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[||]

October 2024

Abstract

The number of index funds increased drastically from 2000 to 2020, partially fueled by the emergence of exchange-traded funds (ETFs). Despite the growing availability of similar products, price dispersion persists, indicating significant market power among index funds. We show evidence that investor inertia along with other market frictions may be restricting competition. To understand the sources and implications of market power, we develop a tractable empirical dynamic model of demand for and supply of index funds that accounts for inertia, information frictions, and heterogeneous preferences. Using a novel identification strategy, we find that inertia is high, with only 1-3% of households updating their portfolio each month. Although inertia is high, its impact on the investment behavior of households is limited because investors struggle to optimize due to high information frictions. We show that although the introduction of ETFs lowered expense ratios through both the cost advantage of ETFs and increased competition, demand-side frictions substantially dampened the 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, the University of Chicago, the University of Florida, and the University of Wisconsin IO/Finance Reading Group.

[†]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 have been touted as a low-cost and transparent alternative to active investment management and have grown in popularity over time. From 2000 to 2020, the number of index funds increased more than six-fold. Much of this growth was driven by the introduction of low-cost exchange-traded funds (ETFs), which was expected to generate a "price war" with mutual funds.[1] However, despite the rapid entry of new products, there remains wide variation in the expense ratios of funds that track the same or similar benchmarks, and many funds with expensive fees have a non-trivial market share.

This paper studies why index funds hold market power and why the introduction of ETFs has not disciplined the market. We document that newly launched low-cost funds take years to grow their market share, suggesting that investors may go long periods without making an active investment choice. This may be due to neglect, including failure to monitor an account, or a conscious choice due to, for example, tax considerations. Meanwhile, if inertia were the only friction, all investors would eventually choose low cost funds within each category. However, the market share of new low-cost funds plateaus after the first few years and expensive funds retain significant market share. One explanation is that even when investors make an active choice, they face search or information frictions (e.g., Hortaçsu and Syverson, 2004). Another explanation is that investors have heterogeneous preferences over differentiated products. These data patterns suggest that both of these frictions—inertia and information frictions—in addition to horizontal differentiation, could be at play in this market, and we need a model that incorporates these mechanisms to understand the driver of market power.

From a policy perspective, it is also important to quantify the role of information frictions and inertia. If information frictions make it difficult for investors to choose low-cost funds, this motivates transparency rules such as the SEC's recent proposal to address misleading or deceptive practices, as well as 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 could have a larger impact on market power.[3] The effectiveness of regulating new product entry, price discrimination, and broker commissions will also depend on the degree of frictions that investors face.[4]

We develop and estimate a new model of index fund supply and demand where investors choose index funds in the presence of inertia, information frictions, and preference heterogeneity. Consumer inertia can create complex incentives when firms are forward-looking, and

[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.

[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.

we develop a tractable framework to model inertia and the interaction of inertia with other frictions. Moreover, we use new data and novel identification strategies to separately identify frictions from preferences. Using our empirical model, we assess how inertia and information frictions affect competition and market power within the index fund market, taking into account firms' supply-side response. We also examine how demand-side frictions interact with the advent of ETFs.

We model index fund demand as a discrete choice problem with the idea that an investor chooses an individual branded index fund within a specific investment class/category (e.g., Lipper Class), conditional on the investor's initial decision to invest in the specific investment class/category.[5] Each period, with some probability, investors are either active (i.e., 'awake') or inactive (i.e., 'asleep'). Inactive investors maintain their holdings from the previous period, while active investors re-optimize their choices. When investors are making an active choice, they are still subject to information frictions which diminish their capacity to select the optimal index fund. We model information frictions and investor inactivity separately given that they appear to have distinct underlying causes and policy implications in this market.[6] In addition, we allow investors to have heterogeneous preferences over index funds such that the product space is horizontally differentiated. Importantly, each of these mechanisms may give rise to market power.

We allow preferences and the level of frictions to vary across two investor types: retail and institutional investors. For example, one might expect information frictions to pose a greater problem for retail investors. The index funds available to each type also differ. Institutional index funds are only purchased by institutions while ETFs are purchased by both institutional and retail investors.

We estimate the model using fund-level data from CRSP over the period 2000 to 2020. We face two key empirical challenges. The first is separately identifying inertia from persistent preference, which is challenging because both can give rise to the same data pattern of persistent demand. That is, an investor may continue holding the same fund every month because they are inactive, or because they have a persistent preference for an unobserved characteristic of the fund.

We use a new strategy to address this problem that is straightforward and flexible to implement. To measure the fraction of consumers who are inactive and do not make an active investment decision in a period, we would ideally like to measure the causal effect of a one dollar increase in a fund's past assets under management (AUM) on its current AUM. We examine persistence of an exogenous shock to investors' past holdings using variation in past monthly fund returns. For example, if a fund experienced strong returns two months ago, this acts as a

[5] Consequently, we abstract away from the investor's more general portfolio choice problem.

[6] 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 on whether inertia and information frictions are linked.

positive exogenous shock to lagged AUM. How these past return shocks translate into current AUM tells us the degree of investor inertia. One 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 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.

Our estimates suggest that roughly 97% (94%) of retail (institutional) investors are inactive each month, which means that 28% of retail investors update their portfolio at least once each year. These estimates are in line with survey responses where investors are asked how often they adjust investments. We also are able to directly validate our estimates using new data from the SEC on new mutual fund sales and redemptions.

With our estimates of inertia in hand, we turn to estimating the preferences of investors. We use our estimates of inertia to calculate active demand each period, and we recover the preferences of investors by estimating a standard Berry (1994) demand system for institutional and retail investors. We estimate the elasticity of demand to be 1.6 and 2.7 for retail investors and institutional investors, respectively. It is important to note that the elasticity of demand we recover is a function of both information frictions and preference heterogeneity.

The second empirical challenge is to separately identify information frictions from preferences. We use additional data based on investors' choices in 401(k) plans. When making investment allocation decisions, 401(k) participants typically choose from a fixed investment menu of mutual funds that is determined by the plan sponsor (e.g., participant's employer). In addition, by law, 401(k) investors observe the full menu and receive expense- and performance-related summary disclosures (Kronlund et al., 2021). Since investors choose from a simplified menu with relative transparency, 401(k) plans are a setting with minimal information frictions.[7] 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, although investors often still view index funds as horizontally differentiated products. Even if consumers face some frictions in this setting, our estimates are useful as they provide an upper bound on the extent of preference heterogeneity (or a lower bound on the extent of information frictions). We find that the demand elasticity is 4.2 in this setting. For retail investors, this suggests that information frictions are roughly 1.6 times more important than preference heterogeneity in the index fund market.

With our estimates of consumer demand for index funds, we return to the original question of why competition does not eliminate market power. Answering this question requires us to specify the supply side. In our model, index fund managers compete for assets in a dynamic and differentiated Nash Bertrand expense-ratio setting game. Consistent with the data, index fund

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

managers are multi-product issuers and potentially price discriminate across institutional and retail investors by separately offering funds that are only available to institutional investors. Index fund managers set expense ratios to maximize the present discounted value of future profits while accounting for investor inertia, information frictions, and preference heterogeneity.

We show that these frictions have distinct implications for the optimal pricing behavior of index fund managers. Inertia has two potentially offsetting effects for how index fund managers set steady-state prices (e.g., Beggs and Klemperer, 1992). On the one hand, inertia increases the incentive to invest in new consumers, as demand from investors today will be more persistent into the future. On the other hand, an increase in inertia makes the demand curve more inelastic, which incentivizes managers to charge a higher expense ratio and harvest existing investors. We derive a simple expression for steady state index fund pricing with inertia for our framework. We show that with efficient capital markets (e.g., expected returns equal required returns), the "invest" and "harvest" incentives perfectly offset such that inertia has no impact on the pricing behavior of managers; however, in practice, we show that we would expect the presence of consumer inertia to increase steady state prices. In contrast, an increase in either information frictions or preference heterogeneity always incentivizes managers to increase their expense ratios.

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 29 (14) basis points, which implies an average (median) markup of 32 (22) basis points.

We use our model estimates to first quantify the effect of demand-side frictions investors face. We find that information frictions are a key source of market power. Eliminating these frictions by, for instance, mandating additional disclosure, would lower average expense ratios for retail investors by 32%. Although 97% of retail investors are inactive each month, inertia has a more modest impact on the expense ratios paid by consumers, reducing them by 12%. The limited demand-side response, which is perhaps surprising, can be attributed to the fairly severe information frictions that hinder investors' ability to make optimal investments when making an active choice. As a result, merely enabling investors to re-optimize more frequently (i.e., removing inertia) offers little value if the investors are not effectively optimizing in the first place. In contrast, eliminating both inertia and information frictions greatly impacts the prices retail investors pay. In this case, the average expense ratio retail investors pay falls by 71%, and the standard deviation of prices falls by 45%.

Institutional investors are more active, face less information frictions, and are more price-elastic than retail investors, so fund managers may face incentives to price discriminate between them. We find that eliminating price discrimination would lower the average expense ratio paid by retail investors by 23%. 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.[8]

Lastly, we consider how the introduction of ETFs impacted market competition and interacted with inertia and information frictions. ETFs differ from mutual funds in two crucial aspects relevant to competition. First, the marginal operating cost of an ETF is typically lower than that of a comparable index mutual fund (Jiang et al., 2023). Second, ETFs are inherently available to all investors, thereby precluding price discrimination across retail and institutional investors. Our counterfactual results indicate that the introduction of ETFs lowered the average retail index fund expenses by 19%; roughly half of the effect comes from the cost advantage of ETFs, and the other half comes from the competition channel. However, the introduction of ETFs would have had a larger effect in the absence of information frictions and inertia since these frictions limit the uptake of low-cost ETFs. These findings underscore the substantial impact of product innovation on the competitive dynamics within the industry and its interaction with frictions faced by investors.

Related Literature

It is well documented that even when financial products are similar, there is often large price dispersion and consumers often fail to choose the lowest price (see Campbell (2016) and Clark et al. (2021) for an overview).[9] 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 explore the role of search costs. Using a novel empirical approach, the authors find that they can rationalize the observed dispersion in prices with modest search costs. 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.[10] Similarly, Janssen and Thiel (2024) study how search frictions and preferences contribute to persistent investment in active funds.

While these types of models are quite flexible, Hortaçsu and Syverson (2004) note that their measure of search costs potentially captures additional factors such as individual preference heterogeneity (e.g., horizontal differentiation) and switching costs, which cannot be separately identified from information costs in their setting. While information frictions, including search costs, are one explanation for why individuals choose expensive financial products, some have

[8] 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 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`.

[9] 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.

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

argued that information frictions are unlikely to fully explain choice behavior in certain settings (Woodward and Hall, 2012; Grubb, 2015).

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 more generally (Kiser, 2002).[11] 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 incorporating these key frictions. We also propose 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.

Our paper also relates to the growing literature at the intersection of industrial organization and finance. In related work, An et al. (2021) develop a structural model of the ETF market that incorporates two-tiered competition of index providers and ETF managers. The authors document that the index providers (e.g., S&P Dow Jones) that create and license indices for ETF's to track also have substantial market power in addition to the ETF managers. Investors also have preferences over different brands of indices. Baker et al. (2022) and Egan et al. (2022) use similar demand-side approaches to recover investor expectations in the index fund market. 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 growing literature using a demand system approach to asset pricing (Koijen and Yogo, 2019a).[12] Our work also relates to the growing literature using IO-type demand systems (e.g., Berry (1994), Berry et al. (1995), etc.) in other 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). We show how to extend these types of frameworks to quantify the role of various frictions. For example, we factor in managerial incentives within 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 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.

[11]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).

[12]Other examples include Koijen and Yogo (2019b); Bretscher et al. (2020); Benetton and Compiani (2021) and Haddad et al. (2021).

2 Data and Motivating Evidence

2.1 Data

Our base index fund data set comes from CRSP mutual fund and covers the period 2000 to 2020. We restrict our attention to funds classified in CRSP as index funds, including both mutual funds and ETFs.[13] We observe monthly data on total net assets and returns and quarterly information on other fund characteristics such as expense ratios and Lipper classification. Index funds in the data are defined at the share class level, which implies many of the funds in our data share common investment portfolios with other funds in our data. 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 because we are interested in how share classes, in particular retail and institutional share classes, contribute to the observed price dispersion in the index fund market.

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 different 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. Following Hortaçsu and Syverson (2004) we construct load-adjusted expense ratios. We add 1/3rd of all loads to the expense ratio because, given our estimates of inertia, investors update their portfolios roughly once every three years on average. This adjustment has a very minor impact on expense ratios because most funds do not have front or rear loads, especially in more recent years.[14] For example, 90% of index funds in our sample did not have any loads in 2020.

[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]We use load adjusted expense ratios in our analysis but note that our findings are quantitatively similar if we do not adjust expense ratios for loads or if we drop all funds with loads.

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. Observations are at the fund-by-month level. The variables *Retail Mutual Fund, Institutional Mutual Fund,* and *ETF* are all indicator variables.

One dimension we are particularly interested in is understanding the behavior of institutional versus retail investors. For index funds structured as mutual funds, we can observe in CRSP whether the fund is an institutional fund or retail mutual fund. In contrast, exchange traded funds may be purchased by either institutional or retail investors. We use quarterly institutional holdings data (13F) to determine the share of ETF assets held by institutional versus retail investors. Roughly 35% of the funds in our sample are classified as retail mutual funds, 26% are classified as institutional mutual funds, and the remaining 38% are classified as ETFs.

Employer-sponsored investment accounts provide a simplified menu of fund offerings 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. For index funds that are relatively homogeneous, the presence of price

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

dispersion provides initial evidence consistent with market power.

Figure 1 Panel (a) shows the growth in the number of index mutual funds and ETFs over the period. The number of mutual funds increased by almost 3-fold. ETFs were almost non-existent at the start of the period, but outnumbered mutual funds by the end of the period. While some of the new funds were differentiated, Panel (c) focuses on the most homogeneous index funds and shows that the number of similar funds increased over time as well. Finally, Panel (b) and (d) show that the market size also drastically increased over the period.

Figure 1: Growth of Index Funds

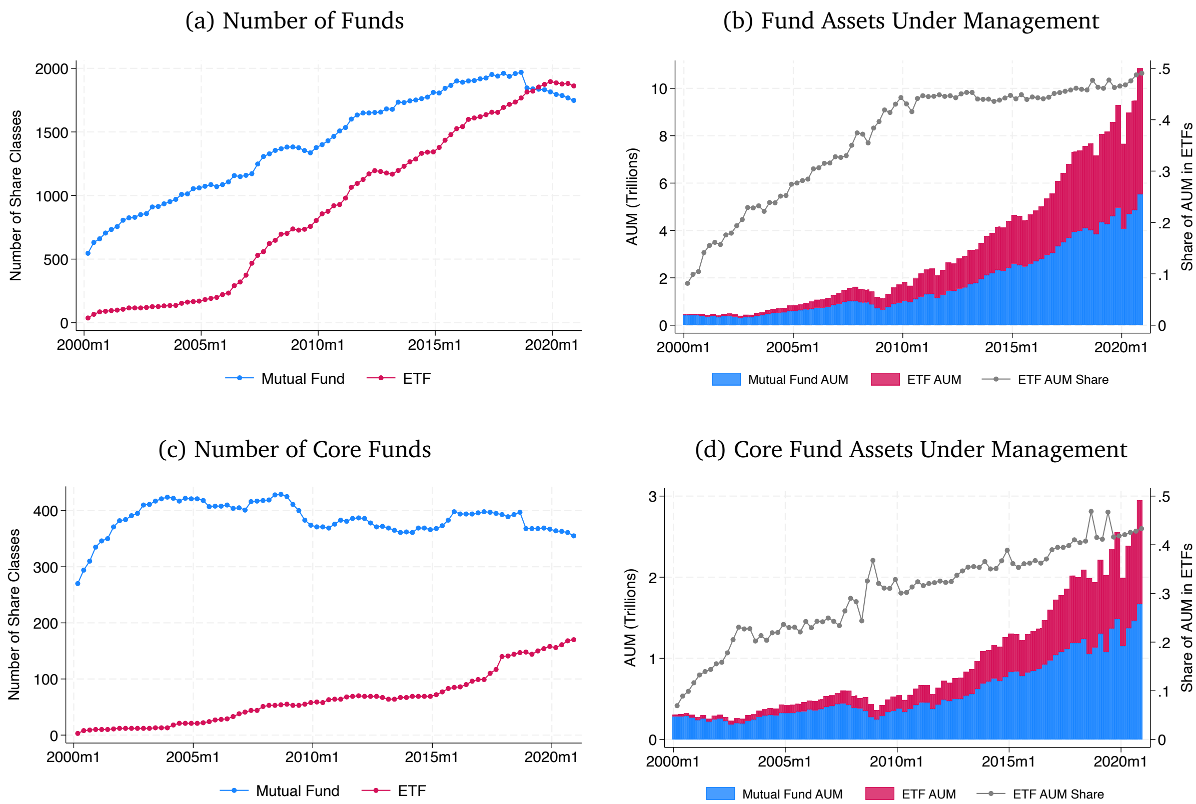


Figure 1 displays the growth of index funds over time. Panel (a) displays the number of index mutual funds and ETFs over time. Panel (b) displays the AUM of index mutual funds and ETFs, along with the ETF share of total index fund AUM. Panels (c) and (d) present the same figures, but for a restricted set of ten core lipper classes, which include the S&P 500, and Large/Mid/Small Cap x Value/Neutral/Growth.

All of these facts would seem to imply that market power should decrease over the period and the market should be converging to perfect competition. Therefore, it is surprising that price dispersion remained relatively constant over time. Figure 2 displays the distribution of fund expense ratios over the period. Panels (a) and (b) display the equal-weighted and asset-weighted distribution of expense ratios for our full sample. Panel (a) indicates that the average index fund expense ratios have fallen 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. However, the interdecile range has remained relatively constant at 150-160 basis points over the bulk of our sample.

Figure 2: Distribution of Fund Expense Ratios over Time

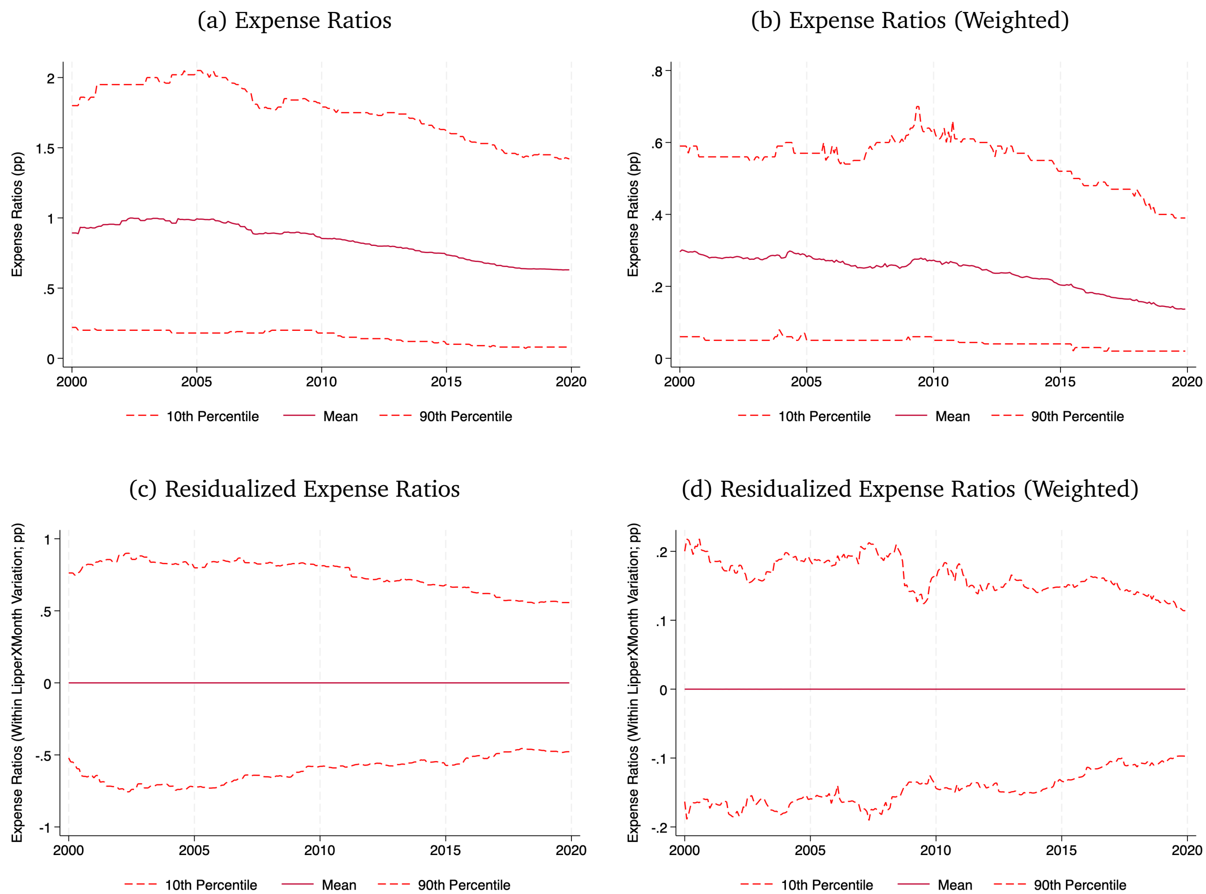


Figure 2 displays the distribution of index fund expense ratios over time. Panels (a) and (b) display the equal weighted and asset-weighted distribution of expense ratios. Panels (c) and (d) display the equal weighted and asset-weighted distribution of residualized expense ratios, where we residualize 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 distribution of expense ratios (Panel a) with the asset-weighted distribution of expense ratios (Panel b) provides prima facie evidence about investors' elasticity of demand. The asset-weighted distribution is shifted downwards relative to the equal-weighted distribution of expense ratios, 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 expensive index funds.

Dispersion in expense ratios could be driven in part by product differentiation related to the underlying types of index funds/asset classes. For example, it could be the case that index funds classified in Lipper as commodities based metals funds are more expensive than funds classified in Lipper as S&P 500 index funds. Thus, some of the observed dispersion in expense ratios in Panels (a) and (b) is potentially attributable to these types of observable fund differences. To account for these differences we residualize expense ratios by regressing them on Lipper Class-by-month fixed effects. Figure 2 Panels (c) and (d) plot the residualized expense ratios. These fixed effects explain 35% of the variation in fund expense ratios and 79% of the variation in fund returns. Thus, even after accounting for differences across Lipper Classes, there is still a substantial amount of variation in expense ratios. The results indicate that over the whole sample funds in the 90th percentile were 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 secular decline in average expense ratios, inertia could be quite costly for those 82-88% of investors that 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 by examining how fund flows respond to the introduction of new low-cost funds. Since inertia prevents investors from switching to cheaper new funds, the sluggishness of outflows into these funds provides a preliminary test of whether inertia plays an important role in this market.

Figure 3 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 over time. We focus on funds that survive 5 or more years to avoid selection issues. Consistent with there being a large degree of inertia, demand for inexpensive funds is initially low and 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 only 5%. The fact that demand for low cost funds is not higher in the long run when most

[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.

investors have made an active choice may be a result of information frictions or preference heterogeneity.

Figure 3: Market Share of a Newly Launched Low Cost Fund

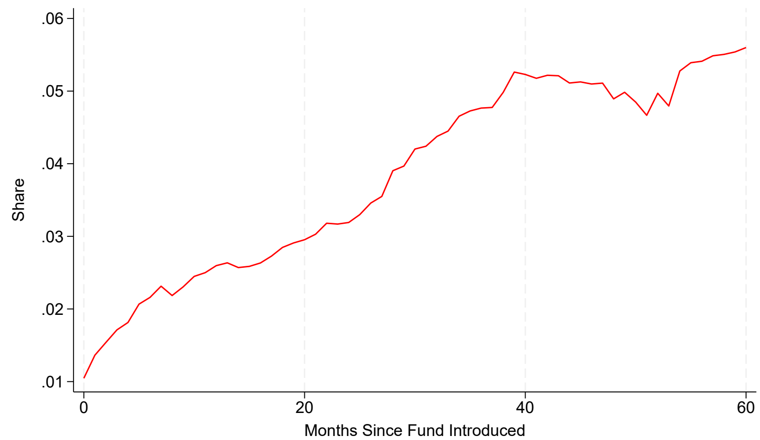


Figure 3 displays the average market share of a newly launched low-cost fund which survives at least 5 years by month since introduction. Low-cost funds are defined as those in the bottom quartile of the price distribution in their lipper class at the time of launch.

Information Frictions: Given that the expense ratios of 401(k) plans are more transparent, investors may be more price sensitive. We examine price dispersion for 401(k) plans in Figure A1. There is a large difference between the equal-weighted distribution of expense ratios (Panel a) and the asset-weighted distribution of expense ratios (Panel b), suggesting that investors in 401(k) plans are quite price sensitive. While the interdecile range of asset-weighted expense ratios is about 40 basis points for index funds, the interdecile range is only 8 basis points at the end of the sample for 401(k) plans. Given that preference heterogeneity is likely similar, lower price dispersion in 401(k) plans relative to the broader index fund market provides initial evidence that information frictions play an important role in the index fund market.

Price Discrimination: Index fund managers will often create multiple funds and ETFs that share the same index/underlying portfolio. In particular, mutual funds often have a class structure which allows intermediaries to explicitly price discriminate across investor types. The fund manager will then typically offer a less expensive version of the mutual fund to institutions, who are more price sensitive, and a more expensive version of the mutual fund to retail investors, who are less price sensitive. For a given underlying portfolio (identified in the data as crsp_portno) and moment in time, we calculate the difference between the average expense ratio of retail mutual funds and that of institutional mutual funds. Figure 4 displays the distribution of this difference for those portfolios that are held by at least one retail and one institutional fund. The results indicate that, on average, an institutional fund charges an expense ratio that is 94 basis point lower 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 4: Within Portfolio Variation in Expense Ratios

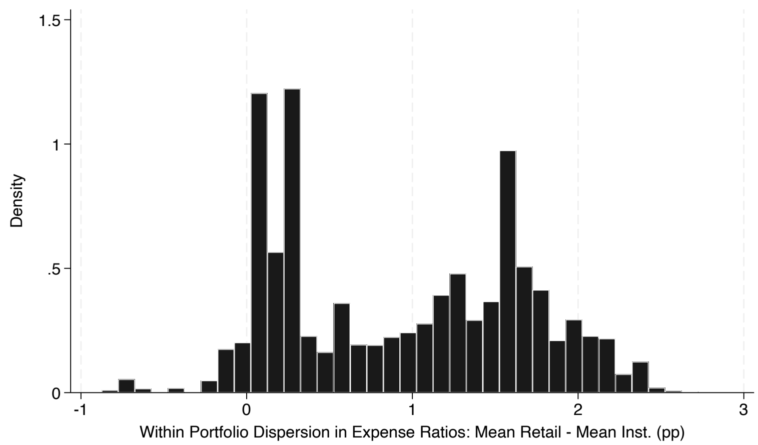


Figure 4 displays the within portfolio dispersion in expense ratios across retail and institutional funds. We focus on differences between retail and institutional funds. For a given underlying portfolio (identified in the data as $crsp\_portno$) and moment in time, we calculate the difference in the 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 empirical model of supply and demand for index funds. Our objective is to use the model to provide new insight into the mechanisms driving the price dispersion we observe in the data. The baseline version of the model includes two types of agents: heterogeneous investors who possess demand for index funds and index fund managers who create a suite of available index funds. We also consider an extension of the model in Appendix D where we introduce financial advisers who potentially distort the investment decisions of investors due to conflicts of interest.

Motivated by the evidence in Section 2.2, we focus on four mechanisms that may explain why investors buy expensive index funds and why funds have market power in our baseline framework. First, investors have heterogeneous preferences over index funds such that index funds are horizontally differentiated. Second, investors face information frictions and have imperfect information about fund characteristics when choosing a fund. Third, investors exhibit inertia and do not actively update their portfolios every period. Fourth, index fund managers are able to price discriminate across institutional and retail investors.

3.1 Investors: Demand for Index Funds

We model an investor's demand for index funds as a discrete choice problem. We consider an investor's index fund choice conditional on her decision to invest in a specific investment category/asset class (e.g., small-cap value, mid-cap growth, etc.), which allows us to abstract away from the investor's portfolio choice problem. For example, we model an investor's decision to invest in a particular Vanguard S&P 500 Index fund over a similar BlackRock S&P 500 Index fund, but do not model the investor's initial decision of whether and how much to invest in S&P 500 Index Funds. In our counterfactuals, we assume that changes in expense ratios affect demand for funds within a category/asset class but do not cause investors to switch category/asset classes.

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 ability to purchase certain types of funds. For example, 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]

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

3.1.2 Information Frictions

Investors may not research all funds and may not fully understand fund characteristics such as the expense ratio. We assume that each investor potentially 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 model of information frictions is consistent with influential work on rational inattention in discrete choice models (Matějka and McKay, 2015). A growing literature has argued that the rational inattention model can explain investor behavior and demand for complex financial products (e.g., Kacperczyk et al., 2016; Brown and Jeon, 2023). As we show in Appendix B, one can consider a model in which each individual has a prior about the payoff of each fund and then optimally researches funds given costly information acquisition. This model yields an expression for expected utility that is equivalent to Eq. (1) when individuals have a homogeneous prior across options and share the same cost of information within type. 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:

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

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:

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

The set $\mathcal{J}_{T(i),m(j),t}$ denotes the investor's consideration set: the set of index funds available to a type $T(i)$ 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 ($\theta_{T(i)}$) as well as decompose the error term into two components: one due to information frictions ($\sigma_{\nu,T(i)}$) and the other due to product differentiation ($\sigma_{\epsilon,T(i)}$).

3.1.4 Inertia

We consider the possibility that investors suffer from inertia. In each period, there is some probability an investor will be active and some probability the investor will be 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 in a given period is heterogeneous across investor types but is constant across investors conditional on their type. The probability a type T investor is inactive in a given period is ϕ_T and the probability she is active is $1-\phi_T$. 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). Alternatively, it is possible that the ϕ_T is endogenous and responds to market conditions, a situation we consider in Section 4.1.

We assume that when active investors choose a fund, 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 ϕ_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, denoted $AUM_{T,j,t}$ is given by:

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

The term $AUM^{Inactive}_{T,j,t} \equiv \phi_T AUM_{T,j,t-1}(1+r_{m(j),t-1})$ 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 $AUM^{Active}_{T,j,t} \equiv (1-\phi_T)M_{T,m(j),t}s_{T,j,t}$ 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 only purchased by retail investors and institutional mutual funds are only purchased by institutional investors. Both retail and institutional investors can purchase ETFs.

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 funds earn a markup of $p_{j,t} - c_j$ for each dollar of assets collected, where c_j is the marginal cost of operating the fund.

We assume that index fund managers play a differentiated, multi-product, dynamic, Nash-Bertrand, expense ratio setting game. 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/expense ratios $\mathbf{p}_{k,t}$, $\mathbf{p}_{k,t+1}$,... to maximize the presented discounted value of future profits discounted by β.

$$\max_{\mathbf{p}_{k,t},\mathbf{p}_{k,t+1}\cdots|\mathbf{p}_{-k,t},\mathbf{p}_{-k,t+1}\cdots} \sum_{\tau=t}^{\infty} \beta^{\tau-t} \sum_{j \in \mathcal{K}_{k,m}} (AUM_{R,j,\tau} + AUM_{I,j,\tau})(p_{j,\tau} - c_j). \quad (5)$$

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, approximately 70% of funds have a market share smaller than 1%. Given the many funds with very small market share, it is unlikely that fund managers internalize their future strategies. Second, prices appear quite sticky. As of 2020, only 3.5% funds (weighted by assets) charged an expense ratio that was more than 10 basis points lower than what the fund charged five years previously in 2015. It is important to note that even with this assumption, firms set prices while fully accounting for how current demand impacts future demand and profitability due to inertia.

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:

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

The first order condition is fairly standard except for the term $\sum_{\tau=t+1}^{\infty} \left(\beta(1+\tilde{r}_{m(j),\tau})\phi_R\right)^{\tau-t}(p_{j,\tau} - c_j)$, which captures the effects of inertia. For every investor the fund attracts today, there is a ϕ_R chance the investor will remain in the subsequent period, a ϕ_R^2 chance the investor remains for at least two periods, and so on. Furthermore, inactive investors' assets are expected to grow based on fund expected returns, $\tilde{r}_{m(j),\tau}$.[18]

In the static problem (e.g. no inertia), a firm's assets today do not impact 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, $\frac{\partial AUM_{R,j,\tau}}{\partial AUM_{R,j,t}} = ((1+\tilde{r}_{j,t})\phi_R)^{\tau-t}$. Thus, when setting prices, firms account for how changing prices impact both current and future demand. In order to attract new investors who will be inactive in the future, firms may want to set lower prices the more inertia is present. 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:

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

Consequently, an increase in inertia will make a fund's current assets less sensitive to expense ratios, which, all else equal, 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 $p_{j,t} - c_j + \sum_{\tau=t+1}^{\infty}(\beta(1+\tilde{r}_{m(j)})\phi_R)^{\tau-t}(p_{j,\tau} - c_j) = \frac{p_j - c_j}{1-\beta\phi_R(1+\tilde{r}_{m(j)})}$, the manager's first order condition simplifies to:

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

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

[18] For ease of exposition, we assume that fund expected returns is constant over time and constant across index funds in a given market (i.e. $\tilde{r}_{j,t} = \tilde{r}_{m(j)}$).

zero (i.e., $\phi_R = 0$), the first order condition simplifies to a standard static first order condition. Second, given a growth-adjusted discount factor of 1 (i.e., $\beta(1+\tilde{r}_{m(j)}) = 1$ 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, our model implies that these forces perfectly offset when the growth-adjusted discount factor is equal to 1. 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 $\beta_m = \frac{1}{1+\tilde{r}_m}$.

In practice, we would expect the growth-adjusted discount rate to be positive such that the discount factor is slightly less than one.[19] Given a growth-adjusted discount factor less than one, the first order condition implies that with inertia, the index fund manager will set a higher markup than in the static model without inertia. As the growth-adjusted discount factor decreases (i.e., the discount rate increases), managers will place more value on extracting profits from current investors than on profits from future demand.

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 an index 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 $p_{j,t}$, given our demand system, in steady state is given by

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

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. Again notice that if either $\phi_I = \phi_R = 0$ or $\beta(1+\tilde{r}_{m(j)}) = 1$, then firm's first order condition for setting prices in the dynamic model simplifies to the standard first order condition for the static model.

[19] For example, due to fund expense ratios the expected fee-adjusted growth rate will be less than the required return.

4 Estimation and Results

We estimate our structural model of demand and supply for index funds using the mutual fund data set described in Section 2.2. On the investor demand side, we first estimate inertia using an instrumental variable strategy. Once investors' inertia is pinned down, we estimate their preference parameters, and then, we use additional data on choices in 401(k) plans to estimate information frictions. Finally, with the demand parameters in hand, we estimate the index fund managers' marginal costs of operating index funds on the supply side.

4.1 Investor Demand

4.1.1 Inertia

Demand for an index fund is comprised of new demand from active investors and past demand from inactive investors. Following Eq. (4), the total assets under management held by investors of type T of fund j at time t is equal to:

$$AUM_{T,j,t} = \phi_T AUM_{T,j,t-1}(1 + r_{j,t}) + AUM^{Active}_{T,j,t}. \tag{8}$$

In principle, one could estimate the fraction of inactive consumers (ϕ_T) by simply regressing current assets under management on lagged assets under management scaled by returns.

One challenge in estimating Eq. (8) is that lagged assets under management $AUM_{T,j,t-1}$ are potentially endogenous and correlated with $AUM^{Active}_{T,j,t}$, which is unobserved. For example, if investor preferences for funds are correlated over time, then we would expect lagged assets to be positively correlated with the assets held by active consumers. This endogeneity bias would cause us to overestimate the fraction of inactive investors. To address the endogeneity issue, we need an instrument that is correlated with lagged assets but that is uncorrelated with contemporaneous demand by active consumers.

One potential instrument we consider is past returns. The instrument will be relevant provided at least some investors are inactive each period and do not re-balance their portfolios. The instrument will be exogenous provided that past returns are uncorrelated with current demand from active investors. In other words, the instrument will be valid as long as active investors do not chase returns. While in a rational benchmark model we might not expect investors to chase returns, there is a long literature suggesting that at least some investors chase returns. To allow for return chasing, we assume that investors that chase returns do so based on 1-month, 3-month, 6-month, 12-month, and year-to-date cumulative returns, which are the returns that are typically reported by index funds. We then instrument for lagged assets using the past twelve monthly returns. The idea is that conditional reported returns, the choice of active investors is not affected by past monthly returns. We also consider additional specifications where we include market-by-time fixed effects to help control for return chasing with the idea that investors primarily chase returns at the index/Lipper Class level rather than the fund level.[20]

We estimate the fraction of inactive consumers using the following empirical analog of Eq. (8):

$$\ln AUM_{T,j,t} = \phi_T(i) \ln AUM_{T,j,t-1}(1 + r_{j,t}) + X'_{j,t-1}\Gamma + \iota_{T,j,t}, \tag{9}$$

where observations are at the fund-by-month-by-investor type level.[21] The key independent variable of interest is $\ln AUM_{T,j,t-1}$. Importantly, we instrument for lagged assets using the past twelve monthly returns while simultaneously controlling for 1-,3-,6-, 12-month, and year-to-date cumulative returns. In all our specifications, we also 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. The estimate ϕ_T measures the causal impact of an exogenous change in lagged AUM on current AUM, which we attribute to inertia. For example, a 1% exogenous increase AUM last period leads to a ϕ_T percent increase in AUM this period.[22]

We report the corresponding estimates in Table 2. Panels (a) and (b) present the results for retail investors and institutional investors, respectively, and the baseline results are in column (2). For retail investors, a 1% increase in lagged assets under management causes a 0.973% increase in assets under management today in the baseline. In other words, we estimate that 97.3% of retail investors are inactive each month. Put differently, our estimates imply that roughly 28% ($= 1 - 0.973^{12}$) of retail investors update their portfolios at least once a year. We include year-month fixed effects in our baseline specification; however, columns (3) show that estimated inertia is quite similar when including year-month-market fixed effects.

[20]In our specifications where we include market-by-time fixed effects, we control for 1-month and year-to-date cumulative returns because the market-by-time fixed effects capture much of the variation in returns.

[21]Note that we estimate our empirical specification in logs rather than levels. This is because we implicitly assume that inactive investors may passively allocate money to their account each period based on their previous holdings. For example, even if inactive investors do not actively choose index funds each period they may passively allocate their savings to their existing holdings each period. Modeling AUM in terms of logs also has the additional benefit for tractability that the implied active assets under management will always be positive. In Appendix Table A1 we estimate the corresponding regression in levels and find quantitatively similar estimates of inertia.

[22]We attribute this effect to inertia, thereby implicitly assuming that conditional on the expected returns and quality of the fund, investors do not care about lag fund size.

Table 2: Investor Inertia

(a) Retail Investors

VARIABLES	(1)	(2)	(3)
Lag AUM	0.989***	0.973***	0.979***
	(0.000)	(0.016)	(0.012)
Observations	345,689	342,407	340,453
R-squared	0.984	0.983	0.981
IV		X	X
Year-Month FE	X	X	
Year-Month-Mkt FE			X

(b) Institutional Investors

VARIABLES	(1)	(2)	(3)
Lag AUM	0.992***	0.943***	0.977***
	(0.000)	(0.013)	(0.012)
Observations	324,085	320,308	318,488
R-squared	0.986	0.984	0.984
IV		X	X
Year-Month FE	X	X	
Year-Month-Mkt FE			X

Note: Table 2 displays the estimates corresponding to a linear regression model (Eq. 9). Observations are at the index fund-by-month level. The dependent variable is log assets under management. In Panel (a) we restrict our attention to retail investors/AUM and in Panel (b) we restrict our attention to institutional investors/AUM. As described in the text, we address the endogeneity of *Lag AUM* using an instrumental variables approach in columns (2)-(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, 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. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

We find somewhat lower rates of inertia for institutional index fund investors. Roughly 51% ($= 1 - 0.943^{12}$) of institutional investors update their portfolios at least once a year in the baseline. It is also useful to compare our OLS estimates versus our IV estimates (column 1 versus 2). As expected, we estimate a smaller fraction of inactive investors once we account for the endogeneity of lagged assets under management.

In the Appendix, we validate our estimates of investor inertia using new data on fund sales and redemptions, which allows us to construct an alternative measure of inertia. Since 2019, the SEC requires mutual funds to report their monthly sales and redemptions as part of the N-PORT filings. We first use these data to calculate a lower bound of the active share of investors at the market level as the total value of new sales (redemptions) divided by total AUM. We find that for the median market, the total number of new shares purchased (shares redeemed) relative to total AUM is 2.1% (1.8%), which is consistent with our estimates of investor inertia (see Appendix Figure A2).

We can also use these data to directly estimate the share of active investors. Note that new sales for fund j at time t are equal to the share of active investors who choose fund j at time t but were holding a different fund in the previous period:

$$New\,Sales_{T,j,t} = (1-\phi)\,AUM_{T,-j,t-1}s_{T,j,t}.$$

where $AUM_{T,-j,t-1}$ denotes the AUM of all funds in market $m(j)$ excluding fund j at time $t-1$. We can calculate $1-\phi$ at the market level as $\sum_{j\in\mathcal{J}_{T,m,t}}\frac{New\,Sales_{T,j,t}}{AUM_{T,-j,t-1}}$. For the median retail (institutional) market, we estimate that 2.41% (1.69%) of investors are active each month (i.e. $\phi = 0.9759$), which is quantitatively similar to our baseline estimates.

The new sales data are also helpful in estimating Eq. (8) and (9) because the primary endogeneity concern arises from the fact that we do not observe active fund sales. We can use these new fund sales data to effectively control for active sales and directly address the endogeneity concern. In the Appendix, we re-estimate Eq. (9) by flexibly controlling for new fund sales. Again, our estimates of inertia for both institutional and retail investors are quantitatively similar to our baseline estimates.

One concern is that inertia is endogenous and varies across markets. If it is costly to research funds, investors may only make an active choice if price dispersion is high enough and there are large gains from making an active choice. We examine heterogeneity in inertia by price dispersion in the market. Appendix Table A3 shows there is no statistically significant difference in the share of inactive investors between markets with high and low price dispersion. If anything, inertia is higher when there is more price dispersion. This is not consistent with inertia and information frictions being linked. We also examine other dimensions of heterogeneity in inertia across funds and find modest evidence of heterogeneity in inertia by fund characteristics.[23] These results motivate our assumption that inertia is exogenous in our context.

[23]We allow inertia to vary with whether a fund has a front-, a rear-, or no-load since one might expect investors that purchase funds with loads suffer from greater inertia. As seen in Appendix Table A4, retail investors that purchase funds with rear-loads are slightly more likely to be inactive than investors that purchase funds without rear loads. We also examine whether inertia varies with past returns. One might expect that investors in funds that experienced positive returns to have stickier demand as a result of potential tax consequences. We find some modest evidence that retail investor inertia increases after a fund has experienced positive returns, but its economic magnitudes are quite small.

4.1.2 Active Investor Demand

These estimates of inertia help us separately determine the choices of active and inactive investors. Then, we estimate both retail and institutional investor demand using the revealed choices of active investors following our framework in Section 3.1. It is important to focus on the choices of active investors because these choices reflect investors' current preferences and information frictions with respect to available index funds. In contrast, because investors suffer from inertia, total assets are driven in part by the past choices of inactive investors, which depend on past preferences and product characteristics.

Given our framework (Eq. 3) and following Berry (1994), the market share of fund j in market m among active type T investors can be written in logs as:

$$\ln s_{T,j,t} = \frac{1}{\sigma_{\eta,T}}\left(-p_{j,t} + X'_{j,t}\theta_T + \xi_{T,j,t}\right) - \ln\left(\sum_{l\in\mathcal{J}_{T,m(j),t}} \exp\left(\frac{-p_{l,t} + X'_{l,t}\theta_T + \xi_{T,l,t}}{\sigma_{\eta,T}}\right)\right). \quad (10)$$

We estimate the corresponding regression specification:

$$\ln s_{T,j,t} = -\underbrace{\alpha_T}_{\frac{1}{\sigma_{\eta,T}}} p_{jt} + X'_{j,t}\underbrace{\Gamma_T}_{\frac{\theta_{T(i)}}{\sigma_{\eta,T}}} + \underbrace{\mu_{T,m(j),t}}_{-\ln\left(\sum_{l\in\mathcal{J}_{T,m(j),t}} \exp\left(\frac{-p_{l,t}+X'_{l,t}\theta_T+\xi_{T,l,t}}{\sigma_{\eta,T}}\right)\right)} + \underbrace{\zeta_{T,j,t}}_{\frac{\xi_{T,j,t}}{\sigma_{\eta,T}}}. \quad (11)$$

Observations are at the fund-by-month level. Importantly, we include market-by-time fixed effects $\mu_{T,m(j),t}$, which absorb the non-linear term in Eq. (10). Consequently, we can estimate Eq. (11) using linear methods to recover investors' demand parameters. We also control for 1-, 3-, 6-, 12-month, and year-to-date 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 whether the fund is an ETF, has a front load, or has a rear load.

As described above, to estimate the model we first need to calculate market shares among active investors. We use our estimates of inertia to calculate total assets held of fund j by active investors of type T at time t as

$$AUM^{Active}_{T,j,t} = \exp\left(\frac{\ln AUM_{T,j,t} - \hat{\phi}_T(i)\ln AUM_{T,j,t-1}(1+r_{j,t})}{1-\hat{\phi}_T}\right).$$

We then compute the market share among active type T investors for each fund j at time t in market m as:

$$s_{T,j,t} = \frac{AUM^{Active}_{T,j,t}}{\sum_{l\in\mathcal{J}_{T,m(j),t}} AUM^{Active}_{T,l,t}}.$$

This provides us with an estimate of active market shares, which is the dependent variable in our main demand specifications.

One additional challenge in estimating Eq. (11) using ordinary least squares (OLS) is that fund expense ratios ($p_{j,t}$) are potentially endogenous. For example, if an index fund manager anticipates high latent demand for their fund (e.g., high $\xi_{T,j,t}$), they may find it optimal to charge a higher expense ratio. This type of behavior would cause our OLS estimates of investors' sensitivity to prices α_T to be biased downwards such that investors appear less sensitive to prices than they actually are. To account for the potential endogeneity of expense ratios, we instrument for expense ratios with Hausman (1996) instruments. Specifically, we instrument for the expense ratio that an index fund manager k charges on its fund j at time t using the average expense ratio that fund manager k charges on all of its funds in other markets at time t (i.e., excluding market $m(j)$). 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 idea is that 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 provided that the fee that BlackRock charges on its high-yield bond fund is uncorrelated with demand shocks for BlackRock's large-cap value fund.

Table 3 displays our baseline demand estimates. We report the estimated perceived utility parameters for our retail investor sample in columns (1)-(2) and the estimates for our institutional investor sample in columns (3)-(4). In each specification, as expected, we estimate a negative and significant relationship between expense ratios and demand. In the bottom row of the table, we report the corresponding elasticity of demand. We estimate an elasticity of demand ranging from 1.2-2.7, depending on the exact specification. Consistent with intuition, we find that institutional investor demand is substantially more elastic than retail investor demand; institutional investors demand is roughly 69% more elastic than retail investor demand (column 2 vs 4). As described above, part of this could be due to the fact that retail investors have more severe information frictions such that $\sigma_{\eta,R} > \sigma_{\eta,I}$.[24]

[24] In Appendix Table A5 we estimate our demand specification where we use data on new mutual fund sales. Since 2019, funds are required to report the total net asset value of new shares purchased (excluding reinvestments of dividends and distributions). We find qualitatively similar estimates using these alternative data.

Table 3: Investor Preferences when Actively Demanding Index Funds

VARIABLES	(1)	(2)	(3)	(4)
Expense Ratio	-206.840***	-293.871***	-296.677***	-489.639***
	(2.264)	(8.484)	(2.337)	(9.302)
Observations	346,817	128,532	322,102	133,530
R-squared	0.153	0.068	0.256	0.139
Year-Month-Mkt FE	X	X	X	X
IV		X		X
Retail Sample	X	X		
Inst. Sample			X	X
Elasticity of Demand	1.2	1.6	1.7	2.7

Note: Table 3 displays the estimates corresponding to a linear regression model (Eq. 11). 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. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

We examine the robustness of the demand estimates to including cost shifters as instruments, which further addresses potential endogeneity of 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 (see Appendix Table A6). An alternative way to construct a cost shifter is to proxy fund managers' markup with the sum of adviser fees and distribution fees, and then use the difference between expense ratio and the markup proxy as an instrument following Janssen and Thiel (2024) (see Appendix Table A7). We find very similar elasticity of demand for both retail and institutional investors. In addition, we explore preferences for top fund managers by including an indicator for whether the fund manager is among the top three firms measured by total assets (BlackRock, State Street Bank, and Vanguard), and also find similar elasticity of demand (see Appendix Table A8).

4.1.3 Information Frictions

A challenge in the literature is that it can be difficult, if not impossible, to separately identify information frictions from preference heterogeneity. To address this, we estimate our demand system in the unique 401(k) setting where information frictions are likely to be close to zero. Investors choose 401(k) investments from a fixed menu of roughly 10 to 30 options. And within each broad asset class (e.g., large cap US equities), an average plan typically offers two index funds.[25] For the funds in the 401(k) menu, employers are also required to provide expense ratio- and performance-related disclosures, which are designed to be transparent for investors and shown to be effective in reducing information frictions (Kronlund et al., 2021). Therefore, we argue that investors have information on the full menu of 401(k) funds. Comparing demand by active investors in the 401(k) setting and in the broader index fund market provides insight into the magnitude of information frictions.

Given that 401(k) investors do not face information frictions, ($\nu_{i,j,t} = 0$), investors select the fund that maximizes their indirect utility ($u_{i,j,t}$) rather than their selecting the fund that maximizes their perceived utility ($\tilde{u}_{i,j,t}$). Therefore, following Eq. (3), the market share of fund j in 401(k) plan l is given by:

$$S_{j,l,t} = \frac{\exp\left(\frac{-p_{j,t}+X'_{j,t}\beta+\xi_{j,t}}{\sigma_\epsilon}\right)}{\sum_{n\in\mathcal{J}_{l,m(j),t}}\exp\left(\frac{-p_{n,t}+X'_{n,t}\beta+\xi_{n,t}}{\sigma_\epsilon}\right)}, \tag{12}$$

where $\mathcal{J}_{l,m,t}$ corresponds to the index funds in market m that are available in 401(k) menu for plan l at time t. As described in the previous section, we can then directly estimate Eq. (12) following Berry (1994) using our 401(k) plan data. Here, we recover the term σ_ϵ which measures the importance of product heterogeneity. In contrast, we previously recovered the term σ_η, which is a function of both information frictions and product heterogeneity. This allows us to separately identify information frictions and product heterogeneity.

Similar to the previous section, we include 401(k)-plan-by-market-by-time fixed effects, and use both OLS and instruments for expense ratios.[26] Due to concerns about investor inertia, we restrict our attention to newly created 401(k) plans in the first year they were introduced, when all investors were active and there was no inertia.

Table 4 displays the estimates. Column (1) displays our OLS estimates and column (2) displays our IV estimates. In both specifications, we find a negative relationship between demand and fund expense ratios. The elasticity of demand in our preferred specification (column 2) is 4.2, which is substantially higher than the retail investors' elasticity of demand in our main sample but closer to the institutional investors' elasticity of demand (Table 3). Given that information frictions are likely negligible in the 401(k) plan setting, this difference in demand elasticity implies significant information frictions for retail investors. Also, given that unobserved product heterogeneity is likely similar for institutional and retail funds,[27] these results suggest that information frictions help explain why retail investors have a lower elasticity of

[25]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.

[26]Given the nature of the 401(k) data, we construct our Hausman instruments following Egan et al. (2021) where we construct the instrument for fund j in plan l as the average expense ratio of all other funds offered by manager $k(j)$ that do not appear on the plan l's 401(k) menu.

[27]For example, roughly 80% of retail index funds are available to institutional investors through alternative share classes.

demand than institutional investors in Section 4.1.2.

Table 4: Demand for Index Funds In 401(k) Plans

VARIABLES	(1)	(2)
Expense Ratio	-616.101***	-743.290***
	(52.096)	(110.898)
Observations	2,020	2,016
R-squared	0.552	0.099
PlanxMarketxYear FE	X	X
IV		X
Elasticity of Demand	3.5	4.2

Note: Table 4 displays the estimates corresponding to a linear regression model. Observations are at the index fund-by-year-by-401(k) plan level. Robust standard errors are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

We separately identify information frictions and preference heterogeneity by exploiting exogenous variation in information frictions in the 401(k) setting. One might be worried that the sample of index funds available in 401(k) may be different from the sample of index fund available more generally, so there is less unobserved heterogeneity in our sample of 401(k) funds. To address this concern, we re-estimate our baseline specification from Section 4.1.2, restricting our data set to fund-year observations that correspond to our 401(k) sample (Appendix Table A9). The results suggest that unobserved heterogeneity is not smaller in our 401(k) sample. An additional concern is that investors may still face some information frictions choosing from a transparent menu of 401(k) funds. To this end, our estimates may be a lower bound on information frictions. We discuss the implications for our counterfactual exercises in Section 5.

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

$$\Omega_{(l,m)}(\mathbf{p}) = \begin{cases} -\frac{1-\phi_T}{1-\beta(1+\tilde{r}_{m(l)})\phi_T}\frac{\partial s_l}{\partial p_m}(\mathbf{p}_t) & \text{if } l, m \in \mathcal{K}_{k,m} \\ 0 & \text{else} \end{cases}$$

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 $\beta \times (1+\tilde{r})$, 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, $\beta \times (1+\tilde{r}) = 0.999$. 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 in the Appendix.[28] To account for outliers, we report the winsorized distribution of marginal costs where we winsorize costs at the 5% level.[29] The mean (median) marginal cost is 29 (14) basis points, and the mean (median) markup is 32 (22) basis points.[30]

5 Counterfactuals

We conduct counterfactual simulations based on our model estimates with two 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. For example, the SEC's recently proposed "Names Rule" aims to foster improved market transparency and mitigate informational frictions due to deceptive or misleading fund names. The effectiveness of such policy would hinge on the magnitude of the effect information frictions have on equilibrium expense ratios.

Second, 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, there remains a question of why substantial price dispersion persists in the market. We examine how demand-side frictions interact with this increased competition.

[28] See Appendix Figure A3. For computational ease, we restrict our attention to those index funds with a market share such that $s_{R,j,t} + s_{I,j,t} \geq 1e-6$.

[29] 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`.

[30] 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 (77%) is unexplained by these fixed effects.

In particular, we ask whether the competitive effect of ETF entry was softened by inertia and information frictions.

For each counterfactual, we first consider a partial equilibrium analysis where we keep fund expense ratios fixed thereby ignoring the potential supply-side response. We then consider a general equilibrium analysis where we allow 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.[31]

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.

First, we consider eliminating inertia in the model such that $\phi_R = \phi_I = 0$. Given that we find only about 3% of retail investors update their portfolio each month, one might expect this to have a large effect on expense ratios. Counterfactual results imply that removing inertia reduces average expense ratios from 34 basis points to 31 basis points for retail investors when considering the partial equilibrium impact holding expense ratios fixed. The counterfactual distribution of expense ratios is shown in Panel (a) of Figure A4 and summarized in the first column of Table 5. The fact that eliminating inertia only reduces expense ratios by 3 basis points is surprising. As discussed further below, removing inertia has a modest effect in part due to the fact that retail investors are not very good at optimally selecting index funds in the first place. In other words, allowing them to select funds more frequently (i.e., removing inertia) does not have a substantial impact on the funds investors choose given the magnitude of information frictions.

When considering the supply-side response, eliminating inertia reduces average expense ratio by roughly 4 basis points from 34 to 30 basis points. This is shown in the third column of Table 5. Recall from Section 3, that if index fund managers use a growth-adjusted discount factor of 1 (i.e., $(1+\tilde{r}_{m(j)})\beta = 1$), 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.

[31] See Appendix C for full details on how we implement each counterfactual.

Table 5: Quantifying Frictions: Mean and Standard Deviation of Expense Ratios

Panel A: Retail Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.34	0.40		
Counterfactuals	Without Supply Response		With Supply Response	
No Inertia	0.31	0.40	0.30	0.40
No Info Frictions	0.25	0.32	0.23	0.38
No Inertia or Info Frictions	0.20	0.25	0.10	0.22
No Px Discrimination			0.26	0.37
No Inertia, Info Frictions, or Px Discrimination			0.10	0.22

Panel B: Institutional Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.27	0.29		
Counterfactuals	Without Supply Response		With Supply Response	
No Inertia	0.24	0.27	0.24	0.27
No Info Frictions	0.21	0.23	0.16	0.26
No Inertia or Info Frictions	0.18	0.20	0.10	0.17
No Px Discrimination			0.20	0.29
No Inertia, Info Frictions, or Px Discrimination			0.09	0.17

Note: Table 5 displays the mean and standard deviation of asset-weighted expense ratios investors pay in each counterfactual.

Overall, our results imply that removing investor inertia lowers average expense ratio by 12% accounting for the supply-side response. Interestingly, there is little effect on price dispersion, as measured by the standard deviation of prices. This is consistent with the idea that some investors are still choosing expensive funds even when they are making an active choice.

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 4).[32] When information frictions are eliminated, we find that the average expense ratio retail investors pay falls by 26% from 34 basis points to 25 basis points in the partial equilibrium; this falls further to 23 basis points after accounting for the supply-side response (Figure A4 Panel (b) and Table 5 Panel A). Thus, removing information frictions has a larger effect than removing

[32] 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. We argue that the 401(k) setting provides a setting with minimal information frictions, providing a benchmark for demand without information frictions. 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.

inertia despite the magnitude of inertia.[33]

Third, we consider the counterfactual distributions of expenses that investors would pay if both inertia and information frictions are simultaneously eliminated, displayed in Panels (c) of Figures A4 and summarized in Table 5. Here, eliminating inertia has a much larger effect on the expense ratios retail investors pay after information frictions are eliminated. For example, consider the partial equilibrium where the supply-side remains fixed. If we eliminate inertia but investors still face information frictions, the average expense ratio retail investors pay falls by a negligible amount from 34 to 31 basis points. In contrast, if we eliminate inertia and investors face no information frictions, the average expense ratio retail investors pay falls to 20 basis points. This result is intuitive: removing inertia and allowing investors to shop for index funds more frequently is more valuable when investors are good at shopping for index funds. This is also reflected in the standard deviation of prices which falls from 40 basis points to 25 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 10 basis points.

Collectively, the results suggest that the high level of inertia may not be as costly as it appears at first glance, given that investors face relatively severe information frictions. The benefits of searching more frequently are especially limited for retail investors because they encounter significant information frictions, which may rationalize their infrequent searching (i.e., high inertia). This may also explain why institutional investors, who face lower information frictions, search more frequently (i.e., have substantially lower inertia).[34]

While we focus on retail investors, we also report the corresponding findings for institutional investors in Figure A5 and Panel B of Table 5. We find qualitatively similar results: eliminating inertia would lower the average expense ratios institutional investors pay by 11%, from 27 basis points to 24 basis points; eliminating information frictions would lower their average ratios by 41%, to 16 basis points; eliminating both would lower their average expense ratios by 63%, to 10 basis points. Quantitatively, these effects are slightly smaller than those of retail investors, consistent with the finding in Section 4 that institutional investors face less frictions in general.

These frictions and their differences between institutional and retail investors have important implications for the supply side. To illustrate this, we consider a counterfactual policy in which price discrimination is banned in the index fund market. In Section 4, we find institutional investors are more active, more price-elastic and face less information frictions than retail investors. As illustrated in Section 2.2, managers' ability to price discriminate and charge

[33]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 A9 and Table A11). Compared to eliminating information frictions entirely, reducing information frictions by half leads to two-thirds of the decrease in expense ratios in the partial equilibrium, and over 70% of the effect in the general equilibrium.

[34]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. For example, we can allow inertia and information frictions to be correlated or assume inertia to be a function of choice set characteristics. Estimating such a model would require richer variation in information frictions, however.

different prices to retail and institutional investors contributes meaningfully to the dispersion in fund expense ratios. We implement this counterfactual by requiring index fund managers to charge the same expense ratio for all funds that share the same underlying portfolio defined by portfolio identifier reported by CRSP and then by solving for a new equilibrium, implying expense ratio are the same for institutional and retail investors buying the same portfolio. The dashed-gray line in Panels (d) of Figures A4 and A5 (summarized in Table 5) indicates that eliminating price discrimination would lower the expenses retail (institutional) investors pay by 24% (26%), to 26 (20) basis points.

To summarize, Figure A7 shows the counterfactual effect of removing inertia, information frictions, and price discrimination sequentially. While removing inertia decreases average retail expense ratios by 11%, removing information frictions decreases average expense ratios by another 58%. The effect of eliminating price discrimination has minimal effect on the expenses retail investors pay once we have already eliminated information frictions and inertia. The reason for this is that price discrimination is not very effective when investors are good at shopping for index funds (i.e., investors do not face information frictions and are never inert).

In Appendix D, we extend the model to explore 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 idea that there is less scope for agency issues in the index fund market compared to other settings analyzed in the literature.

5.2 The Introduction of ETFs

We examine how the introduction of ETFs impacts market competition and whether frictions faced by investors play an important role in shaping 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 A8 and report means and standard deviations in Table 6. The results imply that retail index mutual fund expense ratios decrease by 19% (from 42 to 34 basis points) with the introduction of ETFs. Part of the effect is driven by the cost advantage of ETFs and the other part comes from the competitive effect of ETFs. When ETFs instead have the same asset-weighted average marginal cost as mutual funds in the same lipper class and month, the average retail expense ratio would be 4 basis points lower than in the baseline case without ETFs. This suggests that the competitive effect of ETFs accounts for half of the overall effect for retail funds, while the cost effect accounts for the other half.

Table 6: 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	0.42	0.48	0.34	0.40
No Inertia	0.38	0.48	0.30	0.40
No Info Frictions	0.24	0.42	0.23	0.38
No Inertia or Info Frictions	0.15	0.32	0.10	0.22
No ETF Cost Advantage			0.38	0.49

Panel B: Institutional Investors

	Mean	Std. Dev.	Mean	Std. Dev.
	Without ETF		With ETF	
Baseline	0.35	0.37	0.27	0.29
No Inertia	0.33	0.35	0.24	0.27
No Info Frictions	0.21	0.32	0.16	0.26
No Inertia or Info Frictions	0.17	0.27	0.10	0.17
No ETF Cost Advantage			0.30	0.37

Note: Table 6 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. In order to examine whether frictions dampened the effect of ETF entry, we simulate introducing ETFs while eliminating frictions in Table 6. Introducing ETFs while also removing inertia decreases retail mutual fund expense ratios by 29% (from 42 to 30 basis points). Introducing ETFs while eliminating both inertia and information frictions magnifies the effect, as retail expense ratios decline to 10 basis points. In other words, introducing ETFs would have had led to a reduction in expense ratios about 4 times as large if frictions were also removed. The results highlight how frictions dampen the competitive effects of new products.

6 Conclusion

We quantify the underlying frictions in the index fund market and show how they support an equilibrium with substantial market power. We develop a model in which investors have inertia, are subject to information frictions, and have heterogeneous preferences. The model provides sharp insights into how each friction impacts both demand for and the supply of index funds. Using a novel instrumental variables strategy based on historical returns, we show how we can separately identify inertia from investor preferences and show how data from 401(k) choices can be used to separately identify preference heterogeneity from information frictions.

Our estimates imply that both inertia and information frictions give firms significant market power and play a major role in explaining the observed price dispersion in the index fund market. Despite the fact that a large fraction of investors are inactive, removing information frictions is more important than removing inertia in counterfactual simulations. Removing information frictions decreases the average expense ratios paid by retail investors by roughly 32%, from 34 to 23 basis points. These information frictions have distinct implications from preference heterogeneity given that the presence of information frictions implies that investors are not obtaining welfare gains from variety. This suggests that disclosure policies, rule-making that reduces misleading practices, or further development of comparison tools such as FINRA's Fund Analyzer, could lead to a meaningful reduction in market power and increase welfare. In contrast, we estimate that average expense ratios would decrease by 12%, from 34 to 30 basis points, if inertia were eliminated. Interestingly the effects of inertia and information frictions are interrelated. Inertia becomes much more costly for investors when information frictions are reduced, suggesting that it may be beneficial for policy makers to focus on information frictions first.

Supply-side factors also play a significant role, albeit secondary to demand-side frictions. Our analysis reveals that mutual fund providers capitalize on inertia and information frictions to engage in price discrimination 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. We also find that the introduction of ETFs produced two pivotal impacts on the market: firstly, they brought in a new low-cost technology; and secondly, they curtailed the market power of existing mutual funds. The impact of ETFs would have been much larger had there been no inertia and information frictions.

Overall, our results provide new insight into why investors purchase expensive index fund and why the entry of new funds did not eliminate price dispersion. Many markets likely present similar issues and the results highlight the importance of identifying how frictions interact and the underlying 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 (2023). Endogenous information and simplifying insurance choice. Technical report.

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*.

A Additional Figures and Tables

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

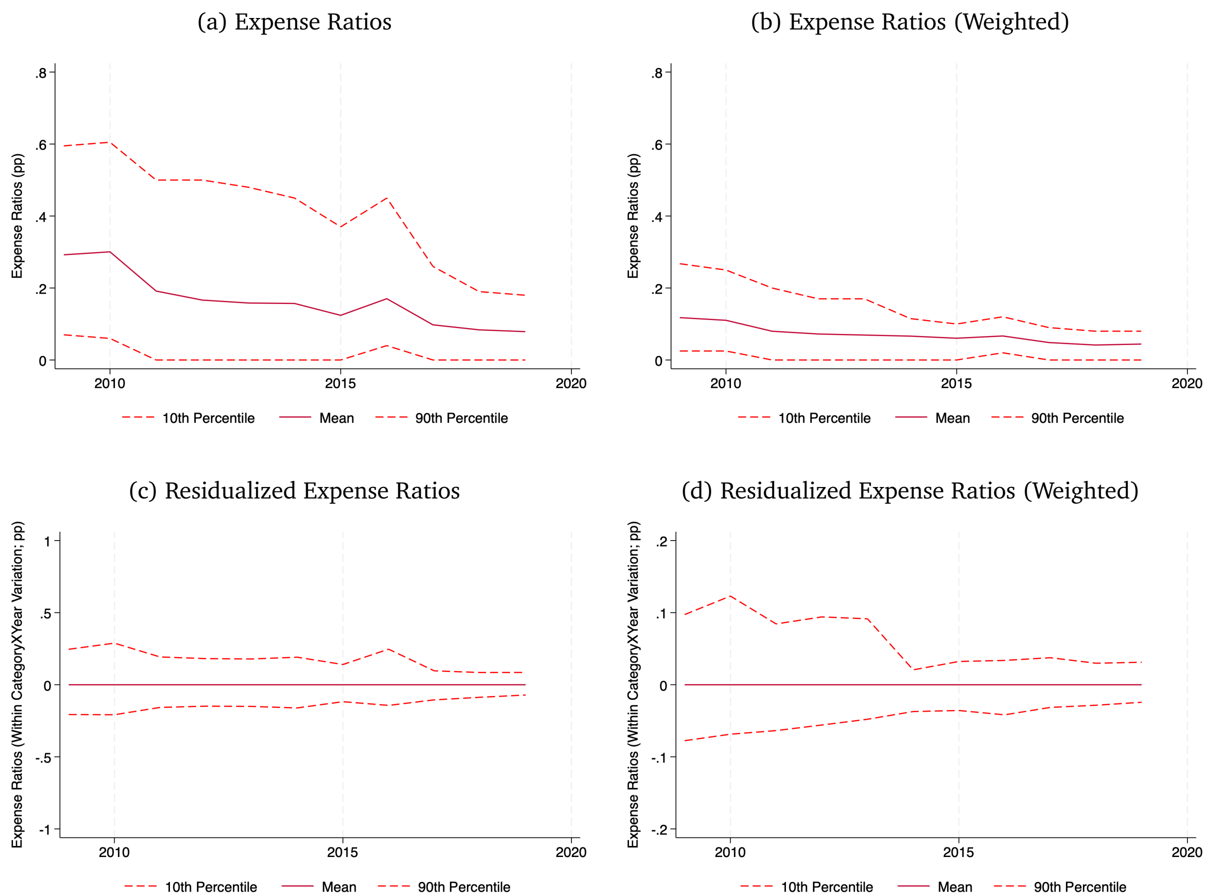
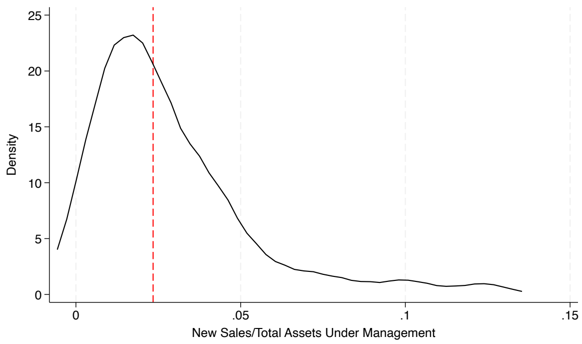


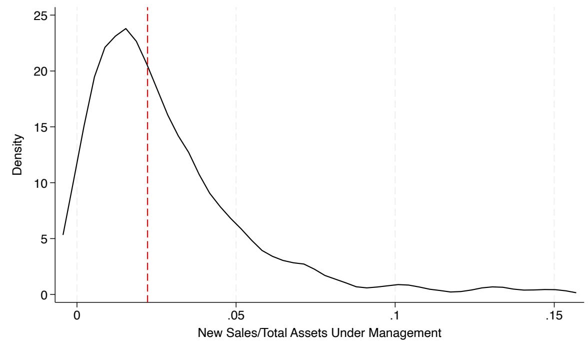
Figure A1 displays the distribution of index fund expense ratios in 401(k) plans over time. Panels (a) and (b) display the equal weighted and asset-weighted distribution of expense ratios. Panels (c) and (d) display the equal weighted and asset-weighted distribution of residualized expense ratios, where 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.

Figure A2: Mutual Fund Sales and Redemptions

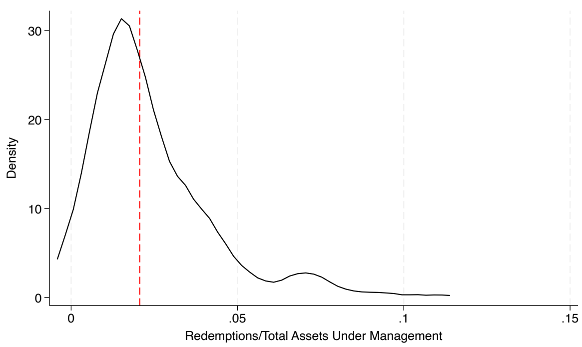
(a) New Shares/AUM (Retail Investors)



(b) New Shares/AUM (Institutional Investors)



(c) Redemptions/AUM (Retail Investors)



(d) Redemptions/AUM (Institutional Investors)

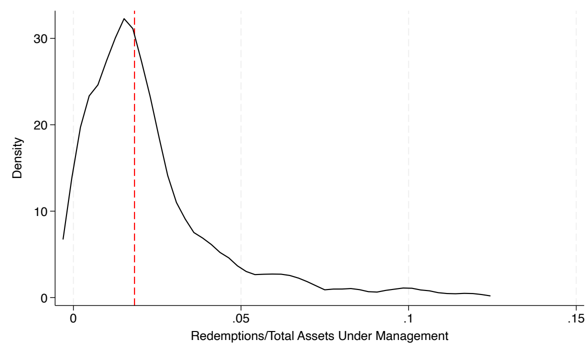


Figure A2a and Figure A2b display the distribution of the total net asset value of new shares purchased relative to total AUM calculated at the market-by-month level over the period 2019-2020 for retail and institutional investors. To account for outliers we restrict the data set to those observations with positive sales, and we censor the distribution at the 95% level. Figure A2c and Figure A2d display the distribution of the total net asset value of shares redeemed relative to total AUM calculated at the market-by-month level over the period 2019-2020. To account for outliers we restrict the data set to those observations with positive redemptions, and we censor the distribution at the 95% level. Data are from Morningstar. The red dashed lined in each figure corresponds to the median observation.

Figure A3: Estimated Marginal Costs and Markups

(a) Distribution of Marginal Costs (b) Distribution of Markups

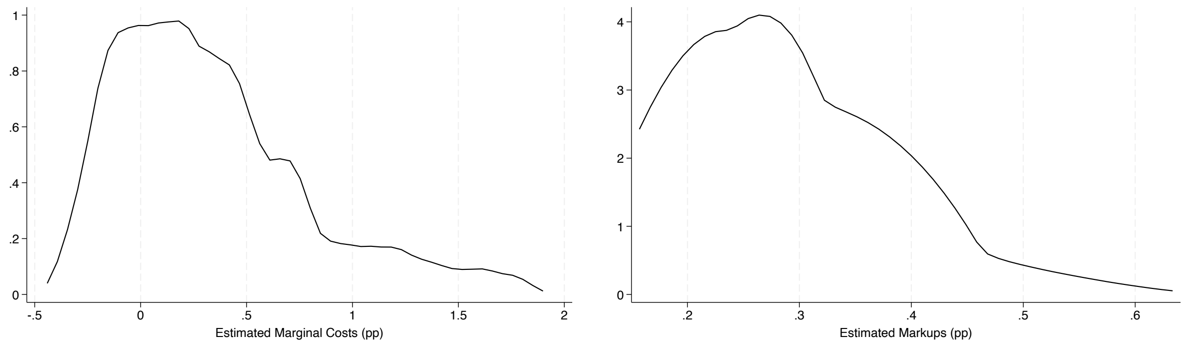


Figure A3 displays the estimated equal-weighted distributions of marginal costs and markups. To account for outliers, both distributions are censored at the 5% and 95% level. Panel (a) displays the density of marginal costs, and panel (b) displays the density of markups.

Figure A4: Counterfactuals: Retail Investors

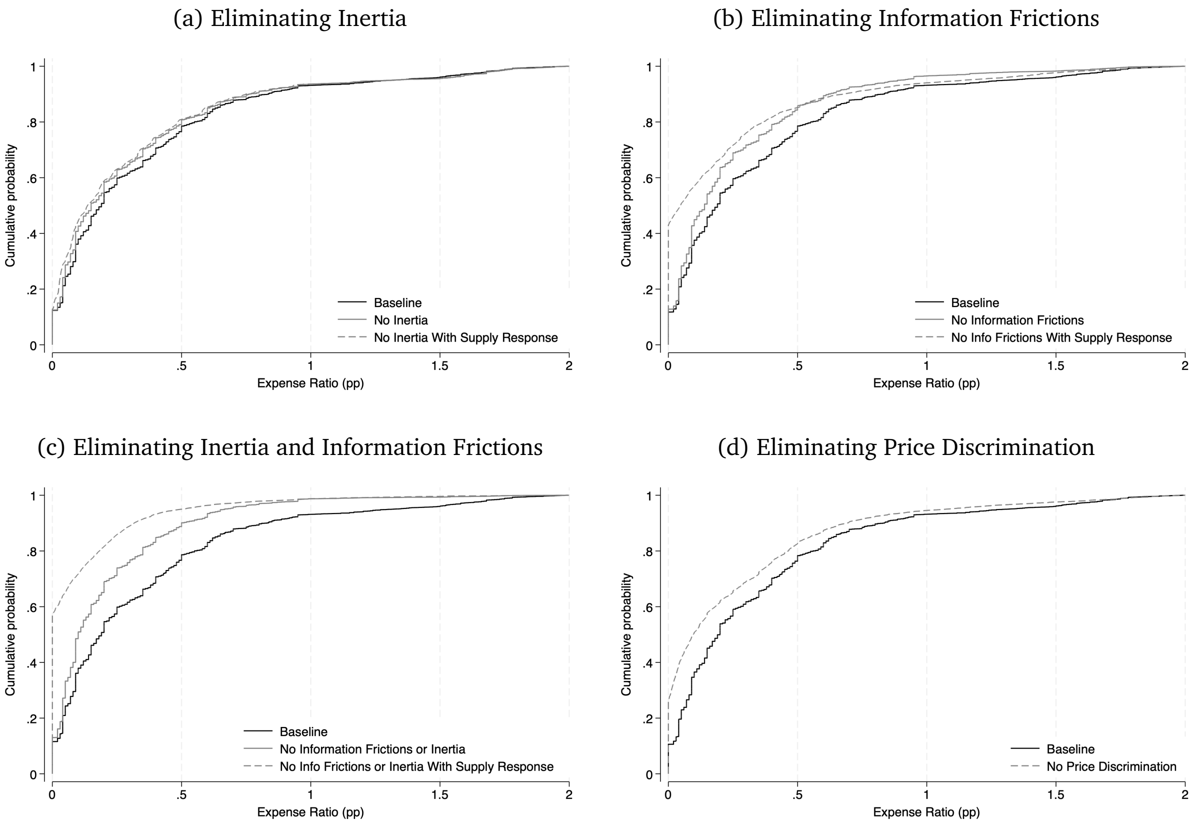


Figure A4 displays the estimated distribution of asset-weighted expense ratios in counterfactual analysis where we eliminate inertia, information frictions and price discrimination.

Figure A5: Counterfactuals: Institutional Investors

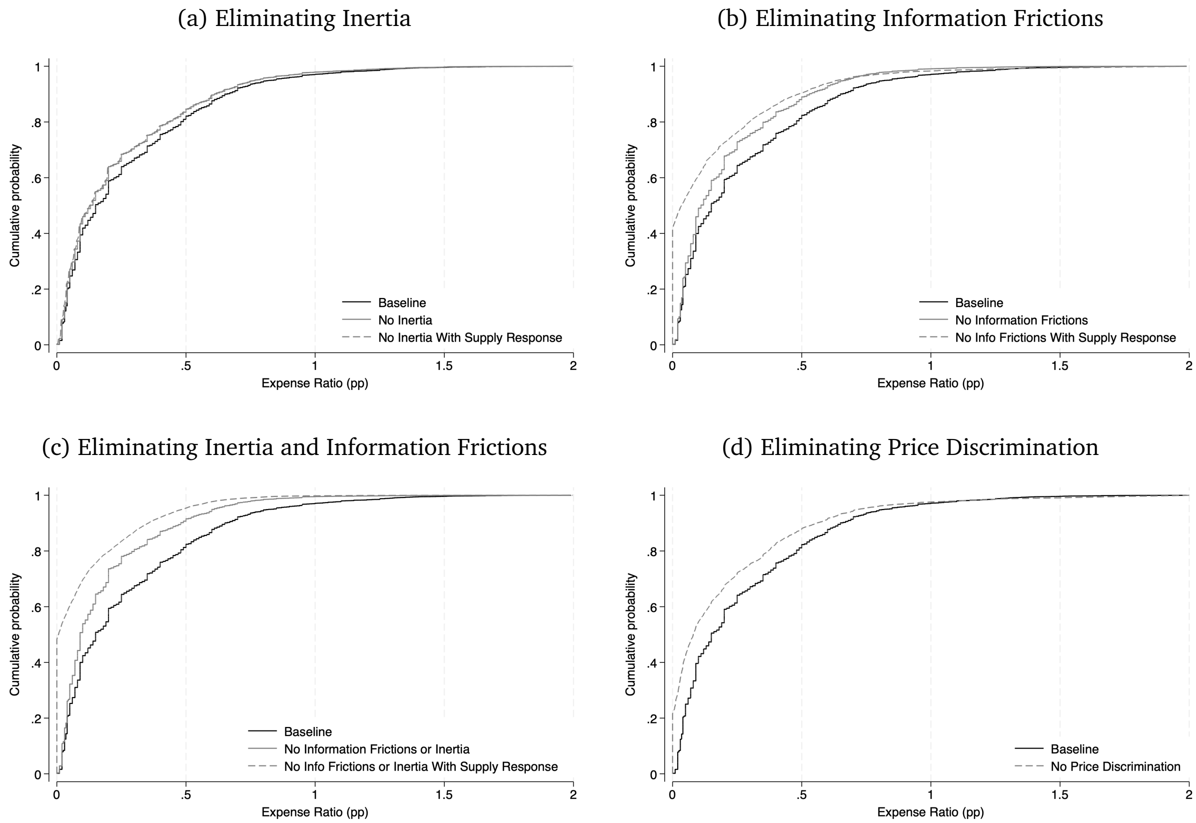


Figure A5 displays the estimated distribution of asset-weighted expense ratios in counterfactual analysis where we eliminate inertia, information frictions and price discrimination.

Figure A6: Counterfactuals: Remove All Frictions

(a) Retail Investors (b) Institutional Investors

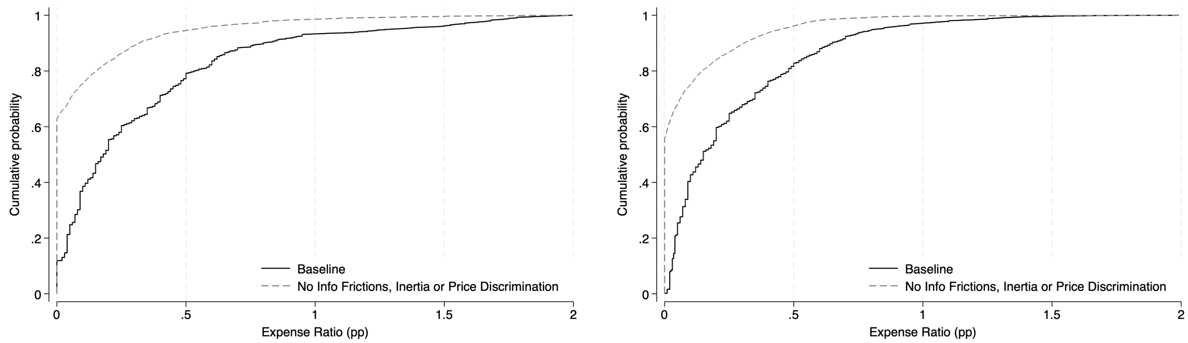


Figure A6 displays the estimated distribution of asset-weighted expense ratios in counterfactual analysis where we eliminate inertia, information frictions and price discrimination.

Figure A7: Sequential Decomposition by Mechanism

(a) Retail Investors (b) Institutional Investors

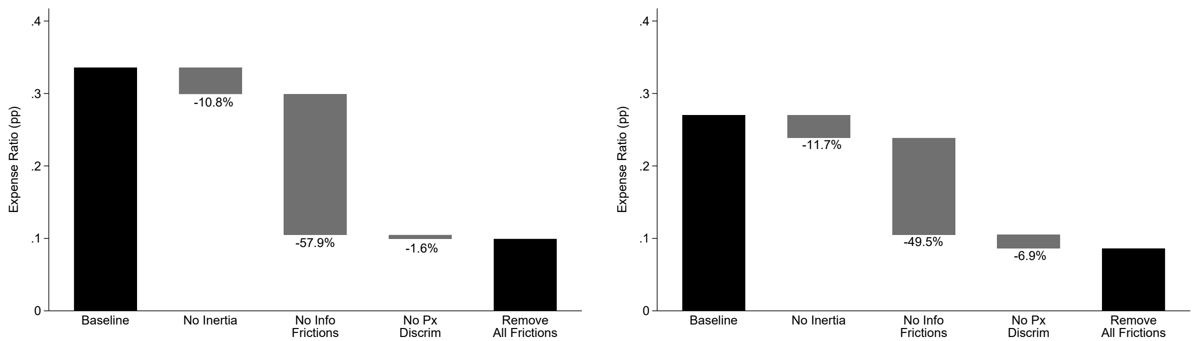


Figure A7 displays the mean asset-weighted expense ratios investors pay after sequentially removing inertia, information frictions, and price discrimination. All counterfactuals account for supply response.

Figure A8: Counterfactuals: The Introduction of ETFs

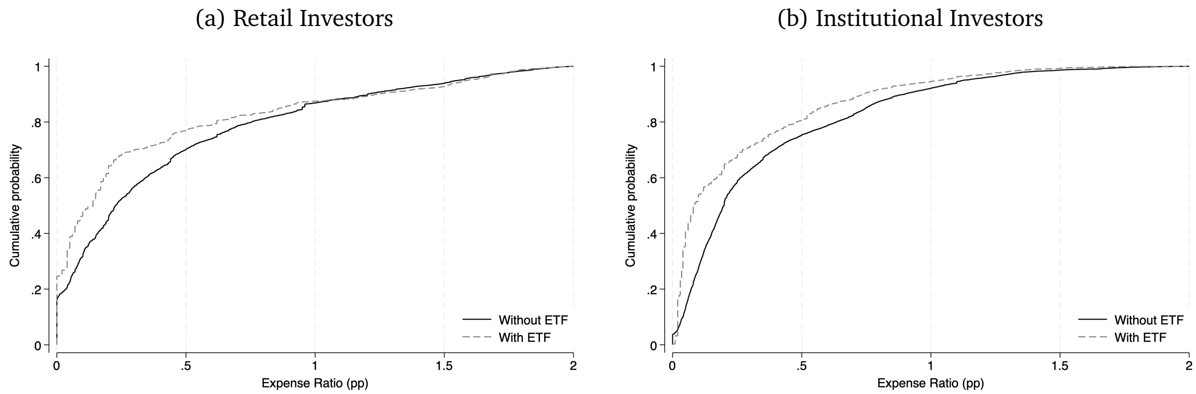
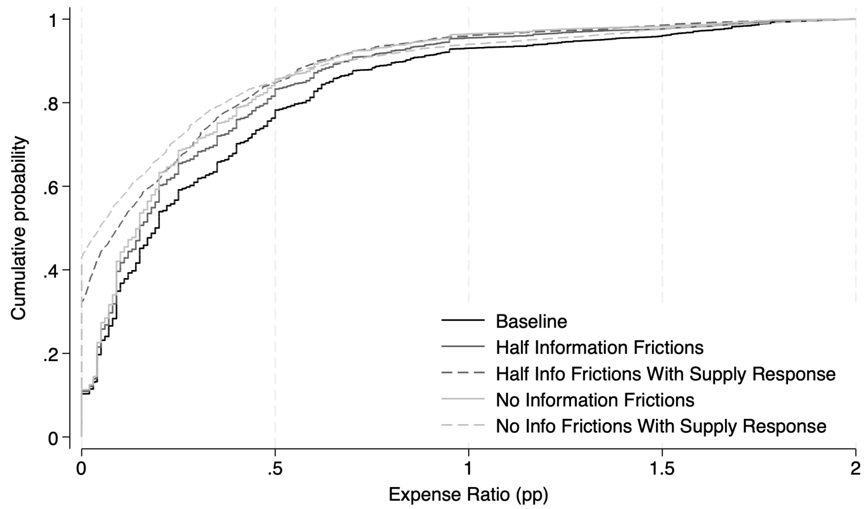


Figure A4 displays the estimated distribution of mutual fund expense ratios in counterfactual analysis where we eliminate ETFs.

Figure A9: Counterfactuals: Reducing Information Frictions

(a) Retail Investors



(b) Institutional Investors

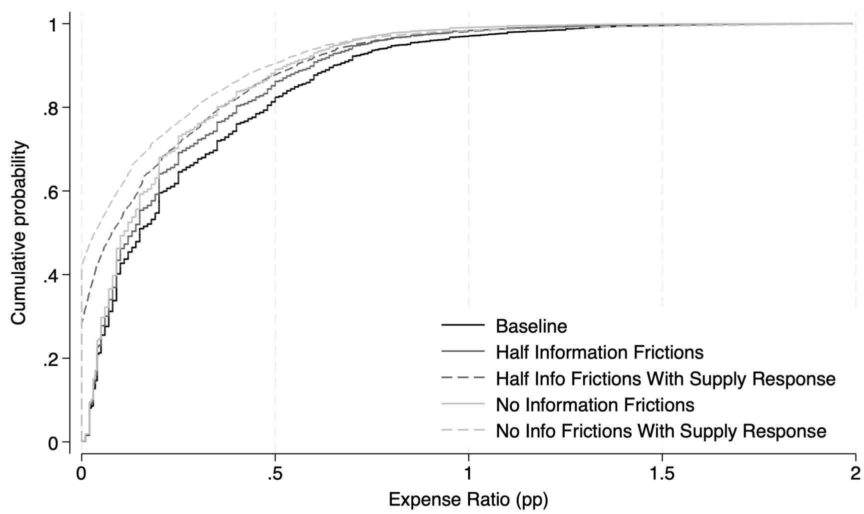
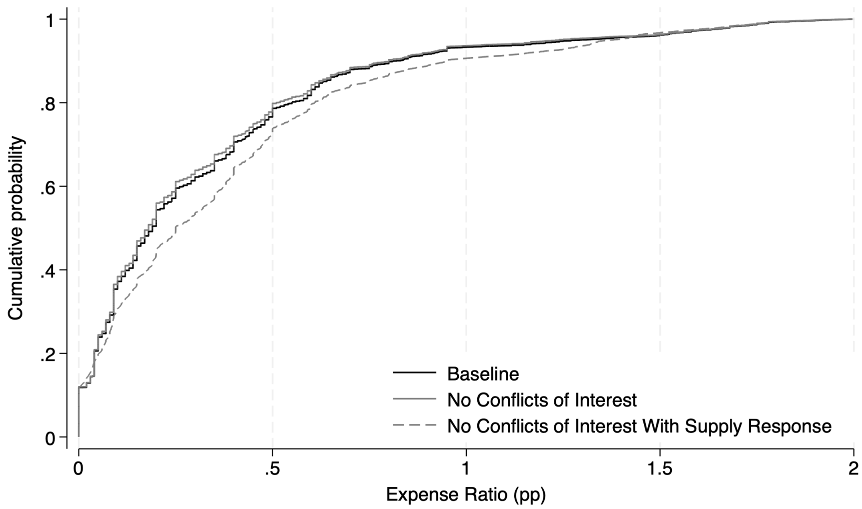


Figure A9 displays the estimated distribution of expense ratios in counterfactual analysis where we reduce information frictions first by half and then entirely.

Figure A10: Counterfactuals: Eliminating Conflicts of Interest

(a) Retail Investors



(b) Institutional Investors

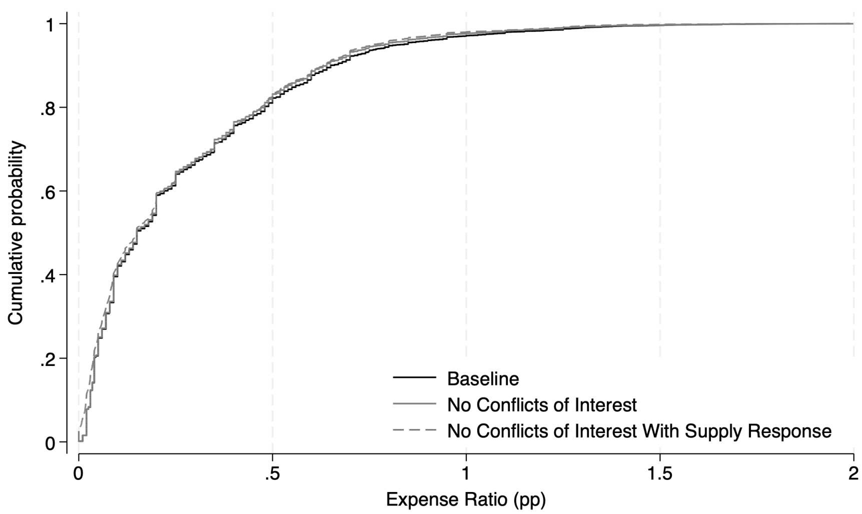


Figure A10 displays the estimated distribution of expense ratios in counterfactual analysis where we eliminate conflicts of interest.

Table A1: Investor Inertia - Estimation in Levels

(a) Retail Investors

VARIABLES	(1)	(2)	(3)
Lag AUM	0.972***	0.978***	0.978***
	(0.002)	(0.003)	(0.003)
Observations	371,660	357,129	355,111
R-squared	0.998	0.998	0.998
IV		X	X
Year-Month FE	X	X	
Year-Month-Mkt FE			X

(b) Institutional Investors

VARIABLES	(1)	(2)	(3)
Lag AUM	0.987***	0.991***	0.991***
	(0.001)	(0.002)	(0.002)
Observations	324,099	313,415	311,598
R-squared	0.998	0.998	0.998
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. 9) that we estimate in levels rather than logs. Observations are at the index fund-by-month level. The dependent variable is assets under management. The independent variable of interest is $AUM_{j,T,t-1}(1+r_{j,t})(1+g)$, where $r_{j,t}$ reflects the monthly return of the fund and g is the average monthly growth rate of AUM held in index funds. 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)-(4) 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)-(3) we control for 1-, 3-, 6-, 12-month, and year-to-date cumulative returns. In columns (4), 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.

Table A2: Investor Inertia - Controlling for New Sales

VARIABLES	(1)	(2)	(3)	(4)
Lag AUM	0.972***	0.953***	0.985***	0.979***
	(0.002)	(0.003)	(0.001)	(0.002)
Observations	19,207	18,950	19,845	19,614
R-squared	0.985	0.987	0.991	0.992
Retail Sample	X	X		
Inst. Sample			X	X
Control Function	X	X	X	X
Year-Month FE	X		X	
Year-Month-Mkt FE		X		X

Note: Table A2 displays the estimates corresponding to a linear regression model (Eq. 9). Observations are at the index fund-by-month level, where we restrict our attention to ETFs. The dependent variable is log assets under management. As described in the text, when estimating inertia, we address the endogeneity of *Lag AUM* by controlling for market shares of the funds within new mutual fund sales, which helps control for product/investment quality. Specifically, we control for up to 4-th order polynomials of log market shares within new mutual fund sales. 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; whether the fund is an ETF, has a front load, or has a rear load; and 1-, 3-, 6-, 12-month, and year-to-date cumulative returns. Robust standard errors are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A3: Investor Inertia Heterogeneity by Price Dispersion

VARIABLES	(1)	(2)	(3)	(4)
Lag AUM	1.012***	0.949***	0.963***	0.947***
	(0.016)	(0.030)	(0.020)	(0.016)
Observations	169,372	173,035	124,280	196,028
R-squared	0.984	0.981	0.984	0.986
Retail Sample	X	X		
Inst. Sample			X	X
IV	X	X	X	X
Year-Month FE	X	X	X	X
Price Dispersion	High	Low	High	Low

Note: Table A3 displays the estimates corresponding to a linear regression model (Eq. 9) by above and below median price dispersion in the market. Observations are at the index fund-by-month level. The dependent variable is log assets under management. As described in the text, we address the endogeneity of *Lag AUM* using an instrumental variable approach. 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; whether the fund is an ETF, has a front load, or has a rear load; and 1-, 3-, 6-, 12-month, and year-to-date cumulative returns. Robust standard errors are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A4: Investor Inertia Heterogeneity by Load Type

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Lag AUM	0.971***	0.966***	1.042***	0.943***	0.941***	1.010***
	(0.017)	(0.017)	(0.017)	(0.013)	(0.013)	(0.010)
Lag AUM x Has Front Load	0.011			-0.038		
	(0.010)			(0.090)		
Lag AUM x Has Rear Load		0.020**			-0.006	
		(0.008)			(0.010)	
Lag AUM x 1 Year Return			0.041***			0.026***
			(0.015)			(0.010)
Observations	342,407	342,407	342,407	320,308	320,308	320,308
R-squared	0.983	0.983	0.981	0.984	0.984	0.986
Retail Sample	X	X	X			
Inst. Sample				X	X	X
IV	X	X	X	X	X	X
Year-Month FE	X	X	X	X	X	X

Note: Table A4 displays the estimates corresponding to a linear regression model (Eq. 9). Observations are at the index fund-by-month level. The dependent variable is log assets under management. As described in the text, we address the endogeneity of *Lag AUM* using an instrumental variable approach. 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; whether the fund is an ETF, has a front load, or has a rear load; and 1-, 3-, 6-, 12-month, and year-to-date cumulative returns. Robust standard errors are in parenthesis. *** p<0.01, ** p<0.05, * p<0.10.

Table A5: Estimated Investor Preferences Using New Sales Data

VARIABLES	(1)	(2)	(3)	(4)
Expense Ratio	-255.259***	-460.942***	-433.055***	-1,007.444***
	(4.374)	(38.056)	(9.411)	(83.907)
Observations	8,141	3,841	8,317	6,253
R-squared	0.552	0.449	0.402	0.034
Year-Month-Mkt FE	X	X	X	X
IV		X		X
Retail Sample	X	X		
Inst. Sample			X	X
Elasticity of Demand	1.4	2.6	2.4	5.6

Note: Table A5 displays the estimates corresponding to a linear regression model (Eq. 11), where we compute market shares using new sales. 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 A6: Robustness of Investor Preferences Including Spread Instrument

VARIABLES	(1)	(2)	(3)	(4)
Expense Ratio	-332.128***	-245.194***	-565.372***	-502.266***
	(15.182)	(18.705)	(15.453)	(18.290)
Observations	55,248	55,248	62,578	62,578
R-squared	0.074	0.081	0.115	0.141
Year-Month-Mkt FE	X	X	X	X
Other Firm Assets		X		X
IV	X	X	X	X
Retail Sample	X	X		
Inst. Sample			X	X
Elasticity of Demand	1.9	1.4	3.2	2.8

Note: Table A6 displays the estimates corresponding to an instrumental variable regression model (Eq. 11). In all specifications, we use 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-, 12-month, and year-to-date cumulative returns; and whether the fund is an ETF, has a front load, or has a rear load. In columns (2) and (4) we also control for other firm assets since this may affect trading cost. Robust standard errors are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A7: Robustness of Investor Preferences Including Markup Instrument

VARIABLES	(1)	(2)	(3)	(4)
Expense Ratio	-228.867***	-198.075***	-405.809***	-384.147***
	(5.802)	(6.240)	(6.730)	(6.876)
Observations	320,508	320,508	302,861	302,861
R-squared	0.039	0.043	0.097	0.107
Year-Month-Mkt FE	X	X	X	X
Other Firm Assets		X		X
IV	X	X	X	X
Retail Sample	X	X		
Inst. Sample			X	X
Elasticity of Demand	1.3	1.1	2.3	2.2

Note: Table A7 displays the estimates corresponding to an instrumental variable regression model (Eq. 11). In all specifications, 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 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. In columns (2) and (4) we also control for other firm assets since this may affect trading cost. Robust standard errors are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A8: Robustness of Investor Preferences including Top 3 Firm Indicator

VARIABLES	(1)	(2)	(3)	(4)
Expense Ratio	-187.946***	-231.891***	-278.471***	-496.956***
	(2.325)	(10.413)	(2.412)	(10.547)
Top 3 Firm	1.111***	1.326***	0.535***	-0.133***
	(0.037)	(0.072)	(0.017)	(0.037)
Observations	346,817	128,532	322,102	133,530
R-squared	0.155	0.077	0.258	0.137
Year-Month-Mkt FE	X	X	X	X
IV		X		X
Retail Sample	X	X		
Inst. Sample			X	X
Elasticity of Demand	1.1	1.3	1.6	2.8

Note: Table A8 displays the estimates corresponding to a linear regression model (Eq. 11) including an indicator for whether the fund manager is one of the top three firms measured by total assets (BlackRock, State Street Bank, and Vanguard). Observations are at the index fund-by-month-by-investor type (i.e., retail vs. institutional) level. In columns (2) and (4) we use the Hausman instrument as an instrument for expense ratios. 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; whether the fund is an ETF, has a front load, or has a rear load; and the top 3 indicator. Robust standard errors are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A9: Estimated Investor Preferences Using the Sample of Funds in the 401(k) Data

VARIABLES	(1)	(2)	(3)	(4)
Expense Ratio	-289.622***	-267.214	-440.183***	-199.519*
	(39.996)	(174.003)	(26.158)	(119.357)
Observations	3,909	2,661	2,624	1,697
R-squared	0.370	0.163	0.607	0.386
Year-Month-Mkt FE	X	X	X	X
IV		X		X
Retail Sample	X	X		
Inst. Sample			X	X
Elasticity of Demand	1.6	1.5	2.5	1.1

Note: Table A9 displays the estimates corresponding to a linear regression model (Eq. 11), where we restrict the sample of fund-year observations to be the same as in our 401(k) sample. 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 A10: Active Demand for Index Funds: Accounting for Brokers

VARIABLES	(1)	(2)	(3)	(4)
Expense Ratio	-360.407***	-330.318***	-522.260***	-514.971***
	(14.098)	(15.236)	(10.282)	(10.660)
12b-1 Fees	189.212***	122.686***	348.209***	227.531***
	(16.566)	(19.940)	(16.226)	(20.150)
Observations	128,532	125,315	133,530	130,876
R-squared	0.071	0.068	0.148	0.150
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	1.9	2.9	2.9
ω	0.34	0.27	0.40	0.31

Note: Table A10 displays the estimates corresponding to a linear regression model (Eq. 15). 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 A11: Reducing Information Frictions: Mean and Standard Deviation of Expense Ratios

Panel A: Retail Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.34	0.40		
Counterfactuals	Without Supply Response		With Supply Response	
Half Info Frictions	0.28	0.35	0.23	0.33
No Info Frictions	0.25	0.32	0.23	0.38

Panel B: Institutional Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.27	0.29		
Counterfactuals	Without Supply Response		With Supply Response	
Half Info Frictions	0.23	0.25	0.19	0.26
No Info Frictions	0.21	0.23	0.16	0.26

Note: Table A11 displays the mean and standard deviation of asset-weighted expense ratios investors pay in each counterfactual.

Table A12: Eliminating Conflicts of Interest: Mean and Standard Deviation of Expense Ratios

Panel A: Retail Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.34	0.40		
Counterfactuals	Without Supply Response		With Supply Response	
No Conflicts of Interest	0.32	0.39	0.39	0.42

Panel B: Institutional Investors

	Mean	Std. Dev.	Mean	Std. Dev.
Baseline	0.27	0.29		
Counterfactuals	Without Supply Response		With Supply Response	
No Conflicts of Interest	0.26	0.27	0.25	0.27

Note: Table A12 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 with 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, one can 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 to reduce 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.[35]

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

C 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:

$$\mathcal{S}_{T,j,t}(\phi, \mathbf{p}, \sigma_\nu) = \sum_{\tau=0}^{\infty}(1-\phi)\phi^\tau s_{T,j,t-\tau}(\mathbf{p}_{t-\tau}, \sigma_\nu). \tag{13}$$

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. (13) 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.

D 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.

D.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:

$$s_{T,j,t} = \frac{\exp\left(\frac{\frac{\omega_T}{1-\omega_T} f_{j,t} - p_{j,t} + X'_{j,t}\theta_T + \xi_{j,T(i),t}}{\sigma_{\eta,T(i)}}\right)}{\sum_{l \in \mathcal{J}_{T,m(j),t}} \exp\left(\frac{\frac{\omega_T}{1-\omega_T} f_{l,t} - p_{l,t} + X'_{l,t}\theta_T + \xi_{l,T(i),t}}{\sigma_{\eta,T(i)}}\right)}, \tag{14}$$

which is the core of our estimation strategy.

D.2 Estimation

We estimate Eq. (14) 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):

$$\ln s_{T,j,t} = \underbrace{\varpi_T}_{\frac{\omega_T}{\sigma_{\eta,T}(1-\omega_T)}} f_{jt} - \underbrace{\alpha_T}_{\frac{1}{\sigma_{\eta,T}}} p_{jt} - X'_{j,t}\underbrace{\Gamma_T}_{\frac{\theta_{T(i)}}{\sigma_{\eta,T}}} + \underbrace{\mu_{T,m(j),t}}_{\ln\left(\sum_{l \in \mathcal{J}_{T,m(j),t}} \exp\left(\frac{\frac{\omega_T}{1-\omega_T} f_{l,t} - p_{l,t} + X'_{l,t}\theta_T + \xi_{l,T(i),t}}{\sigma_{\eta,T(i)}}\right)\right)} + \underbrace{\zeta_{T,j,t}}_{\frac{\xi_{T,j,t}}{\sigma_{\eta,T}}}. \tag{15}$$

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.[36] Because brokers are also compensated with front and rear loads, we calculate load-adjusted 12b-1 fees where we add 1/3rd of total loads to the 12b-1 fees.

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.

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

We report our corresponding estimates in Table A10. 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 1.9 to 2.0 for retail investors and 2.9 for institutional investors, which are consistent with our baseline demand estimates (Table 3). 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 (1) 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 0.52 ($= 0.34/(1-0.34)$) percentage point increase in expense ratios. In other words, the estimates suggest that brokers place roughly 2 times ($= (1-0.34)/0.34$) 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 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 A10 and Table A12). If we keep the product space fixed, eliminating conflicts of interest would reduce the expense ratios that retail and institutional investors pay by 2 and 1 basis point, respectively. The small size of the effects can be explained by the fact that the majority of the index funds in our sample (roughly 75%) do not pay 12b-1 fees.