac{\exp((\bar{u}_{i,j,t} + \sigma_{\epsilon,T(i)}\epsilon_{i,j,t})/\lambda_{T(i)})}{\sum_{l\in\mathcal{J}_{T(i),m(j),t}} \exp((\bar{u}_{i,l,t} + \sigma_{\epsilon,T(i)}\epsilon_{i,l,t})/\lambda_{T(i)})}$$

Therefore, it is as if expected utility after information acquisition takes the form

$$\tilde{u}_{i,j,t} = \bar{u}_{i,j,t} + \sigma_{\epsilon,T(i)}\epsilon_{i,j,t} + \lambda_{T(i)}e_{i,j,t}$$

where $e_{i,j,t}$ is distributed Type 1 Extreme Value. This implies that $Var(\nu_{i,j,t})$, the degree of information frictions in Eq. (1), is proportional to the unit cost of information, $\lambda_{T(i)}$, in the rational inattention model presented above.[30]

[30] In the above model, $e_{i,j,t}$ is distributed Type 1 Extreme Value. For the purposes of estimation, we

C Market Growth

This section shows that our inertia estimates are robust to alternative adjustments for the growth of the index fund market. These adjustment rely on specific assumptions about how assets are newly added to the market.

At one extreme, one may assume that newly-added assets come from investors' automatic contributions to their existing holdings. Assume in time t, investors automatically contribute $g_{m,t}$ to their investment accounts in market m. For example, if they have \$100 invested in fund j and $g_{m,t} = 1.01$, then they automatically contribute \$1 to fund j. Then, we can redefine new sales and active demand as

$$\begin{aligned} NewSales_{T,j,t} &= s_{T,j,t}(1-\phi_{T,m(j),t})\sum_{l\neq j} AUM_{T,l,t-1}(1+r_{m(l),t-1}) \\ &\quad + AUM_{jt-1}(1+r_{m(j),t-1})(g_{m(j)t}-1) \qquad (15) \\ AUM^{Active}_{T,j,t} &= s_{T,j,t}(1-\phi_{m(j)t})\sum_{l\in\mathcal{J}_{T,m,t}} AUM_{T,l,t-1}(1+r_{m(l),t-1}) \\ &\quad + AUM_{T,j,t-1}(1+r_{m(j),t-1})(g_{m(j),t}-1). \qquad (16) \end{aligned}$$

At another extreme, one may assume that newly-added assets are entirely from new investors in the index fund market who make active choices, so we can redefine new sales and active demand as

$$\begin{aligned} NewSales_{T,j,t} &= s_{T,j,t}[(1-\phi_{T,m(j),t})\sum_{l\neq j} AUM_{T,l,t-1}(1+r_{m(l),t-1}) \\ &\quad + M_{T,m(j),t} - \sum_{l} AUM_{T,l,t-1}(1+r_{m(l),t-1})] \qquad (17) \\ AUM^{Active}_{T,j,t} &= s_{T,j,t}[(1-\phi_{T,m(j),t})\sum_{l} AUM_{T,l,t-1}(1+r_{m(l),t-1}) \\ &\quad + M_{T,m(j),t} - \sum_{l} AUM_{T,l,t-1}(1+r_{m(l),t-1})]. \qquad (18) \end{aligned}$$

The inertia estimates are robust to these assumptions. Figure A2a and Figure A2b present the inertia distributions estimated under the first assumption: newly-added assets are investors' automatic contributions to their existing holdings. We find that retail (institutional) investors are inactive 95.4% (97.5%) of the time in a median market-

assume that $\nu_{i,j,t}$ is distributed according to Cardell (1997). It should be noted that this assumption is also consistent with rational inattention since any random utility model can be rationalized by a generalized cost function for information (Fosgerau et al., 2020).

month. Figure A2d and Figure A2d present the inertia distribution estimated under the second assumption: newly-added assets are from new active investors. We find that retail (institutional) investors are inactive 94.7% (96.6%) of the time in a median market-month. Our baseline estimates fall between the estimates based on these two extreme assumptions.

D Implementing Counterfactuals

In our counterfactual analyses, we focus on how the distribution of expense ratios changes as a function of inertia, information frictions, and price discrimination. We compute the distribution of expense ratios in each counterfactual where we weight fund expense ratios by the predicted market share multiplied by market size. We compute the predicted market share of fund j at time t among type T investors as a function of inertia, expense ratios and information frictions:

$$\mathcal{S}_{T,j,t}(\phi, \mathbf{p}, \sigma_\nu) = \sum_{\tau=0}^{\infty} (1-\phi)\phi^\tau s_{T,j,t-\tau}(\mathbf{p}_{t-\tau}, \sigma_\nu). \tag{19}$$

The term $(1-\phi)\phi^\tau$ reflects the share of investors that were last active at time $t-\tau$ and $s_{T,j,t-\tau}(\mathbf{p}_{t-\tau}, \sigma_\nu)$ denotes the share of active investors that would purchase fund j at time $t-\tau$ given the vector of expense ratios $\mathbf{p}_{t-\tau}$ and information frictions σ_ν. When computing Eq. (19) we assume that all investors were active in the first month of our sample (i.e. January 2000). A fund's active market share is zero in all months prior to the introduction of the fund. Note that to make the underlying economic mechanisms in our counterfactuals more transparent, we compute predicted market shares under the assumption that the market size is constant over time. For each counterfactual we consider, we then compute the equilibrium vector of expense ratios and predicted market shares.

E Extension: Accounting for Financial Advisers

Previous research has highlighted the importance of brokers/financial advisers in a household's investment decision. To understand how brokers impact the index fund choices of investors, we also consider the extension where we assume that investors choose index funds with the help of a broker.

E.1 Setup

We follow the setup developed in Robles-Garcia (2019) and further used in Egan et al. (2022) where we assume that all financial advisers are ex-ante identical. For each client i, the financial adviser chooses the index fund j from the set $\mathcal{J}_{T(i),m(j),t}$ that maximizes a weighted average of the financial adviser's and consumer's incentives, denoted $\pi_{i,j,t}$:

$$\pi_{i,j,t} = \omega_{T(i)} f_{j,t} + (1 - \omega_{T(i)})\tilde{u}_{i,j,t}.$$

The variable $f_{j,t}$ measures the commissions a financial adviser earns from selling index fund j, and the parameter $\omega_{T(i)}$ measures conflicts of interest and reflects the weight that financial advisers place on their own financial incentives (i.e., commissions) versus the financial incentives of their clients (i.e., consumer utility). If $\omega_{T(i)} = 0$ then there are no conflicts of interest. We also allow for conflicts of interest to vary potentially across retail and institutional investors. Note that we also assume that financial advisers maximize the perceived utility of investors $\tilde{u}_{i,j,t}$, which implies that financial advisers observe investor-product-specific demand shocks ($\epsilon_{i,j,t}$) and that financial advisers are subject to the same information frictions as investors.

Under the assumption that financial advisers are myopic in the sense that they maximize current flow profits, the market share of active investors of type T investing in fund j is given by:

$$s_{T,j,t} = \frac{\exp\left(\frac{\frac{\omega_T}{1-\omega_T} f_{j,t} - p_{j,t} + X'_{j,t}\theta_T + \xi_{j,T(i),t}}{\sigma_{\eta,T(i)}}\right)}{\sum_{l \in \mathcal{J}_{T,m(j),t}} \exp\left(\frac{\frac{\omega_T}{1-\omega_T} f_{l,t} - p_{l,t} + X'_{l,t}\theta_T + \xi_{l,T(i),t}}{\sigma_{\eta,T(i)}}\right)}, \tag{20}$$

which is the core of our estimation strategy.

E.2 Estimation

We estimate Eq. (20) in terms of log active market shares following our empirical strategy described in Section 4 to recover investors' preferences and the brokers' preferences

(ω_T):

$$\ln s_{T,j,t} = \underbrace{\varpi_T}_{\frac{\omega_T}{\sigma_{\eta,T}(1-\omega_T)}} f_{jt} - \underbrace{\alpha_T}_{\frac{1}{\sigma_{\eta,T}}} p_{jt} - X'_{j,t} \underbrace{\Gamma_T}_{\frac{\theta_{T(i)}}{\sigma_{\eta,T}}} + \underbrace{\mu_{T,m(j),t}}_{\ln\left(\sum_{l \in \mathcal{J}_{T,m(j),t}} \exp\left(\frac{\frac{\omega_T}{1-\omega_T} f_{l,t} - p_{l,t} + X'_{l,t}\theta_T + \xi_{l,T(i),t}}{\sigma_{\eta,T(i)}}\right)\right)} + \underbrace{\zeta_{T,j,t}}_{\frac{\xi_{T,j,t}}{\sigma_{\eta,T}}}. \tag{21}$$

An empirical challenge is how to measure broker commissions. We measure broker incentives using 12b-1 fees. 12b-1 fees are used to compensate financial intermediaries for providing services to investors and to pay advertising and marketing expenditures. Evidence from The Investment Company Institute indicates that, on average, 92% of 12b-1 fees are paid to brokers/financial advisers, 6% are paid to underwriters, and 2% are used for marketing expenditures.[31]

One concern is that 12b-1 fees are potentially endogenous and correlated with unobserved demand shocks. To account for this potential endogeneity, we instrument for the actual 12b-1 fees a fund pays using the maximum contractual 12b-1 fee lagged by one year. Funds are required to report the maximum annual charge deducted from fund assets to pay for distribution and marketing costs (12b-1 fees) which may be larger than the actual fee paid in a given year. We use the maximum contractual 12b-1 fee as an instrument because it appears highly sticky in the data (e.g., the 1-year autocorrelation is 0.96) and we lag it by a year with the idea that contractual fees are uncorrelated with future demand shocks.

We report our corresponding estimates in Table A6. Consistent with intuition, we find a positive relationship between our measures of broker incentives and index fund demand. We also estimate elasticities of demand ranging from 1.8 to 2.0 for retail investors and 2.6 to 2.7 for institutional investors, which are consistent with our baseline demand estimates (Table 2). In the bottom of the panel we report the value of ω, which measures how a broker trades off her private financial incentives with the financial incentives of her client. The results in column (2) indicate that brokers are willing to trade-off a 1 percentage point increase in 12b-1 fees (92% of which are historically paid to brokers) with a 0.49 ($= 0.33/(1-0.33)$) percentage point increase in expense ratios. In other words, the estimates suggest that brokers place roughly 2 times ($= (1-0.33)/0.33$) the weight on their client incentives relative to their own. While still relevant, the conflicts of interest in the index fund market we estimate are smaller than what has been estimated in other markets such as the structured product

[31] `https://www.ici.org/system/files/attachments/fm-v14n2.pdf`

and variable annuity market (Egan, 2019; Egan et al., 2022). This is intuitive because the index fund market is more transparent than each of those markets. One might also expect broker incentives to potentially be more relevant for actively managed funds.

We consider how conflicts of interest impact the expense ratios that both institutional and retail investors pay in equilibrium. We implement this counterfactual by setting 12b-1 fees equal to zero and decreasing marginal costs by the corresponding amount. The results indicate that the effects of conflicts of interest are modest in the index fund market (see Appendix Figure A9 and Table A8). If we keep the product space fixed, eliminating conflicts of interest would reduce the expense ratios that retail and institutional investors pay by 8 and 2 basis point, respectively.