Who Stays, Who Switches: Tax Frictions and the Shift from Mutual Funds

Jessica Goldenring*

March 16, 2026

Link to Latest Version

Abstract

This paper shows that unrealized capital gains generate both persistence and fragility in mutual funds. Using newly compiled fund-level data on unrealized gains from SEC filings, I document how investors decide to stay in a mutual fund or switch to lower-fee, tax-efficient vehicles such as separately managed accounts and exchange-traded funds. Unrealized gains create persistence by locking investors into funds despite higher fees and tax disadvantages, as switching triggers immediate capital gains taxes. I show unrealized gains reduce outflows and enable managers to charge higher fees. However, unrealized gains also create fragility through capital gains distributions: when some investors redeem shares, and the fund liquidates appreciated securities, these realized gains are passed through as taxable income to remaining investors. This creates strategic complementarities in which one investor's redemption and switching decision affects others' incentives to stay or switch. I rationalize these dynamics through a bank-run model with heterogeneous unrealized gains and endogenous distributions that generates multiple equilibria under identical fundamentals that creates financial fragility.

*Columbia Business School; jessica.goldenring@columbia.edu.
I am grateful to my dissertation committee: Kent Daniel, Yiming Ma, Tano Santos, and Jesse Schreger for their continuous guidance and support. I thank Boaz Abramson, Olivier Darmouni, Xavier Giroud, Jane Li, Xuelin Li, Federico Mainardi, Harry Mamaysky, Simon Oh, Tomasz Piskorski, Pari Sastry, Suresh Sundaresan, Dominik Supera, Stijn Van Nieuwerburgh, Laura Veldkamp, and Kairong Xiao for helpful comments. Finally, I am appreciative of all the support from my fellow PhD students at CBS. I acknowledge financial support from the Center of Political Economy Grant.

1 Introduction

Mutual funds (MFs) have long been the dominant investment vehicle through which households obtain diversified exposure to financial assets. However, the past decades have witnessed the rise of alternative investment vehicles that now rival in scale the over \$10 trillion US equity MF industry. In particular, Exchange-Traded Funds (ETFs) have grown to manage over \$5.7 trillion in assets, while separately managed accounts (SMAs) have expanded to \$7.1 trillion and have become increasingly accessible as minimum investment requirements have declined.

A leading explanation behind the rise of these alternative investment vehicles is the advantageous tax treatment they offer relative to MFs. The structure of MFs creates a tax inefficiency: capital gains distributions. Capital gains distributions occur when funds sell appreciated assets for cash. MFs function as pass-through entities, meaning realized gains are distributed to shareholders as taxable income. When some investors redeem their MF shares, the fund might sell holdings to meet those redemptions. If these holdings have appreciated, the realized gains are passed through pro-rata to remaining investors, creating tax liabilities for shareholders who maintain their positions. ETFs largely avoid this problem by using in-kind transactions rather than cash sales to meet redemptions.[1] In a SMA, the manager replicates a portfolio across investors in legally separate accounts, isolating investors from the redemption decisions of other investors in the same strategy.

The structural tax disadvantage of MFs compared to these alternative investment vehicles raises a central puzzle: why do investors remain in MFs? This puzzle deepens given that MFs also charge higher fees than comparable ETFs and SMAs. In this paper, I show that unrealized gains, the appreciation in value of an investment that has not yet been sold and taxed, are key to understanding why MFs persist despite their disadvantages. The unrealized gain mechanism proposed here operates through two channels: lock-in and fragility. Each channel corresponds to a distinct tax consideration that shapes investors' decision to stay or switch.

Consider the lock-in channel first. On one side, unrealized gains in investors' MF shares create switching costs that encourage staying. When investor MF shares have appreciated significantly, exiting triggers an immediate tax liability on those gains. The median equity MF has doubled in value over the past 15 years, meaning that to switch to an SMA or ETF, investors

[1]ETFs have two markets: individual investors trade shares on exchanges (secondary market), while authorized participants create or redeem shares with the fund (primary market) by exchanging securities instead of cash. This setup limits taxable distributions because trades occur between investors and redemptions are done in-kind, avoiding asset sales.

must sell their MF shares and realize substantial gains, paying taxes immediately. Those with large unrealized gains face high exit taxes, creating tax lock-in that keeps investors from leaving despite the fee and tax disadvantages of the MF structure.

On the other side, unrealized gains in the MF's underlying portfolio impose costs on staying through capital gains distributions, encouraging investors to switch. This mechanism is the source of MF fragility. Because MFs are pass-through entities, when the MF sells appreciated holdings to meet redemptions, they typically distribute the realized capital gains pro-rata to all remaining shareholders as taxable income. This creates strategic complementarities: each investor's optimal decision to stay or switch depends on what other investors do. When some investors switch out, their outflows force asset sales, and if the fund sells appreciated holdings, the resulting distributions impose tax liabilities on those who remain, making switching more attractive for them as well. High fund-level unrealized gains thus play a dual role. While they create stability by deterring switching through high exit taxes, they also amplify fragility when outflows occur by generating large distributions that can trigger further switches in a self-reinforcing cycle.

This lock-in channel creates a form of customer captivity, and thus MF pricing power. When investors face high switching costs due to their unrealized gains, funds can charge higher fees without triggering exits. Consistent with this, I find that funds with higher unrealized gains charge significantly higher fees, controlling for performance and fund characteristics. This provides a direct link between tax lock-in and the fee disadvantage of MFs relative to ETFs and SMAs: the same unrealized gains that keep investors locked in enable managers to extract rents through higher fees.

This dual nature of unrealized gains is illustrated in 2020 with Vanguard's Target Retirement Funds. Prior to 2020, Vanguard operated the same target-date strategies through two separate fund structures with different fee schedules and minimums: investor share funds with higher fees but lower investment minimums (0.14% expense ratio, $1,000 minimum) and institutional funds with lower fees but higher investment minimums (0.09% expense ratio, $100 million minimum). Note that these funds were invested in by both taxable and non-taxable investors. When Vanguard reduced the institutional funds' minimums to $5 million in 2020, it enabled smaller investors to access the lower-fee version. Many investors switched from the higher-fee investor funds to the lower-fee institutional funds, triggering substantial redemptions. These asset sales realized substantial gains that were distributed pro-rata to all remaining shareholders who stayed in the investor funds. Distribution yields reached up to 10% of Net Asset Value (NAV) in 2021, creating large tax liabilities for taxable investors who chose to stay. The episode illustrates the strategic complementarities at the heart of the stay-or-switch decision. The SEC later determined that Vanguard had failed to adequately disclose these potential tax consequences and brought an enforcement action in 2025 (SEC Release No. 33-11359).

The tension illustrated in the Vanguard case applies broadly: many MFs have twin offerings in ETFs or SMAs that provide similar portfolio exposure at a lower cost and with greater tax efficiency. A large share of MFs are paired with an ETF or SMA twin, allowing investors to hold the same portfolio through a different vehicle structure. The widespread availability of these lower-cost, tax-efficient alternatives intensifies the question of MF persistence.

A central challenge to this question is that unrealized gains at the investor level are not directly observed. I address this limitation by constructing fund-level measures of unrealized gains from SEC regulatory filings. While fund-level unrealized gains cannot perfectly measure individual investor tax lock-in, they provide a natural proxy because funds with appreciated portfolios have generally delivered returns that create unrealized gains for their shareholders.

The first contribution of this paper is constructing a dataset of fund-level unrealized gains using a language model approach to extract data at scale. MFs disclose aggregate unrealized gains in annual shareholder reports (Form N-CSR), but these disclosures are embedded in unstructured text that varies in format across funds and years, making traditional parsing methods impractical. Large language models enable systematic extraction of this information across thousands of fund-year observations. I merge these annual measures with monthly changes from Form N-PORT to create high-frequency measures from 2019 to 2024. Importantly, the N-PORT data serve as an internal validation of the LLM extraction: the monthly flow measures from N-PORT cumulate to match the annual stock measures extracted from N-CSR, confirming the accuracy of the language model approach across funds and years. This dataset captures both the extent to which fund assets have appreciated relative to their cost basis and the potential taxable gains that would be realized if holdings were sold. The data show substantial variation in unrealized gains across funds and over time, and this heterogeneity is central to understanding the trade-off between remaining in an MF and switching to a lower fee and more tax-efficient vehicle. Specifically, I find that the average MF has about 25% unrealized gains, as a proportion of AUM, with a standard deviation of 19% from 2019 to 2024. Critically, portfolio holdings data alone cannot reveal the unrealized gains of an MF because cost bases are not disclosed. The data therefore illuminate an additional dimension of the MF vehicle beyond traditional holdings analysis.

Using this data, the paper next establishes three facts about the role of unrealized gains. First, funds with higher unrealized gains experience significantly lower outflows showing the lock-in of MF investors. A one standard deviation increase in unrealized gains is associated with approximately a 1 percentage point lower annual outflows, representing about a 7% decrease relative to the mean outflow rate, depending on the specification. This suggests that investors facing substantial unrealized gains in their shares do not switch, even when lower-cost, more tax-efficient alternatives are available.

Second, managers exploit this tax lock-in by extracting higher fees. When unrealized gains are high, managers have higher fees and, likewise, lower fee waivers, which are temporary reductions in fees that lower the net expense ratio paid by investors. A one standard deviation increase in unrealized gains reduces fee waivers by about 0.6 percentage points, equivalent to a 7.1% decrease relative to the mean waiver rate. This pattern indicates that managers recognize when investors are locked in and adjust their fee accordingly.

Together, these two facts suggest that unrealized gains provide stability to MFs: they reduce outflows and allow managers to sustain higher fees despite the inferior tax treatment of MFs relative to other choices. However, this apparent stability masks a source of fragility. The lock-in channel depends on investors choosing to remain in the fund to avoid realizing their unrealized gains. When outflows do occur, the tax consequences fall on those who stay. This creates a potential feedback loop between outflows and capital gains distributions that the third fact reveals.

To understand this feedback mechanism, the final fact examines how fund-level unrealized gains and outflows jointly determine capital gain distribution yields. When MFs experience outflows, they might sell portfolio holdings to meet redemptions, and if these holdings have appreciated, the realized gains are distributed to remaining shareholders as capital gains distributions. Note, high fund-level unrealized gains alone do not generate large capital gain distribution yields; rather, outflows are required to force asset sales that realize the fund's unrealized gains. Similarly, high outflows in funds with low unrealized gains produce modest capital gain distribution yields. Thus it is the interaction that matters: funds with both high fund-level unrealized gains and high outflows have high capital gain distribution yields. This interaction generates strategic complementarity between investors in the fund. As some investors sell shares of a fund with high unrealized gains, remaining investors are more incentivized to sell their shares to avoid receiving capital gains distributions.

To rationalize these empirical patterns, I next develop a three-period bank-run model adapted to the MF setting, extending Diamond and Dybvig (1983). The model adds heterogeneous investor unrealized gains and endogenous capital gains distributions. In the first period, investors invest in the MF. In the second period, investors observe their unrealized gains and decide whether to remain in the MF or switch to a lower-fee, tax-efficient alternative. Investors with low unrealized gains find it optimal to exit and avoid future distributions, while those with high unrealized gains remain locked in to defer realization of the MF shares. However, when investors redeem, the fund must sell assets to meet outflows. These asset sales realize unrealized gains at the fund level, which are distributed pro-rata to remaining shareholders under pass-through taxation rules. In the third period, all investors liquidate and consume. In the model, the MF manager sets fees to maximize revenue.

The model shows the strategic complementarity: each investor's exit decision depends on others' exits, since aggregate outflows determine the distribution burden on those who stay. It is precisely the combination of endogenous capital gains distributions and heterogeneity in investor-level unrealized gains that gives rise to the potential for multiple equilibria. Endogenous distributions create the feedback loop whereby exits force asset sales that impose tax costs on those who remain, while heterogeneity in unrealized gains determines which investors are locked in and which are near the margin of switching. Together, these two features generate multiple equilibria under identical fundamentals. In the stable equilibrium, few investors exit, distributions remain modest, and staying dominates. In the run equilibrium, many investors exit, triggering large distributions that make exit optimal. Identical fund fundamentals can support either outcome depending on investor coordination.

The model serves four purposes. First, it formalizes the mechanisms generating the empirical patterns: how investor-level unrealized gains create lock-in while fund-level unrealized gains amplify distribution sensitivity to outflows. Second, it derives implications for manager fee-setting behavior when investors face heterogeneous switching costs. Third, it demonstrates that identical fund fundamentals can support multiple equilibria, revealing inherent fragility in the MF structure. Fourth, parameterized to the data, the model enables comparative static analyses examining how policy and manager choices affect these dynamics.

To explore the model's implications, I set baseline parameters consistent with observed data and conduct three comparative static exercises. First, I compare how funds meet redemptions by varying which securities they sell. Under proportional liquidation, the fund sells all holdings proportionally, maintaining portfolio weights. Under least-gains-first, the fund prioritizes selling securities with the lowest unrealized gains (securities with high cost-basis). While least-gains-first reduces distributions in the stable equilibrium by prioritizing these high-basis assets, it accelerates fragility during stress. This is because once high-basis assets are depleted, only low-basis securities remain, amplifying distributions when outflows escalate. Second, I examine how capital gains tax rates affect the switching decision. Higher tax rates strengthen lock-in by increasing switching costs, but they also magnify losses in the fragile equilibrium by increasing the after-tax burden of distributions on remaining investors. Finally, I analyze the role of liquidity shocks, such as consumption needs or mandatory retirement withdrawals, that force some investors to exit regardless of tax considerations. The role of retirement investors is particularly important in this context, as MFs remain the predominant vehicle for retirement plans, and these tax-deferred accounts are not exposed to capital gains taxes on distributions. Using liquidity shocks, I model how exits by retirement investors affect the remaining taxable investors who bear the full tax burden of resulting distributions. Even modest forced exits by tax-exempt retirement investors substantially increase distribution burdens on remaining taxable investors- as these outflows trigger asset sales whose tax consequences fall entirely on those who remain.

These findings have important implications given the scale and significance of MFs in household portfolios. In 2024, 54 percent of US households owned MFs (Holden, Schrass, and Bogdan, 2024), making them one of the most widespread investment vehicles in the economy. With over $10 trillion in US equity MF assets, understanding the forces that sustain MFs despite their disadvantages, and the conditions under which they become fragile, matters for millions of investors and the broader financial system. The strategic complementarities documented here suggest that funds with high unrealized gains face coordination risk similar to bank runs, but transmitted through tax externalities rather than asset illiquidity. The analysis reveals a fundamental tension: the same unrealized gains that allow MFs to retain assets despite their disadvantages also make them vulnerable to sudden fragility when redemptions begin.

The paper is organized as follows. Section 2 positions this paper with the existing literature. Section 3 describes the institutional details of MFs, ETFs, and SMAs, and presents the data construction. Section 4 presents the empirical facts that motivate the model, which is presented in Section 5. Section 6 presents the comparative statics. Finally, Section 7 concludes.

2 Related Literature

This paper contributes to several strands of literature on taxation, competition across investment vehicles, and MF fragility. First, this paper builds on seminal research examining how taxation shapes investor behavior and portfolio choice (Domar and Musgrave (1944), Feldstein, Slemrod, and Yitzhaki (1980), Stiglitz (1983), Constantinides (1983)). This early theoretical work establishes that tax timing considerations and lock-in effects fundamentally alter optimal investment strategies. Empirical evidence confirms that household trading behavior responds to tax incentives (Ivković, Poterba, and Weisbenner, 2005), and that unrealized gains create lock-in effects that reduce portfolio turnover and selling propensities (Ivković and Weisbenner, 2009).

Within MFs specifically, Barclay, Pearson, and Weisbach (1998), Dickson, Shoven, and Sialm (2000) and Dammon, Spatt, and Zhang (2001, 2004) analyze how MF structures generate tax inefficiencies, showing that capital gains distributions impose tax costs across investors. In particular, Bergstresser and Poterba (2002) document that investors avoid funds with high unrealized gains to reduce exposure to future distributions. Building on these insights, Sialm and Starks (2012), Sialm, Starks, and Zhang (2015), Sialm and Zhang (2020) emphasize the importance of tax clienteles and tax-efficient management practices, showing that fund flows and performance depend critically on the composition of taxable versus tax-exempt investors. Beggs, Hill-Kleespie, and Liu (2022) show that tax burdens continue to shape investor demand and distort the competitive landscape across funds. More broadly, Mainardi (2025) demonstrates that fiscal reforms generate large and persistent reallocation of household savings across institutions, underscoring the first-order importance of tax policy for asset allocation decisions. Building on Dickson, Shoven, and Sialm (2000), who demonstrate that redemptions impose tax externalities on remaining mutual fund shareholders through forced capital gains distributions, I examine how the availability of tax-efficient alternatives (ETFs and SMAs) transforms these within-fund externalities into a broader competitive dynamic that drives reallocation across investment vehicles. As investors gain access to vehicles that avoid these tax costs entirely, the externalities documented by Dickson, Shoven, and Sialm (2000) accelerate investor exit toward alternative structures, creating fragility in the MF vehicle itself.

Second, this paper contributes to the literature on alternative investment vehicles. Literature on ETFs emphasizes several potential advantages relative to MFs, including lower fees, intraday tradability, and enhanced liquidity (Poterba and Shoven (2002), Ben-David, Franzoni, and Moussawi (2017), Lettau and Madhavan (2018), Helmke (2024)). More recently, Moussawi, Shen, and Velthuis (2025) demonstrate that the dominant advantage of ETFs lies in their superior tax efficiency through in-kind redemptions. This paper takes the tax efficiency advantages of ETFs as established by Moussawi, Shen, and Velthuis (2025) and examines how these differences shape the dynamics of investors switching decisions and MF stability. Moreover, I show how tax lock-in enables MF managers to sustain higher fees: investors with large unrealized gains face switching costs that reduce competitive pressure despite the availability of lower-cost alternatives.

The literature on SMAs has focused primarily on performance comparisons with MFs. Early studies find only modest SMA advantages, largely explained by lower fees and turnover (Elton, Gruber, and Blake, 2013). More recent twin designs show that SMAs managed alongside MFs are often favored over MFs without a twin (Chen, Chen, Johnson, and Sardarli, 2017), and Rohleder, Tentesch, Weh, and Wilkens (2022) find that portfolio divergence within twins reduces risk-adjusted performance. Fedyk (2023) examine investor behavior and show that SMA flows respond primarily to Morningstar ratings rather than risk-adjusted returns. Recently, Zhang, Guo, Cheng, and Wang (2024) extend performance comparisons to ETFs, finding that SMAs underperform ETFs in large-cap strategies but outperform in mid- and small-cap segments on a gross return basis. This paper complements this performance-focused literature by examining how the tax advantages inherent in SMA structure affect investor switching decisions and MF persistence.

Moreover, this paper contributes to the broader literature on strategic complementarities among investors and implications for financial stability. Chen, Goldstein, and Jiang (2010) show that MFs holding illiquid assets are especially vulnerable: when some investors withdraw, funds sell assets to meet redemptions, which reduces the value of the remaining shares and increases the incentive for others to follow suit. Goldstein, Jiang, and Ng (2017) extend this insight to corporate bond funds, demonstrating how fragility is amplified in fixed-income markets where liquidation costs are high. More broadly, this literature connects MF redemptions to asset price dynamics and market stability through illiquidity channels (Coval and Stafford (2007),Chernenko and Sunderam (2020), Jiang, Sun, and Wang (2022) Ma, Xiao, and Zeng (2022)).

Methodologically, this paper relates to models of runs and coordination frictions, following Diamond and Dybvig (1983). Specifically, this paper builds on the literature of strategic complementarities including Chen, Goldstein, and Jiang (2010). The contribution in this paper is identifying a distinct channel for strategic complementarities that operates independent of asset illiquidity. In prior work, redemption costs operate through price impact and fire sales. In this setting, the externality operates directly through the tax code: when investors redeem their shares from a MF, and the fund sells appreciated securities, this realizes capital gains that are distributed pro rata to remaining shareholders. This pass-through taxation structure transforms individual exit decisions into collective costs, generating fragility even when underlying assets are liquid. This distinction has important implications: while illiquidity-driven runs call for liquidity management tools, tax-driven fragility requires different solutions including fund-level tax management practices and policy on taxation rules.

Finally, this paper contributes to the literature on financial innovation, which has historically been attributed to taxes and regulations (Miller, 1986). This paper reexamines this familiar narrative through a new lens: rather than focusing solely on the benefits these innovations provide to adopters, I explore how the same tax forces that drive financial innovation can simultaneously create fragility and strategic complementarities in the incumbent structures left behind. In doing so, I connect the literature on tax-motivated financial innovation to the literature on financial stability, showing that the two are fundamentally intertwined.

3 Institutional Details and Data

3.1 Institutional Details

This paper considers three vehicles that provide professional management of diversified portfolios: MFs, ETFs, and SMAs. While functionally similar in offering diversified exposure to financial assets, SMAs and ETFs have expanded to rival MFs in scale (Figure 1). Their structural differences though create distinct implications for taxation, which this section discusses, while Section 4.2 examines fee differentials.

MFs and ETFs are pooled investment vehicles in which investors purchase shares representing part ownership of a fund's portfolio, with the manager collectively investing on behalf of all shareholders. This pooled structure creates an indirect connection between investors and portfolio assets. However, there is a key structural difference between MFs and ETFs that lies in how shares are traded with investors. Illustrated in Figure 2, MFs operate solely through a primary market where the fund manager directly issues and redeems shares with investors. ETFs, on the other hand, operate through a dual-market structure: individual investors trade shares on a secondary market, while institutional authorized participants (APs) facilitate share creation and redemption in the primary market typically through in-kind transactions.[2]

SMAs offer a third structural alternative. An SMA is an investment vehicle in which a professional manager replicates a specified investment strategy across individual investor portfolios. As shown in Figure 2, unlike pooled vehicles, SMAs provide investors with direct ownership of individual securities. Historically limited to institutions and high-net-worth individuals, SMA minimum investment requirements have declined substantially, with the median falling from millions to $100,000 (Figure 3).[3] The structural differences between investment vehicles create distinct consequences for capital gains taxation.

Tax Implication of Vehicle Structure: Capital Gains Distributions The MF structure is most susceptible to capital gains distributions, which are payments made by a fund to its shareholders representing the fund's net realized gains from selling securities within its portfolio. Because MFs operate solely through a primary market, redemptions require the fund to sell portfolio holdings directly, realizing gains that must be passed through to remaining shareholders. MFs are required to pass through at least 90% of taxable income, including capital gains, to shareholders annually to maintain their pass-through tax status.[4] In practice, most MFs distribute gains only once per year, with 68.9% of MFs averaging a single distribution annually between 1970 and 2024 (Figure A.IV), typically in December (Figure A.III).[5] The magnitude of these distributions depends on whether the securities sold to meet redemptions have appreciated, making unrealized gains central to understanding the stay-or-switch decision.

Two types of unrealized gains shape the decision to stay in a MF or switch to an alternative vehicle. First, fund-level unrealized gains represent the embedded appreciation in the MF's portfolio holdings. When the MF sells appreciated securities through portfolio rebalancing, or to meet investor redemptions, these fund-level gains are realized and passed through to shareholders as taxable distributions.[6] Second, investor-level unrealized gains represent the appreciation of

[2]Trading in-kind means transferring assets in their original form which here means as securities rather than exchanging in cash. Trading in-kind avoids the realization event that triggers capital gains distributions.

[3]This decline reflects technological advances that allow managers to scale SMA offerings more efficiently. Additional institutional details of SMAs are in Appendix D.I.

[4]Under IRC Subchapter M, regulated investment companies (MFs) can avoid corporate taxation by distributing at least 90% of their taxable income to shareholders. The Investment Company Act of 1940 governs their structure and operations.

[5]Additional features of capital gains distributions are discussed in Appendix **??**.

[6]Distributions include both short-term capital gains (taxed at ordinary income rates) and long-term capital gains (taxed at preferential rates). This paper focuses on total capital gains distributions, excluding dividend distributions.

each investor's MF shares, which determines the tax cost of switching to another vehicle. Fund-level gains create distribution risk that encourages switching, while investor-level gains create switching costs that encourage staying.

Although conceptually distinct, these two measures are closely related. MFs with appreciated portfolios have generally delivered strong returns that create unrealized gains for their shareholders. Thus, fund-level unrealized gains provide a natural proxy for the degree of tax overhang faced by the investor base: when a fund reports high fund-level unrealized gains, this suggests that, on average, its investors hold shares with significant embedded appreciation and face higher tax costs upon switching.[7]

A simple example illustrates how unrealized gains generate both the switching costs and the distribution externalities central to this paper. Suppose a MF initially holds two shares of Stock X purchased at $10 each, with two investors each holding one fund share at NAV $10. When Stock X rises to $50, a new investor enters and the MF buys one additional share at $50. The MF now has $80 of unrealized gains ($40 on each low-basis share, zero on the high-basis share), while original investors each hold shares that appreciated by $40.

Suppose one original investor redeems, realizing $40 in gains and paying taxes immediately. To meet the redemption, the MF must sell one Stock X share. Which share it sells determines the tax consequences for remaining investors: selling the high-basis share generates no taxable gain and thus no distribution, while selling a low-basis share realizes $40 that is split among remaining investors, imposing $20 each. The resulting distribution creates a strategic complementarity: by increasing the tax burden of staying, one investor's redemption makes switching more attractive for remaining investors, potentially triggering further redemptions. The MF's liquidation choice thus directly shapes the tax burden on its investors, with high-basis sales minimizing distributions and low-basis sales magnifying them. Moreover, which securities remain in the portfolio after liquidation affects the fund's vulnerability to future redemptions, as I explore formally in the model.

A critical feature of mutual fund taxation is that investors are not taxed twice. When remaining investors receive the $20 distribution in this example, the investor cost basis is adjusted. Thus, when the investor eventually redeems, the investor pays taxes only on the appreciation

[7]Formally, the fund-level unrealized gains ratio equals the value-weighted average of investor-level embedded gain ratios whenever the fund's total portfolio cost basis equals the aggregate amount investors paid for their shares. This condition holds exactly for new investments and reinvested distributions, and diverges only through redemptions where the cost basis of liquidated securities differs from the redeeming investor's cost basis.

since the distribution, not on the full gain from the original purchase.[8] The tax cost of the distribution is not the tax itself, but rather the timing: investors pay taxes earlier than if gains had remained unrealized, forgoing the returns those tax payments could have earned if they remained invested. This accelerated taxation is precisely the externality: one investor's redemption forces others to pay taxes sooner than they otherwise would, reducing the present value of their after-tax wealth even though the total tax liability remains unchanged. With the median equity MF having doubled over the past 15 years (Figure A.II), such switching costs are substantial.

ETFs, while also pass-through entities, largely avoid distributions through their dual-market structure. In the secondary market, ETF shares trade on an exchange between investors, including large institutional investors known as authorized participants (APs), without directly affecting the fund's portfolio or shares outstanding. In the primary market, only APs create or redeem shares with the ETF. These redemptions are typically conducted in-kind, with baskets of underlying securities exchanged rather than cash. Because the ETF delivers securities directly rather than selling them for cash, no realization event occurs and no taxable gains are generated.[9] From 2001 to 2024, 58% of MFs distributed capital gains annually versus only 4.8% of ETFs (Figure 4).[10]

SMAs eliminate distribution risk entirely through direct ownership. Each investor owns securities individually, insulated from others' redemptions. Moreover, this direct ownership also enables tax loss harvesting, where managers can realize losses to offset gains elsewhere in the investor's portfolio or, up to annual limits, ordinary income.[11]

[8] In simplified terms: when the $20 distribution is made, the NAV of each MF share drops by $20 (from $50 to $30). If investors reinvest their $20 distribution, they purchase additional shares at the new $30 NAV (receiving 0.67 shares). Their total cost basis is now $30: the original $10 purchase plus the $20 reinvested distribution. If they immediately sold all shares, they would realize a $20 capital gain ($50 sale proceeds minus $30 cost basis). Thus, the total tax paid is the same: $20 from the distribution plus $20 from the sale equals $40 of taxable gains, which matches the $40 gain they would have realized had there been no distribution.

[9] IRC Section 852(b)(6) exempts in-kind distributions from capital gains recognition. While this provision predates ETFs and technically allows MF in-kind redemptions, Agarwal, Ren, Shen, and Zhao (2022) find MF in-kind redemptions are infrequent and driven by liquidity needs rather than systematic tax management.

[10] See Moussawi, Shen, and Velthuis (2025) and Colon (2017) for detailed analysis of ETF tax efficiency.

[11] Importantly, while MFs and ETFs are required to distribute realized gains, they cannot distribute realized losses. Instead, losses are carried forward within the fund. SMA investors, however, can immediately use realized losses. This timing advantage is valuable because forced gain realization reduces present value through immediate taxation, while accelerated loss realization provides immediate tax benefits.

Twin Investment Vehicles MFs often offer twins with ETFs or SMAs, which is key for this analysis as it allows direct comparison of otherwise identical strategies across vehicle structures.[12] A twin is defined as cases where the same asset management firm offers the same strategy across multiple vehicle types, identified by Morningstar's StrategyId variable.

As shown in Figure 5, a large share of MFs are paired with either an ETF or SMA twin, implying an investor holding a given strategy in an MF could instead access the same strategy through a different vehicle. The pairing pattern is notable: MFs match with either ETFs or SMAs, but rarely both. This distinction stems from the origin of these vehicles: ETFs emerged primarily as passive alternatives, while SMAs developed as active strategies.

The twin structure provides an ideal setting for analyzing investor behavior because they are indistinguishable along key dimensions including management and portfolio composition. Among SMA-MF and ETF-MF twins, 93.7% and 90.8% respectively share at least one manager, with 38.9% and 33.9% respectively sharing identical management teams. The similarity extends to nearly identical portfolio construction: the average SMA-MF twin shares 17 of the top 20 holdings with its MF counterpart (Table A.VIII). These twins isolate how vehicle structure and resulting tax treatment affect investor decisions.

3.2 Data

The primary data source for investment vehicle data is Morningstar Direct, restricted to US-domiciled vehicles in the US equity category group. A key distinction in coverage is that MF and ETF data are collected by Morningstar from mandatory regulatory filings, while SMA data are self-reported by managers at the composite level.[13] This difference reflects the regulatory structure; unlike pooled funds, SMAs are organized as non-pooled vehicles under Rule 3a-4 of the Investment Company Act of 1940 and are exempt from the registration and reporting requirements applied to MFs and ETFs. MF data are reported at the share-class level. Throughout the analysis, I present results at both the share-class and fund levels, with fund-level aggregates value-weighted by total net assets across share classes.

Morningstar reports assets, returns, fees, and categorical variables including inception date, investment style, an index-fund indicator, and the strategy identifier that links twin strategies

[12] Additional papers that use these twins include (Chen, Chen, Johnson, and Sardarli, 2017; Rohleder, Tentesch, Weh, and Wilkens, 2022).

[13] An SMA composite represents the aggregated performance of all client portfolios managed under the same strategy. Figure 2 illustrates an SMA composite made up of three individual SMA accounts.

across vehicles. A detailed description of the data is provided in Appendix B.I. Important to this analysis, for MFs and ETFs, fees are reported directly as net expense ratios in Morningstar. In contrast, SMA fees are not as transparent as MFs and ETFs. More than half of SMAs have fees negotiated between the investor and the manager, and these fees will decrease with minimum investment, as shown in Figure A.V. In line with the literature, I calculate SMA fees as the difference between gross and net returns (Elton, Gruber, and Blake, 2013).

Capital gains distributions data come from both Morningstar and CRSP. Morningstar reports annual aggregates of long-term and short-term distributions, while CRSP records distributions at the event level with distribution dates. The CRSP data enable analysis of distribution timing and its relationship with flows, and Morningstar provides comprehensive universe coverage. Details on distribution construction and matching procedures across databases are in Appendix B.II.

Moreover, this paper incorporates new data on fund-level unrealized gains constructed from SEC filings. While investor-level unrealized gains are not directly observable, fund-level unrealized gains provide a useful proxy for the degree of tax overhang faced by the shareholder base. When a fund reports larger fund-level unrealized gains, this suggests that, on average, its investors hold shares with significant embedded appreciation and therefore face higher tax costs upon redemption.

To construct this measure, I combine two regulatory data sources: Certified Shareholder Reports (SEC Form N-CSR) and Portfolio Holdings Reports (SEC Form N-PORT). Form N-CSR provides annual stock measures of gross unrealized gains at the fund level; while Form N-PORT supplies monthly flow measures of "Net change in unrealized appreciation (depreciation)." Because each filing alone is insufficient, the two are integrated: N-CSR anchors the level of unrealized gains, and N-PORT captures the month-to-month variation. As Form N-PORT reporting began in 2019, this defines the start of the sample, covering US equity MFs from 2019 to 2024.

First, I use Form N-CSR to obtain the gross stock measure of unrealized gains. However, the key empirical challenges is that in practice, these disclosures are highly unstructured: some funds report unrealized gains in clear tabular summaries, while others cite the number in footnotes. To recover these data consistently, I parse and segment all N-CSR filings and then apply large language models (GPT-4o supplemented by GPT-4-turbo) to extract the relevant information.

I then merge the annual levels from N-CSR with the monthly changes reported on Form N-PORT. Because funds file N-CSR at different fiscal year-ends, annual levels alone do not provide comparable cross-fund measures at a given point in time. The monthly N-PORT flows allow me to construct monthly unrealized gains measures across all funds from 2019 to 2024. Moreover, N-PORT filings provide an internal validation of the extraction: cumulative monthly flows between annual N-CSR reports closely match the unrealized gain levels extracted from the filings (Figure A.VII). Appendix B.III describes this process in detail.

To facilitate comparisons across funds, unrealized gains are scaled by total net assets to obtain the unrealized gain ratio:

$$\text{Unrealized Gain}_{i,t} = \frac{\$ \text{ of Unrealized Gains}_{i,t}}{\text{TNA}_{i,t}},$$

where \$ of Unrealized Gains$_{it}$ are the accumulated unrealized gains of fund i, at time t and TNA$_{it}$ is the total net assets of fund i at time t. Intuitively, this measure captures the fraction of a fund's portfolio that would be taxable if liquidated. For example, a ratio of 0.20 implies that 20 percent of assets represent unrealized appreciation that will be taxed if sold. For clarity and consistency, I henceforth refer to fund-level unrealized gains scaled by total fund assets as the unrealized gains ratio. I use unrealized gains and unrealized gains ratio interchangeably to refer to this scaled measure.

The constructed dataset reveals substantial cross-sectional and time-series variation in unrealized gains, illustrated in Figure 6. Figure 6a shows the distribution of unrealized gain ratios across MFs. The average MF has 23.6 percent of unrealized gain. Moreover, the dispersion is wide, with some funds carry significant unrealized gains in excess of 60 percent, while others report negative values corresponding to accumulated capital losses.[14] Figure 6b presents the time-series of unrealized gains. The series exhibits pronounced cyclical fluctuations. Average unrealized gains fell sharply during the COVID-19 market shock in early 2020 and again in 2022 amid equity market declines- before recovering in subsequent periods. The shaded region shows substantial dispersion between funds, with the 10th to 90th percentile range spanning from capital losses to gains exceeding 50 percent of assets at various points in time.

[14]Importantly, while gains are required to be distributed, losses are not and are instead carried forward within the MF.

4 Empirical Facts

This section establishes three empirical facts that reveal how unrealized gains shape MF dynamics. The first fact documents lock-in: funds with higher unrealized gains experience significantly lower outflows, even as alternative vehicles offer lower fees and better tax efficiency. This is consistent with tax-based lock-in: investors facing substantial embedded gains would trigger immediate capital gains taxes by switching. The second fact shows that managers utilize this lock-in. Funds with higher unrealized gains charge higher fees and offer fewer fee waivers, controlling for performance and fund characteristics. When investors cannot leave without a tax penalty, competitive pressure weakens and managers can extract higher rents. The third fact uncovers the fragility embedded in this apparent stability. When redemptions do occur, funds with high unrealized gains must sell appreciated holdings, generating large capital gains distributions that impose tax costs on remaining investors. This interaction between outflows and unrealized gains creates strategic complementarities: one investor's exit raises the tax cost of staying for others, potentially triggering further redemptions. Together, these facts show that unrealized gains simultaneously sustain MFs through lock-in and make them vulnerable to self-reinforcing outflow cycles.

4.1 Unrealized Gains and Outflows

Fact 1. Mutual funds with higher unrealized gains experience less outflows, even as alternative vehicles offer lower fees and better tax efficiency.

Using the constructed measure of unrealized gains, I examine the relationship between unrealized gains and outflows from MFs. Because any redemptions can trigger asset sales that realize capital gains for distribution, I construct annual outflows by summing negative net flows across all months, scaled by lagged assets.[15] This monthly aggregation captures the cumulative redemption pressure that forces liquidations throughout the year.

Figure 7 illustrates the baseline relationship between lagged unrealized gains and MF outflows. The figure shows a clear negative relationship: funds with higher unrealized gains experience lower outflows, consistent with tax-induced lock-in. While this descriptive pattern is suggestive, it may be confounded by other factors that jointly influence both unrealized gains and outflows. To isolate the effect of unrealized gains on outflows, I estimate the following spec-

[15]See Appendix B.I for details.

ification, which introduces controls for past performance and observable fund characteristics:

$$\text{Outflows}_{it} = \beta_1 \,\text{UnrealizedGains}_{i,t-1} + \beta_2 \,\text{Performance}_{i,t-1} + \gamma' \mathbb{X}_{it} + \mathcal{S}_i + \delta_t + \varepsilon_{it}, \qquad (1)$$

where Outflows_{it} denotes the outflow ratio for MF i in year t which is the annual outflows divided by the lagged assets. $\text{UnrealizedGains}_{i,t-1}$ is the lagged unrealized gain ratio. $\text{Performance}_{i,t-1}$ controls for past fund performance using either prior-year returns or alpha estimated from the Fama–French three-factor model over the preceding 12 months.[16] The vector $\mathbb{X}_{it}$ includes standard determinants of fund flows including turnover ratio, log fund age, log total net assets, expense ratio, cash holdings ratio, and an indicator for active management. Style fixed effects $\mathcal{S}_i$ account for differences in investment objective, while time fixed effects δ_t capture market-wide shocks to flows. Standard errors are clustered at the fund level.[17] The regressions are estimated at the annual frequency to avoid seasonal fluctuations.

Table 2 reports the results. Column (1) shows that a one percentage point increase in lagged unrealized gains is associated with 0.113 percentage points lower annual outflows, significant at the 1% level. Column (2) adds time and style fixed effects along with fund characteristics commonly associated with redemption decisions: turnover, age, size, expense ratio, cash holdings, and an active management indicator (Goldstein, Jiang, and Ng, 2017). These characteristics control for operational and structural differences that could independently influence both investor behavior and a fund's tax position. The coefficient on unrealized gains is –0.055, indicating that while fixed effects and observable characteristics explain part of the baseline relationship, a substantial lock-in effect persists.

Columns (3) and (4) add performance controls to address the concern that poorly performing funds may simultaneously accumulate losses (reducing unrealized gains) and experience higher redemptions. Column (3) uses fund alpha estimated from the Fama-French three-factor model while Column (4) uses prior-year returns. Even after controlling for performance, the coefficient remains stable at –0.055 to –0.059, confirming the relationship reflects tax considerations rather than fund performance.

Columns (5) and (6) restrict the sample to MFs that have a twin SMA or ETF offering a similar investment strategy. This restriction is particularly informative because it isolates funds where investors have a clearly available outside option where they could switch to a more

[16] I estimate risk-adjusted returns using the Fama–French three-factor model with 12-month rolling windows. For MFs, I compute net-asset-weighted returns at the fund level. Factor data from Ken French's data library.

[17] Variables are winsorized at the 1% level.

tax-efficient vehicle yielding the same return. Despite this available alternative, the lock-in effect persists with coefficients of –0.050 to –0.055, statistically significant at the 1% and 5% levels respectively. This finding strengthens the tax lock-in interpretation: even when superior alternatives are readily available, investors with unrealized gains remain in the MF.

To interpret the economic magnitude, consider the substantial cross-sectional variation in unrealized gains documented in Figure 6. Using the coefficient from Column (4), a one standard deviation increase in unrealized gains is associated with 1.12 percentage points lower annual outflows, representing a 7.3% decrease relative to the mean outflow rate. For the average fund with $806 million in assets (the median TNA), this translates to retaining an additional $9.0 million per year.

In total, the results provide consistent evidence that unrealized gains constrain outflows. Appendix Tables A.I and A.II replicate the main specifications replacing style fixed effects with fund fixed effects, which absorb all time-invariant fund characteristics. The results are robust: the coefficient on unrealized gains remains negative and statistically significant across all specifications, confirming that the lock-in effect is not driven by unobserved time-invariant heterogeneity at the fund level. Together, these findings suggest that unrealized gains meaningfully constrain investor mobility, reducing competitive pressure on fund managers to keep fees low. The next fact examines whether MFs with higher unrealized gains indeed charge higher fees, consistent with managers extracting rents from investors who face prohibitive switching costs.

4.2 Unrealized Gains and Fees

Fact 2. Higher unrealized gains are associated with higher fees, consistent with managers extracting rents from tax-locked investors.

Although MFs could in principle compensate for their tax disadvantages by offering lower fees, they instead charge more than both SMAs and ETFs. This section begins by showing how MFs consistently impose higher fees than both SMAs and ETFs, adding a structural cost disadvantage on top of the tax burdens associated with capital gains distributions. The previous section established how unrealized gains constrain investor outflows. This finding suggests a strategic relationship: if unrealized gains reduce investor price sensitivity by creating tax switching costs, MF managers may exploit this reduced competition by maintaining higher fees. This section next shows how MFs with higher unrealized gains have higher fees, consistent with managers extracting rents from tax-locked investors.

Figure 8 shows that MF expense ratios are systematically higher than those of SMAs and ETFs. Consistent with the literature, Figure 8 highlights the declining trend in MF fees over time (Yang, 2023; Helmke, 2024). The average expense ratio for all MFs including active and passive funds has declined from 1.50% to 1% in 2024. Despite this decline, MFs continue to charge substantially more than alternative vehicles. Over the same 2000-2024 period, SMA expense ratios have averaged 0.87% and ETF expense ratios have averaged 0.36%, both remaining well below MF fee levels throughout.

These fee differences could reflect fund characteristics rather than vehicle type. To isolate the vehicle effect, I estimate:

$$\text{Expense Ratio}_{i,t} = \alpha + \beta_1 \text{SMA}_i + \beta_2 \text{ETF}_i + \gamma' \mathbb{X}_{i,t} + \mathcal{S}_i + \delta_t + \phi_s + \varepsilon_{i,t}, \tag{2}$$

where where Expense $\text{Ratio}_{i,t}$ is the annual net expense ratio for investment vehicle i in year t. SMA_i and ETF_i are indicator variables equal to one if investment vehicle i is a SMA or an ETF, respectively, with MFs as the omitted category. $\mathbb{X}_{i,t}$ is a vector of fund characteristics including an indicator for index funds, log total net assets, log fund age, prior-year performance and log minimum investment. $\mathcal{S}_i$, δ_t, and ϕ_s denote style, time, and strategy fixed effects, respectively.[18] Regressions are are estimated at the share-class level clustered standard errors.[19]

Table 3 presents results. Columns (1)-(2) establish baseline estimates: SMAs charge 40.8 basis points less than MFs, while ETFs charge 77.4 basis points less, both statistically significant at the 1% level. Column (3) adds controls for fund size, age, distribution channel, and performance. The vehicle fee differential persists, with SMAs charging 41.7 basis points less and ETFs charging 68.9 basis points less than comparable MFs. Columns (4)-(6) show these differentials remain consistent across retail, institutional, and retirement MF share classes. Most importantly, Column (7) uses only twin funds, which are MFs with corresponding ETF or SMA twins offering nearly identical strategies. Even within these closely matched pairs, SMAs charge 18.3 basis points less and ETFs charge 71.3 basis points less. Column (8) sharpens this comparison by additionally absorbing twin fixed effects, isolating variation across vehicle types within the same investment strategy. The fee advantages remain large and statistically significant: SMAs charge 10.5 basis points less and ETFs charge 13.2 basis points less than MFs, confirming

[18] Variables are winsorized at the 1% level.

[19] Share-class level is synonymous to fund level for ETFs and SMA as these investment vehicles do not have multiple share-classes.

that fee differences are a feature of vehicle structure rather than investment strategy or fund characteristics.

While the literature has extensively documented fee differentials across investment vehicles (Chordia, 1996; Hortacsu and Syverson, 2004), the interaction between fees and unrealized capital gains offers a novel perspective on fee extraction in MFs. Unrealized gains create switching costs that enable funds to charge higher fees or reduce fee waivers, effectively taxing locked-in investors. I examine this relationship through two complementary measures: net expense ratios and fee waivers.

I begin by examining the relationship between unrealized gains and net expense ratios through the following specification:

$$\text{ExpenseRatio}_{it} = \beta_1 \text{UnrealizedGains}_{it} + \beta_2 \text{Performance}_{it} + \gamma' \mathbb{X}_{it} + \mathcal{S}_i + \delta_t + \varepsilon_{it}, \tag{3}$$

where ExpenseRatio_{it} is the net expense ratio for MF i in year t. $\text{UnrealizedGains}_{it}$ is the unrealized gains ratio. Performance_{it} controls for past performance using either prior-year returns or alpha estimated from the Fama–French three-factor model over the preceding 12 months. The control vector $\mathbb{X}_{it}$ includes log total net assets and log fund age. $\mathcal{S}_i$ denotes style fixed effects and δ_t denotes year fixed effects. Standard errors are clustered at the fund level. Variables are winsorized at the 1% level.

Table 4 presents the results. Column (1) shows the baseline, Column (2), controls for performance and Columns (3) and (4) add controls for fund size and age. The coefficient in Column (4) implies that a one standard deviation increase in unrealized gains is associated with 0.77 basis points higher net expense ratios, representing a 1.0% increase relative to the sample mean of 0.80%.

To further examine how managers adjust fees in response to unrealized gains, I analyze fee waivers. MF fee adjustments typically occur through fee waivers rather than permanent fee changes, since the latter typically require shareholder approval while waivers can be implemented at management's discretion. Fee waivers represent temporary reductions where managers voluntarily absorb a portion of fund expenses. More details on MF fee waivers are shown in Appendix B.I. I measure the fee waiver relative to the MF's gross expense ratio for comparison across funds. Higher fee waivers mean lower fees to investors. The waiver ratio is defined as:

$$\text{Waiver Ratio}_{it} = \frac{\text{Fee Waiver}_{it}}{\text{Gross Expense Ratio}_{it}},$$

where the fee waiver is the amount waived by the MF for that year relative to the gross expense ratio of the fund for that year. Using this ratio I estimate:

$$\text{Waiver Ratio}_{it} = \beta_1 \text{UnrealizedGains}_{it} + \beta_2 \text{Performance}_{it} + \gamma' \mathbb{X}_{it} + \mathcal{S}_i + \delta_t + \varepsilon_{it}, \tag{4}$$

where Waiver Ratio_{it} is the fee waiver ratio for MF i in year t, defined as fee waivers divided by the gross expense ratio. The results are shown in Table 5. Column (1) provides the baseline estimate. Columns (2) through (5) add fund alpha, total net assets and age. The results provide consistent evidence that unrealized gains reduce fee waivers. Interpreting columns (3) and (4), a one standard deviation increase in unrealized gains is associated with 0.61 percentage points fewer fee waivers, equivalent to a 7.1 percent decline relative to the sample mean of 8.8 percent. This effect persists after controlling for fund performance, size, and age, providing evidence that managers exploit tax-driven investor lock-in through reduced promotional pricing. Note that in column (5) with the addition of both size and age controls, the coefficient declines but this reduction is likely attributed to fund size and age serving as proxies for unrealized gains, as larger and older funds tend to have accumulated more unrealized gains over time, thus controlling for some of the same investor lock-in effects.

Appendix Tables A.III and A.IV extend the main results by replicating all specifications with fund fixed effects, and by separately reporting alpha and return specifications across both the full sample and the twin fund subsample. The results are robust across all specifications: the positive relationship between unrealized gains and net expense ratios and the negative relationship between unrealized gains and fee waivers both persist after absorbing all time-invariant fund heterogeneity through fund fixed effects, confirming that these patterns reflect within-fund variation in unrealized gains rather than unobserved differences across funds.

The evidence demonstrates that managers of funds with higher unrealized gains maintain higher net fees and reduce fee waivers, consistent with extracting rents from tax-locked investors. Together, these facts suggest that unrealized gains stabilize MFs by reducing outflows and enabling higher fees despite inferior tax treatment. However, this stability is fragile. Lock-in works only when investors choose to stay; when outflows occur, the resulting distributions fall on remaining investors, creating a feedback loop that the third fact examines.

4.3 Capital Gains Distributions and Outflows

Fact 3. High unrealized gains amplify the distribution burden of redemptions, creating strategic complementarities where one investor's exit increases others' incentives to leave.

The previous facts established that unrealized gains create investor lock-in effects through the tax costs of switching: selling fund shares triggers capital gains taxes that discourage exit. MF managers exploit this lock-in by maintaining higher fees. However, tax-locked investors face a second, distinct tax cost. Remaining in funds with high unrealized gains exposes them to potentially large capital gains distributions triggered by other investors' redemptions. While the first tax cost (from switching) creates lock-in, this second tax cost (from distributions) creates pressure to exit. Critically, this distribution risk differs from standard liquidity concerns. Investors may want to exit not because they need immediate access to capital, but specifically to avoid the tax burden imposed by other shareholders' redemption decisions.

When some investors do redeem shares from a MF with substantial unrealized gains, the MF sells appreciated securities to meet redemptions, realizing capital gains that are distributed pro rata to all remaining shareholders. This creates a strategic complementarity among investors. Each investor's redemption decision imposes negative externalities on those who remain, as redemptions force the realization and distribution of unrealized gains that would otherwise remain deferred.

This mechanism is unique to the structure of MFs. When one MF investor redeems, all remaining investors bear the tax consequences of the forced asset sales. SMAs eliminate this risk entirely, as investors hold individual portfolios with no exposure to other investors' actions. Similarly, ETFs avoid this problem through their secondary market structure: investors trade shares with other investors rather than redeem directly from the fund, so redemptions do not force the fund to sell underlying securities and realize gains. Thus, while SMAs and ETFs protect investors from distribution risk, MF investors face a collective action problem where each redemption harms those who stay.

The strategic complementarity arises because large distributions require both conditions simultaneously: high unrealized gains provide the unrealized appreciation to be distributed, while high outflows force the fund to realize those gains through asset sales. Figure 9 illustrates this relationship by plotting distribution yields against unrealized gains for funds with high outflows (top tercile) versus low outflows (bottom tercile). To measure the magnitude of distributions, I use the distribution yield, defined as the per-share distribution amount relative to the predistribution NAV:

$$\text{Distribution Yield}_{it} = \frac{\text{Distribution}_{it}}{NAV_{it}^{\text{reivest}} + \text{Distribution}_{it}}.$$

Figure 9 highlights two important features. First, a fund with low unrealized gains will have lower distributions by construction: if positions have no gains, liquidating them generates nothing to distribute. In the figure, MFs with low unrealized gains have distributions consistently below 4%. Second, while funds with high unrealized gains do exhibit somewhat higher baseline distributions, these remain modest in the absence of redemptions. The economically large increases in distribution yields arise only when high unrealized gains coincide with high redemptions. Thus, the figure illustrates that both high unrealized gains and high outflows are jointly essential for large distributions, highlighting the trade-off: greater lock-in, but heightened vulnerability to redemptions.

To formalize this relationship, I estimate two specifications. The first examines the continuous interaction between outflows and unrealized gains, while the second explores how this amplification varies across terciles of unrealized gains during periods of high outflows.

$$\begin{aligned}\text{DistributionYield}_{it} = {} & \beta_1\,\text{Outflows}_{it} + \beta_2\,\text{UnrealizedGains}_{it} \\ & + \beta_3\left(\text{Outflows}_{it} \times \text{UnrealizedGains}_{it}\right) + \gamma'\mathbb{X}_{it} + \mathcal{S}_i + \delta_t + \varepsilon_{it},\end{aligned} \tag{5}$$

where $\text{DistributionYield}_{it}$ is the distribution yield of MF i in year t. Outflows_{it} denotes the outflow ratio for fund i in year t, constructed as the sum of negative net monthly flows over the year scaled by lagged total net assets. $\text{UnrealizedGains}_{it}$ is the unrealized gains ratio. The interaction term $\text{Outflows}_{it} \times \text{UnrealizedGains}_{it}$ captures whether the effect of redemptions on distributions depends on the level of unrealized gains. The control vector $\mathbb{X}_{it}$ includes prior-year performance (returns or alpha from the Fama–French three-factor model over the preceding 12 months), log total net assets, log fund age, turnover ratio, and cash holdings. $\mathcal{S}_i$ denotes style fixed effects and δ_t denotes year fixed effects. Standard errors are clustered at the fund level.[20] The regressions are estimated at the annual frequency, and distributions are aggregated at the yearly level if a MF distributes multiple times. This is consistent with the majority of MFs only distributing once a year and in December as shown by Figures A.IV and A.III.

Table 6 presents the results. Column (1) shows the baseline specification without controls. The coefficient of primary interest is β_3, the interaction term, which captures the amplification

[20] All variables are winsorized at the 1% level.

effect when both conditions are present simultaneously. The interaction coefficient is 0.48, positive and statistically significant at the 1% level. This indicates that the effect of outflows on distribution yields is amplified when funds hold higher unrealized gains. Column (2) adds the full set of controls including performance, fund size, age, turnover, and cash holdings. The interaction coefficient remains stable at 0.38 and highly significant, confirming that the amplification effect is not driven by observable fund characteristics.

To interpret the economic magnitude, consider a one standard deviation increase in unrealized gains. The interaction coefficient of 0.47 in column (2) implies an additional distribution yield of 1.3 percentage points. Given that these distributions represent taxable events for investors in taxable accounts, the magnitudes are economically significant and demonstrate how redemption activity imposes substantial externalities on remaining shareholders when funds carry elevated unrealized gains.

To further examine how this amplification varies across different levels of unrealized gains, I estimate the interaction separately across terciles of unrealized gains interacted with high outflows:

$$\begin{aligned}\text{DistributionYield}_{it} = {} & \beta_1\,\text{HighOutflows}_{it} + \sum_{j=2}^{3}\beta_j\,\mathbb{I}\big(\text{UnrealizedGains}_{Qj_{it}}\big) \\ & + \sum_{j=2}^{3}\beta_{j+2}\left(\text{HighOutflows}_{it}\times\mathbb{I}\big(\text{UnrealizedGains}_{Qj_{it}}\big)\right) \\ & + \gamma'\mathbb{X}_{it} + \mathcal{S}_i + \delta_t + \varepsilon_{it}.\end{aligned} \tag{6}$$

The dependent variable is consistent with Equation 5 as the distribution yield. HighOutflows_{it} is a binary variable equal to one if the MF experiences outflows above the top tercile in that year, capturing periods of unusually high outflow activity. The specification highlights whether MFs in different unrealized gains terciles respond differently to these high outflow periods. $\text{UnrealizedGains}_{it}$ is an indicator for the unrealized gains ratio that has been divided into three groups (j) based on their ratio: Q1 which is low and omitted as the baseline, Q2 of medium unrealized gains, and Q3 of high unrealized gains. The specification includes an interaction term between high redemptions and the top unrealized gains terciles. The interactions capture whether the relationship between high redemptions and distributions yields differs across MFs with varying levels of unrealized gains.

Columns (3) and (4) of Table 6 present the tercile specification results. Column (3) shows the baseline specification while Column (4) adds controls. Focusing on Column (4), the interaction between high outflows and the top tercile of unrealized gains (Q3) has a coefficient of 4.0 percentage points, statistically significant at the 1% level. This means that during periods of high outflows, funds in the top tercile of unrealized gains experience 4.0 percentage points higher distribution yields compared to funds in the bottom tercile facing similar outflow pressures. The interaction for the middle tercile (Q2) shows a coefficient of 2.0 percentage points, indicating that the amplification effect increases monotonically with the level of unrealized gains.

These results demonstrate the strategic complementarity central to the mechanism: redemption decisions amplify taxable distributions precisely when funds are most vulnerable due to unrealized gains. The magnitudes are economically large. A 4.0 percentage point increase in distribution yield for a fund with substantial unrealized gains facing high outflows represents a meaningful tax burden for remaining investors, creating incentives to exit preemptively and potentially triggering self-reinforcing outflow cycles.

5 Model

The three empirical facts reveal how tax frictions shape competition in the investment management industry. Despite the lower fees and superior tax treatment in SMAs and ETFs, many investors remain in MFs due to unrealized capital gains that create switching costs (Fact 1). This tax lock-in allows managers to sustain higher fees, evidenced by higher fees when unrealized gains are high (Fact 2). However, unrealized gains also create fragility: when redemptions occur, funds with high unrealized gains generate amplified distributions that burden remaining investors (Fact 3). The same unrealized gains that reduce outflows in normal times amplify the tax burden when outflows do occur. This dual role motivates a model where identical fund fundamentals can support both stable and fragile outcomes.

To rationalize these patterns, I develop a model of strategic complementarities in the spirit of Diamond and Dybvig (1983). The model features two key elements: (i) capital gains distributions are endogenously determined by redemptions, and (ii) investors differ in their unrealized gains, which generate heterogeneous tax switching costs. In the model, investors weigh avoiding future distributions against the immediate tax cost of switching to an outside option that eliminates the tax externality, such as an SMA or ETF. Investors with low unrealized gains can switch at little cost, while those with high gains face tax lock-in. In the model, the manager sets fees anticipating these dynamics. Redemptions trigger asset sales that generate mandatory capital gains distributions for remaining investors, creating a feedback mechanism: initial exits increase distributions, raising the burden on those who remain and potentially inducing further exits.

The model produces multiple equilibria through strategic complementarities. When few investors exit, distributions remain modest and staying is optimal; when many exit, distributions surge and further exits become optimal. The model thus formalizes a fragility channel where endogenous tax distributions interact with heterogeneous switching costs to generate coordination failures akin to bank runs.

5.1 Setup

The model unfolds over three periods, $t \in \{0, 1, 2\}$. Time $t = 0$, is a pre-period in which all investors are ex-ante identical and allocate one dollar to the MF. The MF invests in a portfolio of securities that appreciates over time, generating unrealized capital gains for investors. No decisions are made at this stage.

At $t = 1$, each investor privately observes their investor unrealized gain and whether they experience a liquidity shock, which determine their redemption decision. The manager sets fees for the MF, anticipating investor responses. Investors can also choose to invest in an outside option like an SMA or ETF. For notational simplicity, I refer to the outside option as the SMA throughout the model. Investors hit by liquidity shocks redeem immediately and consume, while others choose between remaining in the MF or switching to the outside option. Redemptions force the manager to sell assets according to the liquidation strategy, realizing gains that are distributed pro rata to remaining MF investors. At $t = 2$, all remaining positions are liquidated and consumed.

Investors

At $t = 0$, all investors invest one dollar into a MF. At $t = 1$, each investor privately observes two idiosyncratic characteristics: an unrealized capital gain $\alpha_i \in [\underline{\alpha}, \bar{\alpha}]$ and a liquidity shock $\lambda_i \in \{0, 1\}$. These determine the investor's redemptions decision.

The unrealized gain α_i represents the appreciation in the investor's MF position accumulated between $t = 0$ and $t = 1$. Exiting the MF at $t = 1$ requires realizing this gain and paying a capital gains tax rate $\tau \in (0, 1)$. Thus, α_i captures the investor's tax lock-in: a low α_i implies minimal switching costs, while a high α_i implies substantial tax liability upon exit. I assume

α_i is drawn independently across investors from a known distribution $P(\alpha)$ with continuous density $p(\alpha)$ on $[\underline{\alpha}, \bar{\alpha}]$. For the baseline analysis, I assume $\alpha_i \sim \text{Uniform}\,[0, \bar{\alpha}]$ which simplifies the exposition and makes the strategic complementarities most transparent, while preserving the essential heterogeneity in tax frictions. Note that the variation in alpha reflects unrealized capital gains for different investors of the same fund. For example, different investors might have bought the fund at different points in the past. The variation in alpha does not represent variation in fund-level unrealized gains across funds. Later, I illustrate the effect of fund-level unrealized gains by varying the distribution of alphas.

Each investor faces a liquidity shock with probability $\lambda \in [0, 1]$. Investors hit by the shock ($\lambda_i = 1$), must liquidate immediately at $t = 1$ and consume their after-tax proceeds rather than reallocate. Investors not hit by the shock ($\lambda_i = 0$), choose between remaining in the MF or switching to the outside option. By the law of large numbers, exactly λ fraction of investors experience liquidity shocks. The liquidity shock follows Diamond and Dybvig (1983) and captures individual consumption needs.[21]

Asset Manager

The asset manager operates the MF that invests in a portfolio of securities that appreciate over time, generating unrealized capital gains for investors. The manager is responsible for meeting redemptions, distributing capital gains, and collecting fees, but plays no direct discretionary influence on when or why investors redeem. Instead, the manager affects incentives through the fees (f_{MF}) chosen at $t = 0$.

Because the MF is a primary-market vehicle where the manager directly issues shares to investors, the manager observes the distribution of investor unrealized gains from shareholders. At $t = 0$, the manager sets fees with knowledge of this distribution ($P(\alpha)$) and the probability λ of liquidity shocks, but cannot observe individual investor types (α_i, λ_i), which are only revealed at $t = 1$. This prevents the manager from price discriminating between investors based on their realized characteristics.

There is an outside asset vehicle that by assumption, charges a lower fee than the MF ($f_{SMA} < f_{MF}$). For each dollar invested in the MF, the manager promises a gross return R, conditional on the investor holding the investment until final liquidation at time 2. The outside option also provides the same return as the MF as it has the same strategy. The manager earns

[21] The shock can also represent institutional events such as retirement plan menu changes that mandate exits from MFs, which I analyze in Section 6.2.

revenue by charging fees (f_{MF}) on total assets under management.

The MF follows a liquidation strategy $\omega \in \{\text{Proportional}, \text{Least-Gains-First}\}$, which determines how the manager executes redemptions by selecting which portfolio securities to sell. This strategy is fixed at fund inception and reflects the fund's operational structure and investment mandate. Under proportional liquidation, the manager sells all holdings proportionally to their portfolio weights, maintaining the fund's investment composition but realizing gains evenly across the investor unrealized gain distribution. Under least gain first liquidation, the manager prioritizes selling securities with the smallest unrealized gains (highest cost-basis), reducing immediate distributions to remaining investors but concentrating the portfolio in highly appreciated assets. These two approaches are considered because index funds typically follow approximately proportional liquidation to minimize tracking error, while active funds may adopt more tax-efficient approaches as shown in Appendix C.IV.

5.2 Value Functions

This section characterizes investor payoffs and the manager's optimization problem. I first derive terminal wealth for investors under each choice (forced liquidation, remain in MF, or switch to an outside option), then establish the cutoff rule that determines investor sorting. The key insight is that heterogeneity in unrealized gains α_i generates a monotone sorting condition: investors with low gains prefer the outside option to avoid future distributions, while those with high gains remain locked in the MF.

Investor Problem

Investor i's value function depends on their investor unrealized gain α_i, liquidity shock λ_i, the MF's distribution rate δ, and the fee structure:

$$D_i(\alpha_i, \lambda_i; \delta, f_{MF}, f_{SMA}) = \begin{cases} W_{i,t=1}(\alpha_i) & \text{if } \lambda_i = 1, \\ \max\left\{W^{MF}_{i,t=2}(\alpha_i; \delta, f_{MF}),\ W^{SMA}_{i,t=2}(\alpha_i; f_{SMA})\right\} & \text{if } \lambda_i = 0. \end{cases} \tag{7}$$

Investors experiencing liquidity shocks ($\lambda_i = 1$) redeem immediately at $t = 1$, receiving after-tax proceeds:

$$W_{i,t=1}(\alpha_i) = 1 + \alpha_i(1 - \tau), \tag{8}$$

which equals the initial investment plus the investor unrealized gain net of capital gains tax. These investors consume rather than reinvest.

Investors without liquidity shocks ($\lambda_i = 0$) compare terminal wealth at $t = 2$ under two options: stay in the MF or switch to the outside option. The model assumes that a single capital gains tax rate τ applies to all realized gains, whether from switching to the outside option, receiving MF distributions, or final liquidation.

Switching to the outside option at $t = 1$ requires immediate realization of investor unrealized gains α_i and payment of capital gains tax $\tau\alpha_i$. The after-tax proceeds are reinvested in the outside option, which then grows at net return $(R - f_{SMA})$ until liquidation at $t = 2$. Terminal wealth is:

$$W^{SMA}_{i,t=2}(\alpha_i; f_{SMA}) = \underbrace{[1 + \alpha_i(1-\tau)]}_{\substack{\text{After-tax proceeds} \\ \text{reinvested at } t=1}} \cdot \underbrace{[1 + (1-\tau)(R - f_{SMA})]}_{\substack{\text{Net growth from } t=1 \text{ to } t=2 \\ \text{taxed at liquidation}}}. \tag{9}$$

The first term is the after-tax amount available for reinvestment which is also the new cost-basis in the outside option. The second reflects growth net of fees and taxes. Crucially, the outside option wealth depends only on individual characteristics (α_i, f_{SMA}) and is unaffected by other investors' decisions.

Remaining in the MF exposes the investor to distributions generated by others' redemptions. At $t = 1$, redemptions force the manager to sell assets, realizing gains that are distributed to remaining investors. Let $\delta \in [0, 1]$ denote the distribution rate: the fraction of each remaining investor's position that is realized and distributed as taxable income. The investor receives this distribution, pays taxes, and can reinvest the after-tax amount at the MF's net return until $t = 2$.

Terminal wealth from staying in the MF is:

$$\begin{aligned} W^{MF}_{i,t=2}(\alpha_i; \delta, f_{MF}) = {} & \underbrace{1 + (1-\tau)[(1+\alpha_i)(1-\delta)(1 + R - f_{MF}) - 1]}_{\text{After-tax growth of undistributed position}} \\ & + \quad \underbrace{(1-\tau)(1+\alpha_i)\delta}_{\text{After-tax distribution at } t=1} \quad \underbrace{[1 + (1-\tau)(R - f_{MF})]}_{\text{Growth from } t=1 \text{ to } t=2}. \end{aligned} \tag{10}$$

The first line captures the after-tax value of the position that remains after distributions. The term $(1 - \delta)$ is the fraction not distributed, which grows at the MF's net return $(1 + R - f_{MF})$ and is taxed upon liquidation at $t = 2$. The second line is the after-tax value of distributions received at $t = 1$, which are reinvested in the MF and grow (net of fees and taxes) until $t = 2$.

Unlike the outside option wealth, MF wealth depends on the distribution rate δ, which is endogenously determined by the exit decisions of all investors. This creates the strategic complementarity: the more investors who exit, the higher δ becomes, reducing W^{MF} and making exit more attractive for those who remain.

Investor Sorting At $t = 1$, investors without liquidity shocks choose between remaining in the MF or switching to the outside option by comparing terminal wealth under each alternative. Define the switching value function as the difference in terminal wealth from remaining in the MF versus switching to the outside option:

$$g(\alpha_i; \delta, f_{MF}, f_{SMA}) := W^{MF}_{i,t=2}(\alpha_i; \delta, f_{MF}) - W^{SMA}_{i,t=2}(\alpha_i; f_{SMA}). \tag{11}$$

This function captures the net benefit of staying in the MF relative to switching. The key property is monotonicity in α_i:

Lemma 1 (Investor Sorting) *For any distribution rate δ and fees (f_{MF}, f_{SMA}), the switching value function $g(\alpha_i; \delta, f_{MF}, f_{SMA})$ is strictly increasing in α_i. Consequently, there exists a unique cutoff $\alpha^* \in [\underline{\alpha}, \bar{\alpha}]$ such that:*

- *Investors with $\alpha_i \geq \alpha^*$ remain in the MF,*
- *Investors with $\alpha_i < \alpha^*$ switch to the outside option.*

The cutoff α^* is defined by the indifference condition:

$$g(\alpha^*; \delta, f_{MF}, f_{SMA}) = 0. \tag{12}$$

The result follows from the increasing tax burden of switching as unrealized gains rise. Taking the derivative of equation (11) with respect to α_i:

$$\frac{\partial g}{\partial \alpha_i} = \frac{\partial W^{MF}}{\partial \alpha_i} - \frac{\partial W^{SMA}}{\partial \alpha_i}.$$

For the MF, additional unrealized gains increase the exposure to distributions but also the after-tax proceeds from remaining. For the outside option, additional unrealized gains increase the tax cost of switching. The net effect is positive: investors with higher α_i find the MF relatively more attractive because the immediate tax cost of switching dominates the exposure to future distributions. Since $g(\cdot)$ is strictly monotone, it crosses zero at most once, establishing uniqueness of the cutoff.

Investors with high investor unrealized gains face large tax liabilities if they switch, making the lock-in effect dominant. Investors with low gains pay little to switch and can escape future distributions by moving to the outside option. The cutoff α^* represents the marginal investor who is indifferent: those below switch, those above stay. This sorting mechanism is central to the model's dynamics: the cutoff α^* determines both the volume of redemptions and the resulting distribution rate δ, creating the feedback loop that drives fragility.

Asset Manager

Assets under management in each vehicle depend on investor sorting. Given cutoff α^* and liquidity shock probability λ, the MF retains investors with $\alpha_i \geq \alpha^*$ who are not forced to liquidate, while the outside option attracts those with $\alpha_i < \alpha^*$:

$$AUM_{MF,t=1}(\alpha^*, \lambda) = (1-\lambda) \int_{\alpha^*}^{\bar{\alpha}} (1+\alpha)\, p(\alpha)\, d\alpha, \tag{13}$$

$$AUM_{SMA,t=1}(\alpha^*, \lambda) = (1-\lambda) \int_{\underline{\alpha}}^{\alpha^*} [1+\alpha(1-\tau)]\, p(\alpha)\, d\alpha. \tag{14}$$

The MF's AUM is simply the sum of positions $(1+\alpha)$ for all remaining investors. The AUM of the outside option reflects that the switchers must first pay taxes on their gains, reducing the invested wealth to $[1+\alpha(1-\tau)]$, which becomes the new cost basis in the outside option. Note that total assets across vehicles are strictly less than the initial endowment due to tax leakage from switching.

The MF manager's optimization problem is to maximize total fee revenue:

$$V(f_{MF}) = \max_{f_{MF}} f_{MF} \cdot \mathrm{AUM}_{MF,t=1}(\alpha^*(f_{MF}, \delta), \lambda), \tag{15}$$

where $\alpha^*(f_{MF}, \delta)$ is the equilibrium cutoff from equation 12, δ is the distribution rate formalized in the next section, and λ is the probability of liquidity shocks.

Recall that the manager sets fees at $t = 0$ with knowledge of the distribution $P(\alpha)$ and liquidity shock probability λ, but cannot observe individual investor types (α_i, λ_i), which are revealed only at $t = 1$. The manager faces the following tension: higher fees increase per-dollar revenue but raise the cutoff α^*, encouraging more investors to exit. This trade-off is further complicated by the fund's liquidation strategy, which determines how redemptions translate into distributions δ. Strategies that generate larger distributions for a given level of redemptions amplify investor exit, constraining the fees the manager can charge. I formalize the distribution mechanism and liquidation strategies next.

5.3 Capital Gains Distributions

This section formalizes the distribution externality that links investor redemption decisions through strategic complementarities. When investors redeem their shares from the MF, the manager must sell appreciated securities to satisfy outflows, realizing unrealized gains. Because MFs are pass-through vehicles for tax purposes, these realized gains are distributed pro-rata to remaining investors in the model, creating taxable events regardless of whether remaining investors personally redeemed. This mechanism generates a feedback loop: redemptions force asset sales, which raise distributions, which increase the switching cutoff α^*, which triggers further redemptions.

Definition 1 (Capital Gains Distribution Rate) *The distribution rate* $\delta(\alpha^*, \lambda, \omega)$ *is the fraction of each remaining investor's MF position that is realized and distributed as taxable gains at* $t = 1$.

The distribution rate depends on three components: the redemption cutoff α^* (which determines who exits), the liquidity shock probability λ (which determines forced exits), and the liquidation strategy ω (which determines which assets are sold). Given these variables, the distribution rate is:

$$\delta(\alpha^*, \lambda, \omega) = \omega \cdot \frac{L(\alpha^*, \lambda)}{R(\alpha^*, \lambda)}, \tag{16}$$

where

$$L(\alpha^*, \lambda) = \lambda \int_{\underline{\alpha}}^{\bar{\alpha}} (1+\alpha)\, p(\alpha)\, d\alpha + (1-\lambda) \int_{\underline{\alpha}}^{\alpha^*} (1+\alpha)\, p(\alpha)\, d\alpha, \tag{17}$$

$$R(\alpha^*, \lambda) = (1-\lambda) \int_{\alpha^*}^{\bar{\alpha}} (1+\alpha)\, p(\alpha)\, d\alpha, \tag{18}$$

and ω is the realized gains ratio of the liquidated assets. The feasibility constraint requires

$$\delta(\alpha^*, \lambda, \omega) \leq 1.$$

The numerator $L(\alpha^*, \lambda)$ represents total MF liquidations: forced exits from liquidity shocks (λ fraction of all investors) plus voluntary switchers ($(1-\lambda)$ fraction with $\alpha_i < \alpha^*$). The denominator $R(\alpha^*, \lambda)$ is the asset base of investors who remain in the MF after exits. The ratio $L(\alpha^*, \lambda)/R(\alpha^*, \lambda)$ captures the redemption pressure: the amount that must be liquidated per dollar remaining.

As motivated by Fact 3, what matters for distributions is not only redemptions, but also that the fund has appreciated securities that need to be liquidated. In this model, this is captured by ω, the realized gains ratio. This ratio depends on both the amount of unrealized gains in the fund and the selling strategy the manager employs. As shown below, ω can vary with α^* depending on the liquidation strategy.

The distribution rate δ has a natural interpretation: $\delta = 0$ means no distributions, $\delta = 0.10$ means 10% of each remaining investor's position is distributed as taxable gains, and $\delta = 1$ corresponds to complete liquidation. Values $\delta > 1$ violate feasibility as investors cannot be distributed more than their holdings.

Liquidation Strategies

The realized gains ω depend critically on which portfolio securities the manager sells to meet redemptions. I consider two strategies that capture the range of realistic fund manager behavior.

Under proportional liquidation (prop), the manager sells all securities proportionally to their portfolio weights. This maintains tracking error at a minimum but realizes gains evenly across the investor unrealized gain distribution. The realized gains ratio is constant:

$$\omega^{Prop} = \frac{\int_{\underline{\alpha}}^{\bar{\alpha}} \alpha \cdot p(\alpha)\, d\alpha}{\int_{\underline{\alpha}}^{\bar{\alpha}} (1+\alpha) \cdot p(\alpha)\, d\alpha}. \tag{19}$$

For a uniform distribution on $[0, \bar{\alpha}]$, this ratio simplifies down nicely to $\omega^{Prop} = \bar{\alpha}/(2+\bar{\alpha})$. The distribution rate under proportional liquidation increases with redemptions solely through the ratio of those leaving to those remaining.

Under least-gains-first (LGF) liquidation, the manager prioritizes selling securities with the smallest unrealized gains. This minimizes near-term distributions, reducing the immediate tax burden on remaining investors. However, it concentrates the portfolio in highly appreciated assets, so subsequent redemptions must draw from positions with larger gains.

The realized gains ratio under least-gains-first is:

$$\omega^{LGF}(\alpha^*) = \frac{\int_{\underline{\alpha}}^{\alpha^*} \alpha \cdot p(\alpha)\, d\alpha}{\int_{\underline{\alpha}}^{\alpha^*} (1+\alpha) \cdot p(\alpha)\, d\alpha}. \tag{20}$$

Again with a uniform distribution, this yields a nice ratio of $\omega^{LGF}(\alpha^*) = \alpha^*/(2+\alpha^*)$ but it is dependent on α^*. The realized gains ratio is strictly increasing in α^*: as low-basis assets are depleted, the fund is forced to liquidate progressively higher-gain positions, amplifying the distribution rate. Appendix C provides the formal derivations under a uniform distribution of investor unrealized gains.

Appendix C.IV documents that index MFs behave approximately as proportional sellers (with a coefficient of 0.7 relative to perfect proportionality), while active MFs deviate substantially, consistent with tax-loss harvesting and selective realization strategies. This motivates analyzing both liquidation strategies in the model.

The liquidation strategy ω affects asset manager's trade-off indirectly: strategies that generate larger distributions (for a given α^*) further elevate the cutoff, amplifying outflows. The liquidation strategy ω is exogenously set at fund inception, reflecting the fund's investment mandate and operational constraints. The manager sets fees recognizing how ω shapes the distribution externality and thus investor responses.

Strategic Complementarities

The distribution rate function $\delta(\alpha^*, \lambda, \omega)$ embeds the strategic complementarities that drive fragility. The distribution rate being increasing and strictly convex are two critical characteristics:

Lemma 2 (Distribution Convexity) *The distribution rate $\delta(\alpha^*, \lambda, \omega)$ is strictly increasing and strictly convex in α^*:*

$$\frac{\partial \delta}{\partial \alpha^*} > 0, \qquad \frac{\partial^2 \delta}{\partial (\alpha^*)^2} > 0. \tag{21}$$

The increasing property is immediate: as more investors exit (higher α^*), investor liquidations rise while remaining investor assets fall, increasing the distribution rate. Figure 10 illustrates this decomposition: the left panel shows how $L(\alpha^*, \lambda)$ rises and $R(\alpha^*, \lambda)$ falls as α^* increases, while the right panel shows the resulting distribution rates under both liquidation strategies. The convexity is more subtle and arises from two sources. First, even under proportional liquidation with constant ω, the ratio of those who leave versus those who remain is convex in α^* because the numerator grows while the denominator shrinks, creating accelerating per-capita burden. Second, under least-gains-first liquidation, $\omega(\alpha^*)$ itself increases, further amplifying convexity as redemptions force sales of progressively higher-gain assets.

This convexity generates strategic complementarities among investor exit decisions. The strategic complementarity operates through the cross-partial of the switching value function:

$$\frac{\partial^2 g}{\partial \alpha_i \, \partial \alpha^*} = \frac{\partial^2 W^{MF}}{\partial \alpha_i \, \partial \delta} \cdot \frac{\partial \delta}{\partial \alpha^*} < 0. \tag{22}$$

Since $\partial \delta / \partial \alpha^* > 0$ (more exits raise distributions) and $\partial^2 W^{MF} / (\partial \alpha_i \, \partial \delta) < 0$ (higher distributions harm high-α investors more because they have more wealth exposed), the marginal benefit of staying declines as more investors exit. Moreover, because $\partial^2 \delta / \partial (\alpha^*)^2 > 0$, this decline accelerates: each additional redeemer makes staying less attractive at an increasing rate. This cross-partial condition defines strategic complementarities, where an investor's optimal decision depends negatively on the mass of exits by others, generating the possibility of coordination failures and multiple equilibria (Cooper and John, 1988).

Having characterized the distribution mechanism and its convexity properties, I now turn to equilibrium analysis.

5.4 Equilibrium Analysis

This section characterizes equilibrium by linking investor switching behavior and endogenous distributions. The central result is that strategic complementarities can generate multiple cutoffs for identical fund fundamentals: a stable equilibrium with limited switching and modest distributions, and an unstable run equilibrium with cascading redemptions and severe tax burdens. I first formalize the equilibrium condition, then characterize when multiple equilibria arise and their stability properties.

Equilibrium Characterization

Lemma 1 established that investors sort by a cutoff α^* given any distribution rate δ. In equilibrium, δ itself depends on redemptions induced by α^*, creating a fixed-point problem.

Lemma 3 (Equilibrium Fixed Point) *An equilibrium cutoff α^* must satisfy:*

$$g(\alpha^*; f_{MF}, f_{SMA}, \delta(\alpha^*, \lambda, \omega)) = 0, \tag{23}$$

where $g(\cdot)$ is the switching value function from equation (11) *and $\delta(\alpha^*, \lambda, \omega)$ is the endogenous distribution rate from Definition 1.*

This condition states that the marginal investor with gain α^* must be indifferent between staying and switching, given that all investors with $\alpha_i < \alpha^*$ switch and trigger distributions accordingly. The equilibrium set is:

$$\mathcal{A} = \left\{\alpha^* \in [\underline{\alpha}, \bar{\alpha}] : g(\alpha^*; f_{MF}, f_{SMA}, \delta(\alpha^*, \lambda, \omega)) = 0,\ \delta(\alpha^*, \lambda, \omega) \leq 1\right\}. \tag{24}$$

The feasibility constraint $\delta \leq 1$ rules out solutions where distributions exceed the fund's assets. In what follows, I take fees (f_{MF}, f_{SMA}) and liquidation strategy ω as given and characterize how distribution externalities generate multiple equilibria through strategic complementarities.

Multiple Equilibria

Strategic complementarities arise because the distribution rate $\delta(\alpha^*, \lambda, \omega)$ is increasing and convex in α^* (Lemma 2). As more investors exit, distributions imposed on those remaining grow disproportionately, strengthening the incentive to leave. This feedback can produce multiple equilibria where the switching value function $g(\alpha^*)$ crosses zero at several points.

Proposition 1 (Equilibrium Multiplicity) *For a given liquidation strategy, the equilibrium structure depends on the investor unrealized gains distribution $(\underline{\alpha}, \bar{\alpha})$, fee differential $\Delta f = f_{MF} - f_{SMA}$, and liquidity shock probability λ. The parameter space partitions into three regions:*

1. ***Outside Option Dominance**: For sufficiently large Δf relative to $(\bar{\alpha} - \underline{\alpha})$ and λ, a unique equilibrium exists with $\alpha^* = \underline{\alpha}$, where all non-liquidity-shocked investors exit to the outside option.*

2. ***Multiple Equilibria**: For intermediate parameter values, there exist two interior equilibria $\alpha_1^* < \alpha_2^*$ in $(\underline{\alpha}, \bar{\alpha})$ satisfying equation* (23)*:*

    - *Safe equilibrium (α_1^*): Limited switching and modest distributions (locally stable).*

- *Run equilibrium (α_2^*): Mass switching and severe distributions (unstable).*

3. ***MF Dominance**: For sufficiently large $\underline{\alpha}$ relative to Δf and sufficiently small λ, a unique equilibrium exists with $\alpha^* = \bar{\alpha}$, where all non-liquidity-shocked investors remain in the MF.*

See Appendix C.III for complete stability analysis but the intuition is as follows. The three regions reflect the interaction between fee differentials, tax lock-in, and distribution externalities. High unrealized gains create strong lock-in effects that anchor investors in the MF, while high fee differentials push investors toward the outside option. High liquidity shocks amplify distributions, destabilizing the MF.

The multiplicity region emerges when these forces are balanced: lock-in is strong enough to make staying attractive if distributions are low, but not strong enough to prevent exit if distributions surge. In this intermediate range, multiple self-fulfilling beliefs can be sustained. If investors expect few exits, distributions remain modest and staying is optimal. If investors expect mass exits, distributions surge and leaving is optimal.

When multiple equilibria exist, they differ in their stability properties. The safe equilibrium α_1^* is locally stable in that small perturbations in investor beliefs converge back to this equilibrium. The run equilibrium α_2^* is unstable, acting as a threshold where small shocks can push the system either back toward the safe equilibrium or forward into complete liquidation. The instability of the run equilibrium makes it a focal point for coordination failures: once investors expect switching to approach this threshold, the dynamics can tip into a self-fulfilling cascade.

In regions where multiple equilibria exist, the manager faces coordination uncertainty: the same fee structure can support both stable and run outcomes. Let $\rho \in [0, 1]$ denote the probability that investors coordinate on the run equilibrium. When the manager accounts for this uncertainty, expected revenue becomes:

$$V(f_{MF}) = \rho \cdot f_{MF} \cdot \mathrm{AUM}_{MF,t=1}(\alpha_2^*, \lambda) + (1 - \rho) \cdot f_{MF} \cdot \mathrm{AUM}_{MF,t=1}(\alpha_1^*, \lambda), \tag{25}$$

where α_1^* and α_2^* denote the stable and run equilibrium cutoffs characterized in Proposition 1.

In the baseline analysis, I set $\rho = 0$, meaning the manager optimizes fees assuming investors coordinate on the stable equilibrium. This simplification is an important first step: it clarifies how tax lock-in and distribution externalities shape optimal fees, establishes an upper bound on managerial pricing power (as incorporating coordination risk would lead managers to set lower fees), and provides a foundation for understanding coordination failures like the Vanguard case.

The framework naturally extends by relaxing two baseline assumptions. First, managers could account for coordination risk ($\rho > 0$), pricing more conservatively to mitigate the severity of potential runs. Second, managers could face uncertainty over the liquidity shock probability λ, similarly reducing fees to consider unexpectedly large outflows and subsequent distributions.

6 Comparative Statics

Having characterized the conditions under which multiple equilibria arise, I next examine how equilibrium outcomes vary with key parameters to identify which mutual funds are most exposed to fragility. The analysis begins with a baseline parameterization and then explores how equilibrium regions and outcomes change across the parameter space. I study how unrealized gains, liquidation strategies, liquidity shocks, and capital gains tax rates affect equilibrium cutoffs, distributions, and assets under management. The results clarify how managers set fees when investors face heterogeneous switching costs, how investor heterogeneity influences fragility, whether liquidation strategies can mitigate instability, and how policy and institutional shocks, particularly changes in capital gains tax rates and forced exits by tax-exempt retirement investors, affect fund stability.

The baseline parameters are shown in Table 7. The tax rate is set to 20%, reflecting the maximum statutory long-term capital gains rate for most investors. The expected return is 12%, consistent with average annual equity MF returns over the past decade. The liquidity shock probability is set to $\lambda = 0.05$, consistent with Dávila and Goldstein (2023).

The uniform distribution assumption is chosen for tractability and transparency: it yields a clear mapping from fund-level unrealized gain ratios to the investor-level distribution. The bound of unrealized gains $[\underline{\alpha}, \bar{\alpha}]$ is set with $\underline{\alpha} = 0$ and the upper bound $\bar{\alpha}$ is chosen to match key empirical patterns. For the comparative statics and counterfactual exercises that follow, I set $\bar{\alpha} = 0.88$ to align the model with the mean and standard deviation of fund-level unrealized gains, mean outflow ratios, mean distribution rates, and mean fees observed in the data. The parameterization procedure and model fit are detailed in Appendix C.V. This calibrated distribution corresponds to a fund-level unrealized gains ratio of approximately 31%.

To solve the model, I compute the safe equilibrium cutoff α_1^* and the optimal MF fee f_{MF} that maximizes revenue conditional on α_1^*, maintaining the assumption that the manager optimizes under the expectation that investors coordinate on the stable equilibrium. I then verify that multiple equilibria exist at these optimal fees: both α_1^* and α_2^* satisfy equation (23). This demonstrates that identical fund fundamentals and fee structures can support different outcomes depending solely on investor coordination.

6.1 Equilibrium Results and Comparative Statics

Table 8 presents equilibrium outcomes across various parameter configurations, showing cutoffs (α_1^*, α_2^*), assets remaining in the MF, distribution rates, and optimal fees.

The first row establishes the baseline calibration: $\tau = 0.20$, $R = 0.12$, $f_{SMA} = 0.0\%$, $\lambda = 0.05$, and $\bar{\alpha} = 0.88$. This specification generates two stable equilibria. In the safe equilibrium, the cutoff investor has $\alpha_1^* = 0.25$, the fund retains 73.6% of assets, and the distribution rate stands at 10.9%, supporting an optimal fee of 0.22%. In the run equilibrium, strategic exit accelerates: the cutoff drops to $\alpha_2^* = 0.33$, only 65.9% of assets remain, and the distribution rate surges to 15.8%, forcing the optimal fee down to 0.22%.[22] This baseline demonstrates the core fragility: identical fundamentals support both stable flows and self-reinforcing redemptions.

As a benchmark, consider the case with no distributions ($\delta = 0$), eliminating all strategic complementarities. In this setting, the MF can charge an optimal fee of approximately 1.09%, retaining 70% of AUM from investors. In the baseline model with distributions, conditioning on the stable equilibrium, the optimal fee is 0.54%.

The gap between these fees (1.09% vs 0.54%) illustrates the implicit cost of coordination risk. Even when the manager assumes investors coordinate on the stable equilibrium, the presence of distributions and thus the potential for runs constrains pricing power. This suggests that if managers incorporated explicit uncertainty over which equilibrium obtains, or over the liquidity shock probability λ, optimal fees would be even lower. The no-distribution benchmark thus provides an upper bound on MF pricing power absent coordination frictions and demonstrates how the distribution externality constrains fees. I next examine how these outcomes vary systematically with fund characteristics, specifically the distribution of unrealized gains and the manager's liquidation strategy.

[22]The optimal fees are identical across equilibria in this specification because the manager conditions on the safe equilibrium when setting fees. In practice, managers facing coordination uncertainty might set lower fees as insurance against runs.

6.2 Equilibrium Regions

Proposition 1 establishes that the model can generate one or two equilibria depending on fund and investor characteristics. The number of equilibria depends on two forces pulling in opposite directions: the degree of investor lock-in, and the MF fee disadvantage relative to the outside option. When lock-in is sufficiently strong relative to the fee disadvantage, all investors prefer to remain in the MF and a unique stable equilibrium exists. When the fee disadvantage is sufficiently large relative to lock-in, all investors, regardless of their lock-in, prefer the outside option and again a unique equilibrium exists. Between these extremes, strategic complementarities generate two interior equilibria: a stable equilibrium with limited exits and a run equilibrium with cascading redemptions.

A key feature of the model is that it distinguishes between fund-level unrealized gains and investor-level heterogeneity in tax positions. Investor-level unrealized gains allow me to characterize the dispersion of tax lock-in across the fund's investor base, rather than treating all investors as facing identical exit costs. Two funds with the same aggregate unrealized gains may face different stability depending on whether gains are concentrated among a few investors or dispersed uniformly across the base.

To explore this distinction, I fix fund-level unrealized gains at 20% and vary the dispersion in investor tax positions.[23] Figure 11 maps the equilibrium structure across this parameter space. The vertical axis measures dispersion which is the degree to which unrealized gains vary across the fund's investor base. At zero dispersion, all investors hold identical unrealized gains; higher dispersion indicates greater heterogeneity, which I operationalize as lower minimum investor unrealized gains while maintaining the fund's average at 20%. The horizontal axis shows the MF fee disadvantage relative to the zero-fee alternative vehicle.

The dashed vertical line marks a key empirical benchmark: the average fee differential (0.17%) between MFs and their ETF twins, calculated from Morningstar data. I compare MFs to ETFs rather than separately managed accounts because ETF twins are a more passive strategy and thus track the same underlying portfolio as their MF counterparts, holding investment strategy fixed and isolating the pure fee wedge attributable to the fund structure. The figure shows that the empirical fee differential falls in the model's multiple equilibria region across all levels of investor heterogeneity.

The colored regions partition the parameter space into three equilibrium types. The blue

[23] 20% is chosen as it matches the average shown in Figure 6a.

region represents MF dominance: MFs' tax deferral benefits outweigh their fee disadvantage, leading all investors to prefer the MF structure. The green region shows the outside option dominance: fees are sufficiently high or tax lock-in sufficiently weak that all investors prefer the outside option. Between these regions lies the purple zone of multiple equilibria, where strategic complementarities in investor exit decisions generate coordination problems meaning the same fundamentals can support either stable or run equilibrium.

Unrealized Gains

The top panel of Table 8 demonstrates how investor heterogeneity in unrealized gains affects equilibrium outcomes. Higher average unrealized gains create stronger lock-in effects. Increasing the distribution's upper bound from 0.88 to 5.0 allows the fund to retain an additional 9.6 percentage points of assets in the safe equilibrium and 8.9 percentage points in the run equilibrium. Greater unrealized gains also enable the manager to charge higher optimal fees, which increase from 0.23% to 1.05%. However, while higher unrealized gains provide protection against runs, they do not eliminate the multiple-equilibria structure. Throughout the specifications, both safe and run outcomes remain possible across all unrealized gains distributions, with the key difference being the severity of outcomes rather than their existence.

Liquidation Strategy

Table 9 compares equilibrium outcomes under proportional and least-gains-first liquidation across different distribution widths. With the narrow baseline distribution ($\bar{\alpha} = 0.88$), least-gains-first reduces distribution rates in the safe equilibrium (from 11.0% to 5.1%) but achieves only a modest improvement in asset retention (72.4% vs 73.6% under proportional liquidation). This limited advantage reflects the fact that with such a narrow distribution of gains, nearly all securities have similar unrealized gains, leaving little scope for selective liquidation to reduce distributions.

In contrast, with wider distributions of unrealized gains ($\bar{\alpha} = 2.0$), the advantage of least-gains-first becomes pronounced. Distribution rates fall from 14.4% to 7.6%, and asset retention increases from 77.6% to 78.4%. The larger pool of low-basis securities means least-gains-first can meet substantial redemptions by selectively selling the lowest-basis assets first, substantially reducing distributions relative to proportional liquidation.

While least-gains-first generates lower distribution rates at any given α^* (Figure 10), the relevant comparison is the run equilibrium cutoff α_2^*. A lower α_2^* under least-gains-first means that a run equilibrium exists at a smaller mass of exits: fewer investors need to coordinate on leaving to make exit optimal for all remaining investors. In this sense, least-gains-first lowers the threshold at which a coordination failure becomes self-fulfilling: fewer investors need to coordinate on leaving to make exit optimal for all remaining investors, meaning a run occurs at a smaller mass of exits. Finally, the optimal fee increases under least-gains-first (from 0.54% to 0.68%), reflecting the manager's ability to extract higher rents when distributions are suppressed in the safe equilibrium.

Alternative Tax Rates

The current U.S. tax code imposes a top statutory long-term capital gains rate of 20%. Recent policy debates have considered a wide range of capital gains tax rates, from proposals to lower the top rate to around 15% to proposals that would raise it to nearly 40%, making the effect of tax rates on MF stability a policy-relevant question.[24] Rather than focusing on a particular reform, I vary the long-term capital gains rate τ from 10% to 40% and study the effects on equilibria, AUM retention, and distribution rates. The tax rate enters the model in two ways: it raises the cost of exit for investors who switch, and it magnifies the tax burden of distributions for those who remain.

Figure 12 shows three results. Panel (a) demonstrates that higher τ lowers the safe cutoff and increases the run cutoff. This reflects stronger investor lock-in: more investors remain in the MF in the safe equilibrium, and the run cutoff is pushed toward the boundary case of complete exit. Panel (b) shows that in run states, AUM retention falls more steeply at higher τ, consistent with deeper outflows once a run is triggered. Panel (c) shows that distribution rates rise sharply with τ in the run equilibrium, magnifying the tax burden on remaining investors.

Taken together, these results show that higher capital gains taxes push the two equilibria apart: the safe equilibrium moves toward fewer exits while the run equilibrium moves toward more exits. This polarization means that at higher tax rates, funds are either stable with few exits or experience a full run.

[24]For example, the Biden Administration's FY2025 budget proposed taxing long-term capital gains at ordinary income rates for households with income above \$1 million, implying a top statutory rate of 39.6%. In contrast, policy discussions associated with the Project 2025 initiative have included proposals to maintain or reduce capital gains tax rates relative to current law, with some discussions referencing rates around 15%.

Retirement Shocks

The second counterfactual focuses on retirement investors, who defer the realization of capital gains through retirement accounts and are therefore insulated from the distribution externality. Retirement investors are natural first movers when cheaper alternatives arise, for example through changes to the retirement plan menu. I model this by varying the probability of liquidation shock λ, which captures the fraction of investors forced to leave the fund. In the baseline model, these forced exits represent liquidity shocks where investors consume their proceeds. Here, they represent retirement investors who exit to switch to a lower-fee alternative. From the fund's perspective, both cases are equivalent: a fraction λ of investors redeem shares, forcing the fund to sell assets and realize gains that are distributed to remaining taxable shareholders.

The Vanguard Target Retirement Fund episode illustrates this mechanism. When Vanguard lowered the minimum investment of a lower-fee institutional fund, retirement investors exited en masse. The shift produced outflows of roughly 15% of AUM, which forced sales of appreciated assets and triggered large taxable distributions for those who remained. I calibrate the model to this case by setting the MF fee to 0.14% and the alternative vehicle fee to 0.08%.

Figure 13 demonstrates that as λ increases, the safe equilibrium shrinks, AUM retention falls, and distributions rise. Panel (a) shows both cutoffs increasing with λ, compressing the range of investors who remain. Panel (b) shows AUM retention declining in both equilibria, with the safe and run outcomes converging as forced exits dominate. Panel (c) shows distribution rates rising sharply, particularly in the safe equilibrium where the fund was previously stable. The calibration aligns closely with the Vanguard case: distributions reach approximately 10% when outflows hit 15%. While Vanguard ultimately avoided collapse by merging funds, the figure illustrates how retirement shocks can destabilize MFs in the absence of intervention.

7 Conclusion

This paper explains MF persistence and fragility in the face of competition from SMAs and ETFs by focusing on one central channel: taxes as a switching cost. Using new data on unrealized gains, I show that unrealized gains lock investors into MFs by raising the cost of exit, but also amplify capital gains distributions when redemptions occur. This lock-in enables MF managers to sustain higher fees. My model formalizes how these dynamics generate strategic complementarities in redemption decisions, producing multiple equilibria: a stable outcome with modest outflows and a fragile outcome with cascading redemptions. Counterfactual analyses highlight that higher tax rates on capital gains strengthen lock-in while retirement shocks can rapidly destabilize funds by shifting the tax burden onto taxable investors.

The rise of SMAs introduces an additional dimension beyond the scope of this paper: active tax-loss harvesting. While this analysis focuses on avoiding distributions from others' redemptions, SMAs can also generate tax alpha by strategically realizing losses to offset gains, a capability unavailable in pooled vehicles. As technology reduces SMA investment minimums and makes tax-loss harvesting more accessible, this enhanced tax efficiency may further intensify competitive pressure on MFs. Understanding how tax-loss harvesting interacts with the lock-in and fragility mechanisms documented here represents an important direction for future research on the evolution of investment vehicles.

References

Agarwal, V., H. Ren, K. Shen, and H. Zhao (2022): "Redemption in Kind and Mutual Fund Liquidity Management," *The Review of Financial Studies*, 36(6), 2274–2318.

Barclay, M. J., N. D. Pearson, and M. S. Weisbach (1998): "Open-end mutual funds and capital-gains taxes," *Journal of Financial Economics*, 49(1), 3–43.

Beggs, W., A. Hill-Kleespie, and A. Y. Liu (2022): "Mutual fund tax implications when investment advisors manage tax-exempt separate accounts," *Journal of Banking and Finance*, 134.

Ben-David, I., F. Franzoni, and R. Moussawi (2017): "Exchange-Traded Funds," *Annual Review of Financial Economics*, 9(Volume 9, 2017), 169–189.

Bergstresser, D., and J. Poterba (2002): "Do after-tax returns affect mutual fund inflows?," *Journal of Financial Economics*, 63(3), 381–414.

Chaudhuri, S. E., T. C. Burnham, and A. W. Lo (2020): "An Empirical Evaluation of Tax-Loss-Harvesting Alpha," *Financial Analysts Journal*, 76(3), 99–108.

Chen, F., L.-W. Chen, H. Johnson, and S. Sardarli (2017): "Tailored versus Mass Produced: Portfolio Managers Concurrently Managing Separately Managed Accounts and Mutual Funds," *The Financial Review*, 52(4), 531–561.

Chen, Q., I. Goldstein, and W. Jiang (2010): "Payoff complementarities and financial fragility: Evidence from mutual fund outflows," *Journal of Financial Economics*, 97(2), 239–262.

Chernenko, S., and A. Sunderam (2020): "Liquidity transformation in asset management: Evidence from the cash holdings of mutual funds," *Working Paper*.

Chordia, T. (1996): "The structure of mutual fund charges," *Journal of Financial Economics*, 41(1), 3–39.

Colon, J. M. (2017): "The Great ETF Tax Swindle: The Taxation of In-Kind Redemptions," *Penn State Law Review*, 122(1), 1–54.

Constantinides, G. M. (1983): "Capital Market Equilibrium with Personal Tax," *Econometrica*, 51(3), 611–636.

Cooper, R., and A. John (1988): "Coordinating Coordination Failures in Keynesian Models*," *The Quarterly Journal of Economics*, 103(3), 441–463.

Coval, J., and E. Stafford (2007): "Asset fire sales (and purchases) in equity markets," *Journal of Financial Economics*, 86.

Dammon, R. M., C. S. Spatt, and H. H. Zhang (2001): "Optimal Consumption and Investment with Capital Gains Taxes," *The Review of Financial Studies*, 14(3), 583–616.

——— (2004): "Optimal Asset Location and Allocation with Taxable and Tax-Deferred Investing," *The Journal of Finance*, 59(3), 999–1037.

Diamond, D. W., and P. H. Dybvig (1983): "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, 91(3), 401–419.

Dickson, J. M., J. B. Shoven, and C. Sialm (2000): "Tax Externalities of Equity Mutual Funds," *National Tax Journal*, 53(3.2), 607–628.

Domar, E. D., and R. A. Musgrave (1944): "Proportional Income Taxation and Risk-Taking," *The Quarterly Journal of Economics*, 58(3), 388–422.

Dávila, E., and I. Goldstein (2023): "Optimal Deposit Insurance," *Journal of Political Economy*, 131(7), 1676–1730.

Elton, E., M. Gruber, and C. Blake (2013): "The Performance of Separate Accounts and Collective Investment Trusts," *Review of Finance*, 18(5), 1717–1742.

Fedyk, V. (2023): "Where Are the Sophisticated Investors? Evidence From Separate Accounts," *Working Paper*.

Feldstein, M., J. Slemrod, and S. Yitzhaki (1980): "The Effects of Taxation on the Selling of Corporate Stock and the Realization of Capital Gains," *The Quarterly Journal of Economics*, 94(4), 777–791.

Goldstein, I., H. Jiang, and D. T. Ng (2017): "Investor flows and fragility in corporate bond funds," *Journal of Financial Economics*, 126(3), 592–613.

Helmke, A. (2024): "Will ETFs Drive Mutual Funds Extinct?," *Working Paper*.

Holden, S., D. Schrass, and M. Bogdan (2024): "Ownership of Mutual Funds and Shareholder Sentiment, 2024," ICI Research Perspective Vol. 30, No. 8, Investment Company Institute, Washington, DC.

Hortacsu, A., and C. Syverson (2004): "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of the S&P 500 Index Funds," *The Quarterly Journal of Economics*, 119(2), 403–456.

Ivković, Z., J. Poterba, and S. Weisbenner (2005): "Tax-Motivated Trading by Individual Investors," *American Economic Review*, 95(5), 1605–1630.

Ivković, Z., and S. Weisbenner (2009): "Individual investor mutual fund flows," *Journal of Financial Economics*, 92(2), 223–237.

Jiang, H., Z. Sun, and A. Wang (2022): "Does mutual fund illiquidity introduce fragility into asset prices? Evidence from the corporate bond market," *Journal of Financial Economics*, 143, 277–302.

Khang, K., T. Paradise, and J. Dickson (2021): "Tax-Loss Harvesting: An Individual Investor's Perspective," *Financial Analysts Journal*, 77(4), 128–150.

Lettau, M., and A. Madhavan (2018): "Exchange-Traded Funds 101 for Economists," *Journal of Economic Perspectives*, 32(1), 135–54.

Ma, Y., K. Xiao, and Y. Zeng (2022): "Mutual Fund Liquidity Transformation and Reverse Flight to Liquidity," *Review of Financial Studies*, 35(10).

Mainardi, F. (2025): "The Impact of Fiscal Policy on Financial Institutions, Asset Prices, and Household Behavior," *Working Paper*.

Miller, M. H. (1986): "Financial Innovation: The Last Twenty Years and the Next," *The Journal of Financial and Quantitative Analysis*, 21(4), 459–471.

Morris, S., and H. S. Shin (1998): "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *The American Economic Review*, 88(3), 587–597.

Moussawi, R., K. Shen, and R. Velthuis (2025): "The Role of Taxes in the Rise of ETFs," *The Review of Financial Studies*, p. hhaf044.

Pastor, L., R. F. Stambaugh, and L. A. Taylor (2015): "Scale and skill in active management," *Journal of Financial Economics*, 116(1), 23–45.

Poterba, J. M., and J. B. Shoven (2002): "Exchange-Traded Funds: A New Investment Option for Taxable Investors," *American Economic Review*, 92(2), 422–427.

Rohleder, M., H. Tentesch, R. Weh, and M. Wilkens (2022): "Fraternal twins—Should investors be careful?," *Review of Financial Economics*.

Sialm, C., and L. Starks (2012): "Mutual Fund Tax Clienteles," *The Journal of Finance*, 67(4), 1397–1422.

Sialm, C., L. Starks, and H. Zhang (2015): "Defined Contribution Pension Plans: Sticky or Discerning Money?," *The Journal of Finance*, 70(2), 805–838.

Sialm, C., and H. Zhang (2020): "Tax-Efficient Asset Management: Evidence from Equity Mutual Funds," *The Journal of Finance*, 75(2), 735–777.

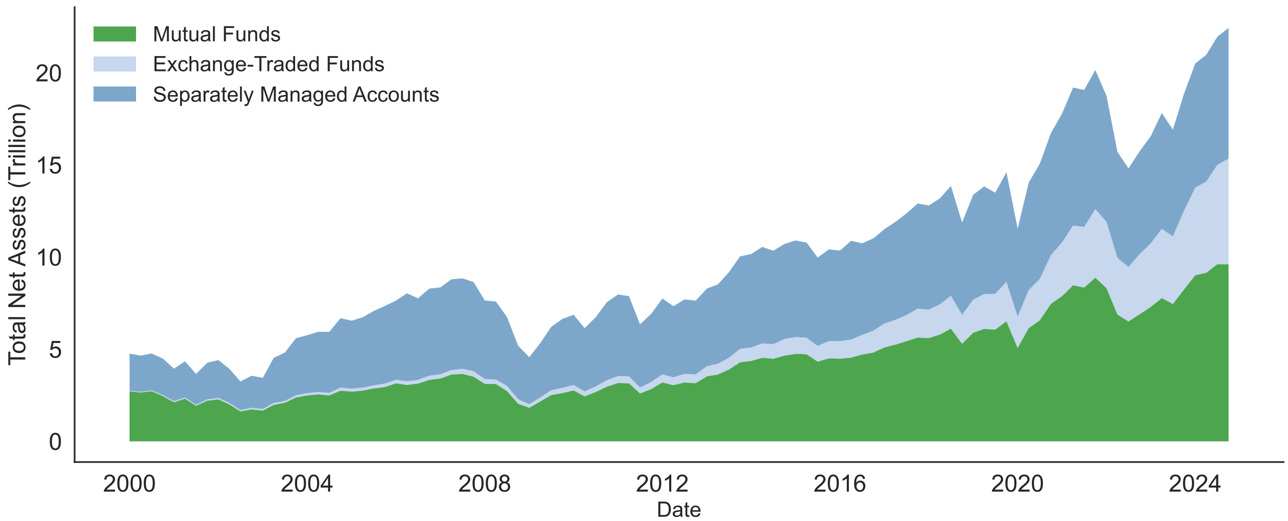
Stiglitz, J. E. (1983): "Some aspects of the taxation of capital gains," *Journal of Public Economics*, 21(2), 257–294.

Yang, H. (2023): "What Determines 401(k) Plan Fees? A Dynamic Model of Transaction Costs and Markups," *Working Paper*.

Zhang, X., T. Guo, Y. Cheng, and H. Wang (2024): "Which Should Be Your Top Pick, Separately Managed Accounts or ETFs?," *Journal of Risk and Financial Management*, 17(5).

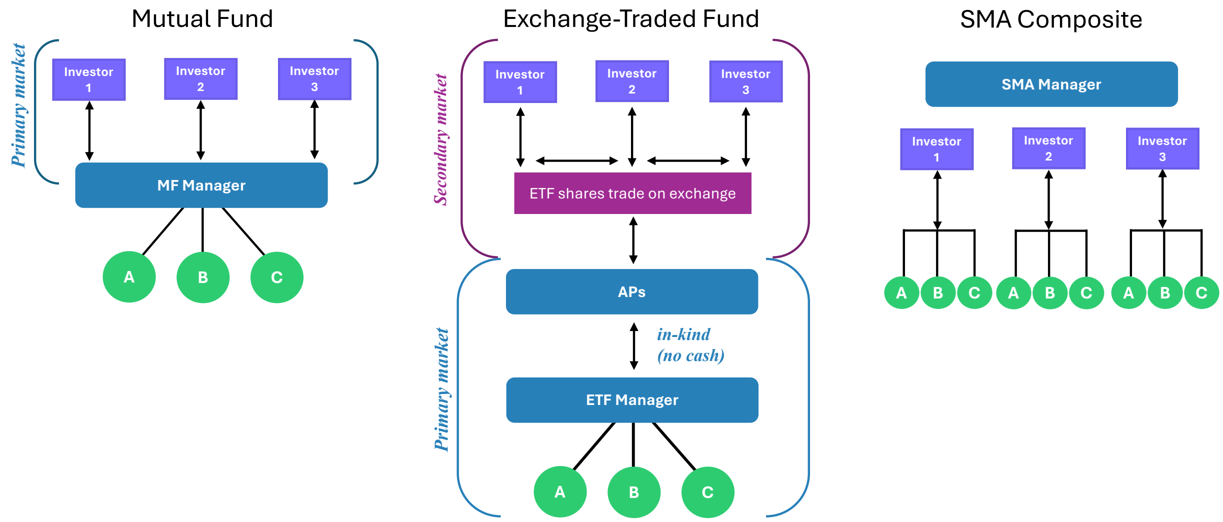
8 Figures

Figure 1. US Equity Total Net Assets by Investment Vehicle



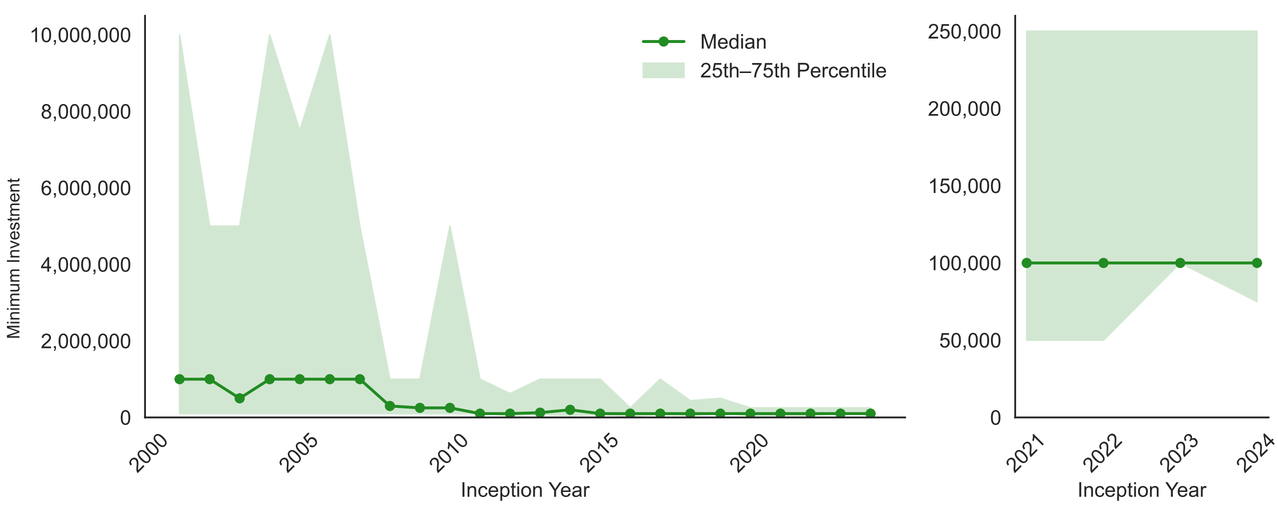
Note: The figure shows the stacked area of US Equity MF, ETF, and SMA total net assets from 2000 to 2024. For MFs, the sample excludes share classes denoted as retirement using the Share Class Type variable.
Source: Morningstar Direct

FIGURE 2. Investment Vehicle Structure



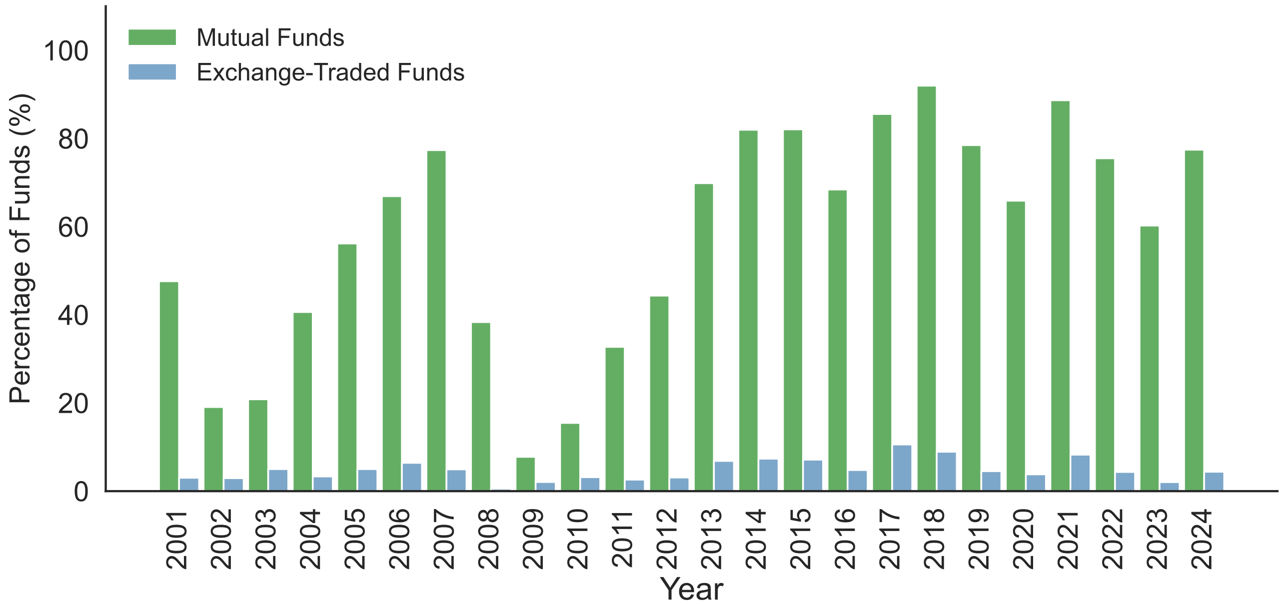
Note: This figure compares the structure of MFs, ETFs, and SMAs. In a MF, investors transact directly with the fund manager, receiving shares representing a claim on the pooled portfolio, and redemptions require the fund to sell underlying securities, potentially realizing taxable gains. In an ETF, individual investors trade shares on a secondary market with other investors or authorized participants (APs), without interacting directly with the fund manager or underlying securities. APs transact with the ETF manager in the primary market through in-kind exchanges, preventing the realization of taxable gains upon redemption. An SMA is a professionally managed account in which the investor directly owns the underlying securities, insulating each investor from the redemption decisions of others.

FIGURE 3. SMA Minimum Investments



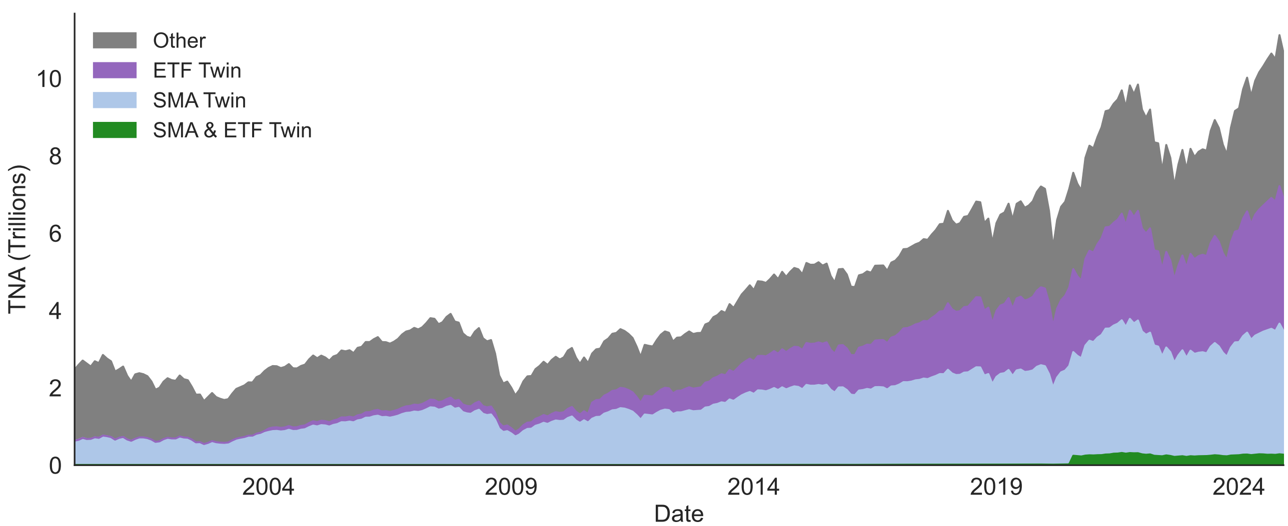
Note: This figure shows the minimum investment of SMAs from 2000 to 2024 for US Equity SMAs. The dark green line shows the median while the light green ranges show the inter-quartile range. The figure is plotted by SMA inception year meaning the minimum investment if the SMA began that year. The right figure zooms to 2021 to 2024 for clarity.
Source: Morningstar Direct

Figure 4. Percent of Equity MFs and ETFs with Capital Gains Distribution



Note: The figure shows the annual percentage of US Equity MFs or ETFs that made capital gains distributions from 2001 to 2024. Capital gains distributions includes short-term and long-term capital gains distributions.
Source: Morningstar Direct

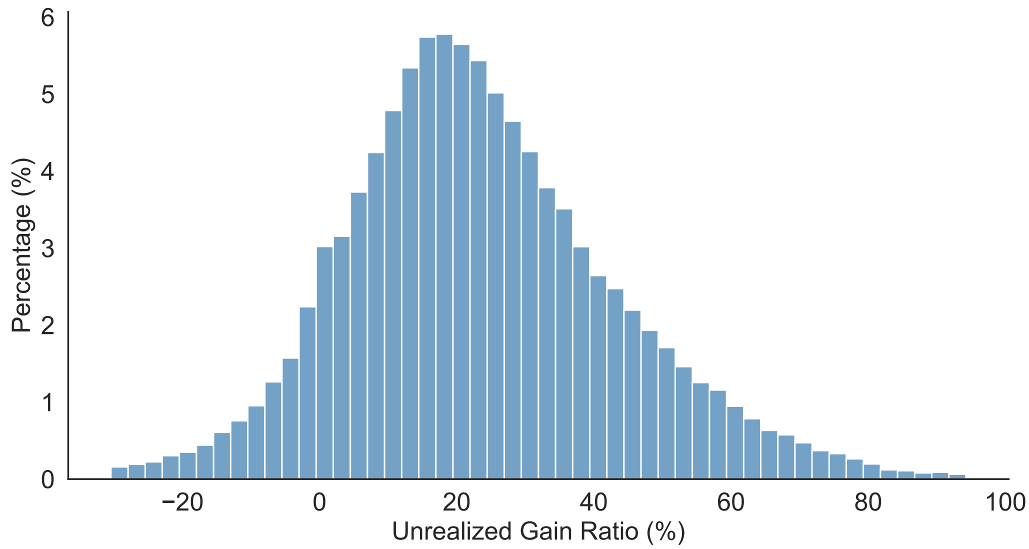
Figure 5. MF Assets by ETF or SMA Twin Strategy



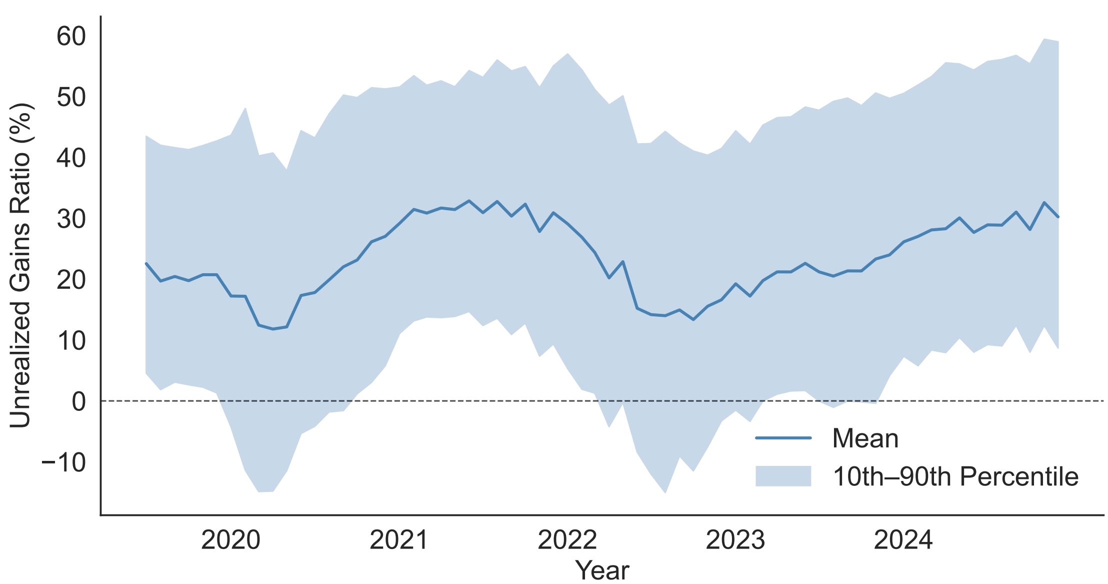
Note: The figure shows the monthly total net assets of US Equity MFs from 2000 to 2024. The plot is grouped by whether the MF has a twin ETF or SMA. A twin is defined by having the same strategy defined by the StrategyId variable in Morningstar.
Source: Morningstar Direct

Figure 6. Equity MFs Unrealized Gains Ratio

(a) Unrealized Gains Ratio Distribution

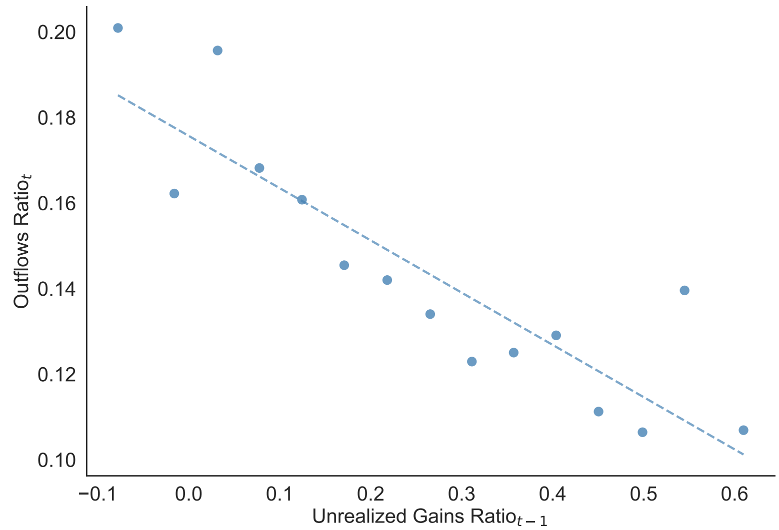


(b) Unrealized Gains Ratio Over Time



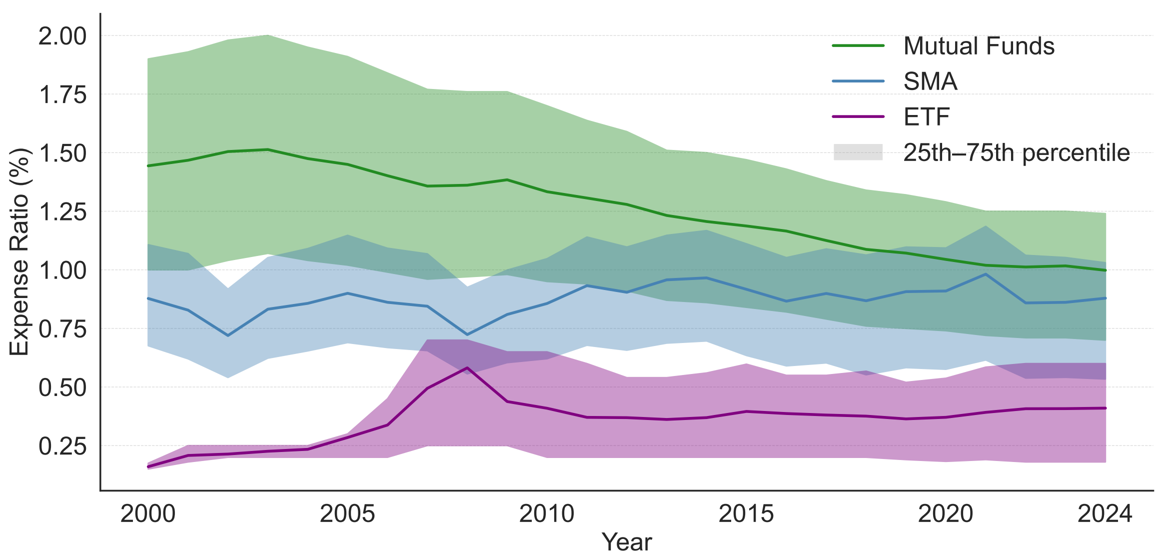
Note: The figures depicts the unrealized gains ratio as a distribution of MFs. Panel (a) shows the panel distribution across time and fund of the unrealized gain ratio. Panel (b) shows the mean ratio along with the 10th and 90th percentiles.
Source: Source: SEC N-PORT and N-CSR

FIGURE 7. Unrealized Gains and Outflows of Equity MFs



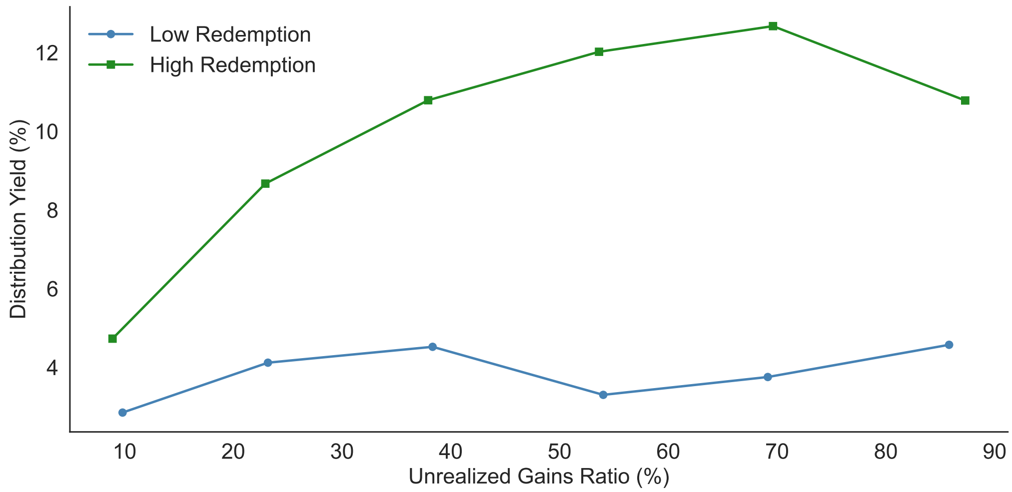
Note: The figure plots the binned relationship between lagged unrealized gains ratios and outflow ratios for MFs. The x-axis shows the lagged unrealized gains ratio, and the y-axis represents the outflows ratio. The scatter points show bin-averaged values across 15 equal-sized bins of unrealized gains. Dashed lines show the fitted linear relationship within each group.
Source: SEC N-PORT, SEC N-CSR, and Morningstar Direct

FIGURE 8. Investment Vehicle Expense Ratios Over Time



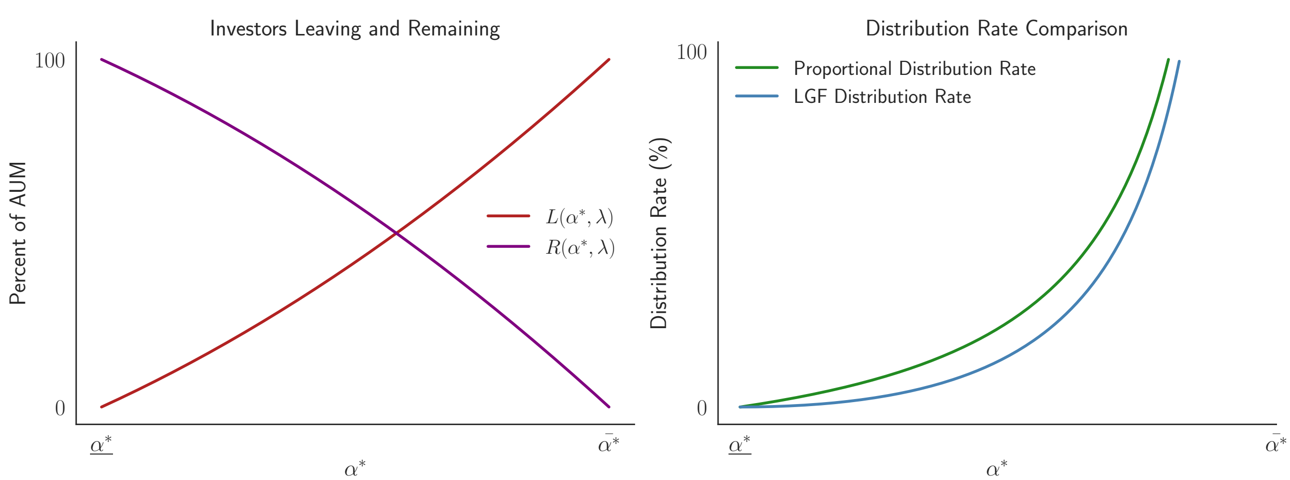
Note: The figure shows the average expense ratio for MF, ETF, and SMAs from 2000 to 2024. The shaded area shows the 25th to 75th percentiles. Active is denoted by the Index Fund variable in Morningstar. SMA expense ratios are computed as the difference between the gross and net return.
Source: Morningstar Direct

Figure 9. Unrealized Gains, Redemptions, and Distribution Yields



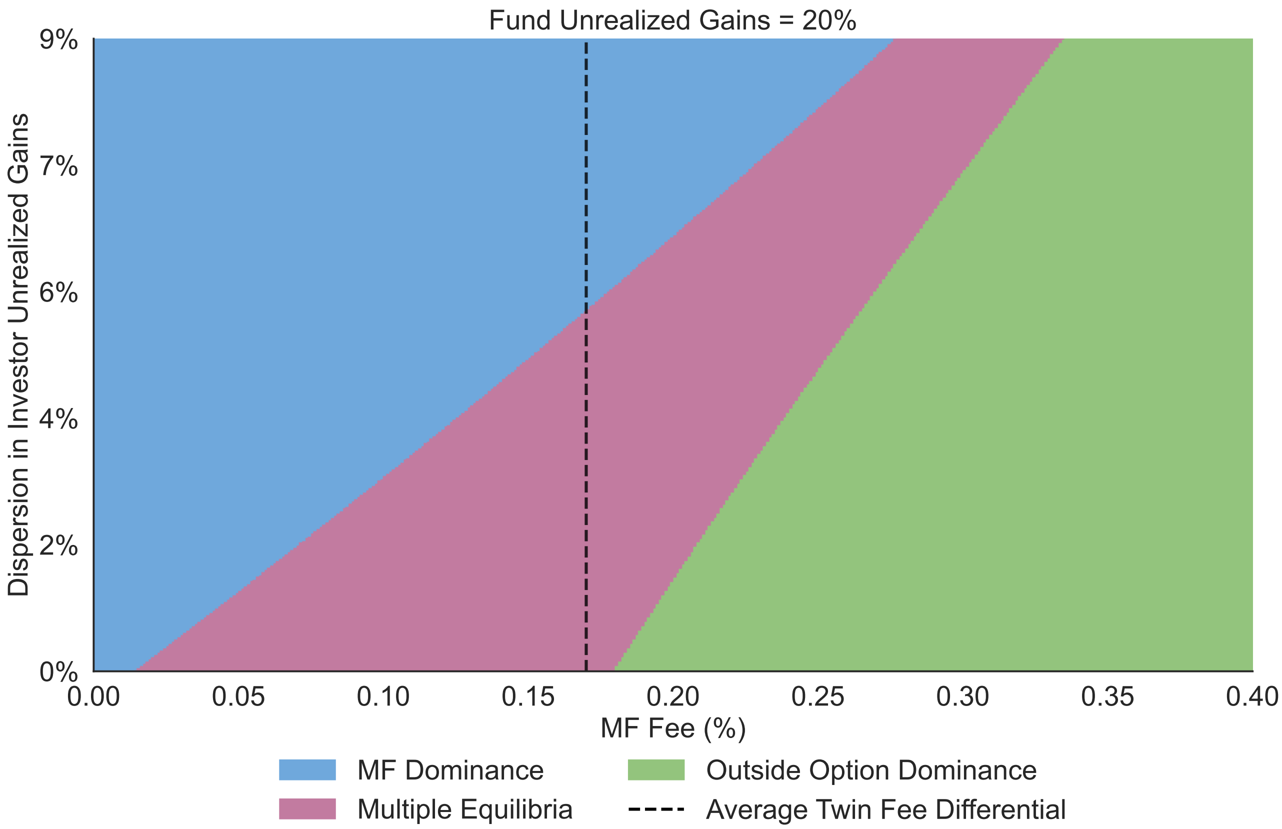
Note: This figure plots the relationship between unrealized gains ratios and distribution yields for equity mutual funds, differentiated by outflow behavior. The unrealized gains ratio is defined as gross unrealized gains relative to total net assets. Distribution yield is calculated as the distribution amount in dollars as a percentage of the fund's net asset value. Funds are classified into low and high redemption groups based on tertile splits of outflow ratios. Data points represent bin means, where bins are constructed using six equal-width intervals spanning from the minimum to maximum unrealized gains ratio among funds with non-negative unrealized gains
Source: SEC N-PORT and N-CSR, CRSP, and Morningstar Direct

Figure 10. Distribution Dynamics: Composition of Remaining Investors and Resulting Distribution Rates



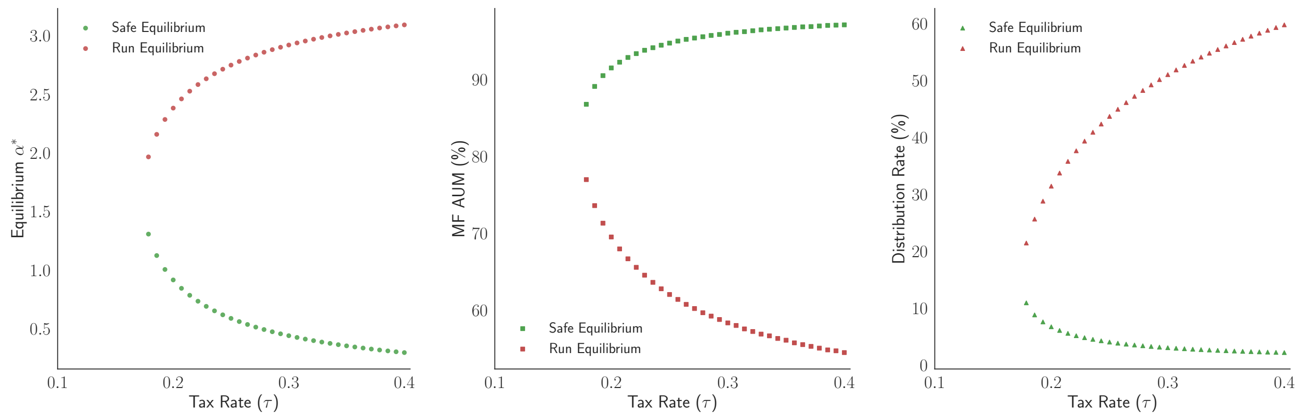
Note: The figure shows the decomposition of the distribution rate. The left panel shows the leaving and remain assets based on the cutoff (α^*). The right panel shows the distribution rate for both liquidation strategies: least-gains first and proportional selling. Illustration under the guiding parameterization ($\tau = 0.20$, $f_{MF} = 0.01$, $f_{SMA} = 0\%$, $R = 12\%$, $\alpha \sim U[0, 5]$).

Figure 11. Equilibrium Regions in Parameter Space



Note: Equilibrium regions for fund-level unrealized gains of 20%. The y-axis shows dispersion in investor tax positions, where 0% indicates homogeneity (all investors identical) and higher values indicate greater heterogeneity. Dispersion corresponds to minimum investor unrealized gains where investors are uniformly distributed over $\alpha \sim U[\alpha_{min}, \alpha_{max}]$ with bounds maintaining the 20% fund average. For each point, I solve $0 = g(\alpha^*; f_{MF}, f_{SMA}, \lambda, \omega)$. Dashed line shows average twin fee differential (0.17%) for ETF twin funds. Parameters: $\tau = 0.20$, $f_{SMA} = 0\%$, $R = 12\%$, $\pi = 1\%$. Results are robust to alternative grid resolutions and convergence tolerances.

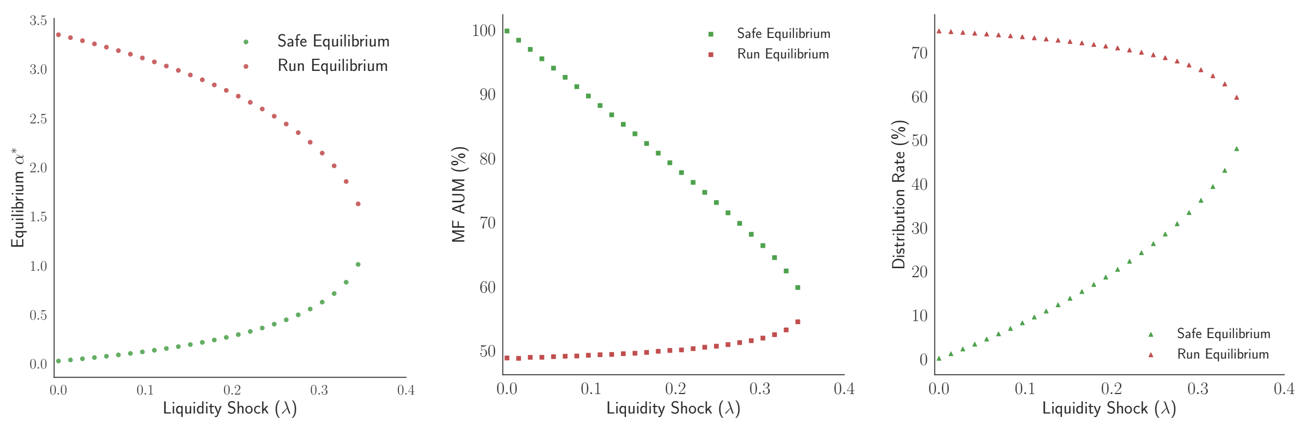
FIGURE 12. Alternative Tax Rates



(a) Safe and Run Equilibria **(b)** AUM Retention **(c)** Distribution Rates

Note: The figure reports policy counterfactuals under alternative long-term capital gains tax rates τ, holding all other model primitives fixed at their baseline values. Panel A shows the values of the safe and run equilibria α^* as the tax rate varies. Panel B shows the corresponding share of MF assets under management (AUM) retained in equilibrium. Panel C shows the capital gains distribution rates at each equilibrium. Other values are: $\tau = 0.20$, gross return $R = 0.12$, MF fee $f_{MF} = 1\%$, SMA fee $f_{SMA} = 0\%$, liquidity shock probability $\lambda = 0.01$, proportional liquidation, and an embedded gains distribution $\alpha \sim U[0, 5]$. Beyond the range where two equilibria are plotted, only the run equilibrium exists at the boundary of the investor gains distribution ($\bar{\alpha}$).

FIGURE 13. Alternative Liquidity Shocks



(a) Safe and Run Equilibria **(b)** AUM Retention **(c)** Distribution Rates

Note: The figure reports counterfactuals under alternative retirement shock probabilities λ, holding all other model primitives fixed at their baseline values. Panel A shows the values of the safe and run equilibria α^* as the tax rate varies. Panel B shows the corresponding share of MF assets under management (AUM) retained in equilibrium. Panel C shows the capital gains distribution rates at each equilibrium. Other values are: tax rate $\tau = 0.20$, gross return $R = 0.12$, mutual fund fee $f_{MF} = 0.14\%$, SMA fee $f_{SMA} = 0.08\%$, proportional liquidation, and an embedded gains distribution $\alpha \sim U[0, 5]$. Beyond the range where two equilibria are plotted, only the run equilibrium exists at the boundary of the investor gains distribution ($\bar{\alpha}$).

9 Tables

Table 1. Sample Summary Statistics

	Mean	SD	P25	Median	P75
Distribution yield (%)	5.60	6.27	0.05	3.90	8.44
Outflows ratio (%)	15.27	16.34	6.36	11.25	18.47
Unrealized gains ratio (%)	25.60	19.04	13.02	23.33	36.65
Return (%)	12.73	19.31	3.93	15.86	25.52
Alpha (annualized, %)	-2.06	6.08	-5.99	-1.74	1.79
Turnover (%)	37.84	24.20	18.00	34.00	54.87
Net expense ratio (%)	0.80	0.38	0.60	0.85	1.02
Gross expense ratio (%)	0.90	0.46	0.66	0.91	1.11
Fee waiver (%)	8.58	15.05	0.00	1.12	11.13
Cash holdings (%)	2.04	2.45	0.60	1.43	2.61
TNA (millions USD)	6299.37	38852.57	209.64	806.77	2866.31
Fund age (years)	22.94	13.74	13.76	22.25	28.62

Note: The table reports summary statistics for the mutual fund sample. Ratios are expressed in percent for presentation, but enter the regressions in decimal form (e.g., 0.154 = 15.4). Fund age is measured in years and TNA is reported in millions of U.S. dollars. All variables are winsorized at the 1st and 99th percentiles.
Source: Morningstar Direct, SEC N-CSR, and SEC N-PORT.

TABLE 2. Unrealized Gains and Outflows

	Dependent variable: Outflow Ratio					
	Full Sample				Twin Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Unrealized Gain Ratio$_{t-1}$	-0.113***	-0.055***	-0.055***	-0.059***	-0.050**	-0.055***
	(0.013)	(0.016)	(0.017)	(0.017)	(0.020)	(0.020)
Alpha$_{t-1}$			-0.091***		-0.102***	
			(0.031)		(0.037)	
Return$_{t-1}$				-0.046***		-0.068***
				(0.016)		(0.019)
Observations	4235	4235	4235	4235	2558	2558
R^2	0.027	0.068	0.074	0.073	0.075	0.076
Time FE	No	Yes	Yes	Yes	Yes	Yes
Style FE	No	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes

Note: The table reports results from estimating Equation 1:

$$\text{Outflows}_{it} = \beta_1\, \text{UnrealizedGains}_{i,t-1} + \beta_2\, \text{Performance}_{i,t-1} + \gamma' \mathbb{X}_{it} + \mathcal{S}_i + \delta_t + \varepsilon_{it},$$

where i indexes mutual funds and t denotes years. The dependent variable Outflows$_{it}$ is the absolute value of annual outflows, computed as the sum of negative monthly flows scaled by lagged total net assets. UnrealizedGains$_{i,t-1}$ is the lagged unrealized gains ratio, defined as gross unrealized gains divided by total net assets. Performance$_{i,t-1}$ is measured either as the prior-year return or as the Fama–French three-factor alpha. The control vector $\mathbb{X}_{it}$ includes turnover ratio, log fund age, log total net assets, expense ratio, and cash holdings. Style ($\mathcal{S}_i$) and year fixed effects (δ_t) are included and denoted as Style FE and Time FE in the table. Standard errors are clustered at the fund level. The sample covers U.S. equity mutual funds from 2019–2024. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Source: Morningstar Direct, SEC N-CSR, and SEC N-PORT.

Table 3. Expense Ratio Determinants

	Full Sample			By MF Share Class			Twin	
	(1)	(2)	(3)	Retail (4)	Institutional (5)	Retirement (6)	(7)	(8)
SMA	-0.362***	-0.398***	-0.344***	-0.800***	-0.043***	-0.065***	-0.183***	-0.105***
	(0.009)	(0.009)	(0.012)	(0.020)	(0.011)	(0.020)	(0.020)	(0.039)
ETF	-0.769***	-0.792***	-0.833***	-1.264***	-0.483***	-0.695***	-0.713***	-0.132***
	(0.010)	(0.010)	(0.012)	(0.021)	(0.017)	(0.016)	(0.068)	(0.050)
MF Index Fund		-0.599***	-0.593***	-0.453***	-0.604***	-0.428***	-0.654***	-0.011
		(0.021)	(0.020)	(0.038)	(0.021)	(0.052)	(0.039)	(0.060)
Alpha			0.177***	0.287***	0.357***	0.410***	0.112***	-0.154***
			(0.008)	(0.010)	(0.011)	(0.010)	(0.012)	(0.025)
Log TNA			-0.010***	-0.005***	-0.006***	-0.004***	-0.010***	-0.065***
			(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)
Log Age			-0.001	-0.018***	-0.016***	-0.017***	-0.008**	0.166***
			(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.009)
Log Minimum Investment			-0.005***	-0.003	-0.001	-0.014***	-0.010***	-0.010***
			(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	232569	232569	219195	126916	108808	97837	98400	98393
R^2	0.122	0.186	0.206	0.391	0.127	0.133	0.186	0.385
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Style FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Twin FE	No	No	No	No	No	No	No	Yes

Note: The table reports regression coefficients from Equation 2:

$$\text{Expense Ratio}_{i,t} = \alpha + \beta_1 SMA_i + \beta_2 ETF_i + \gamma' X_{i,t} + \mathcal{S}_i + \delta_t + \varepsilon_{i,t},$$

where the dependent variable Expense Ratio_{it} is the expense ratio for investment vehicle i in year t. The key regressors include indicators for vehicle type: SMA_i equals one for separately managed accounts, and ETF_i equals one for exchange-traded funds, with mutual fund share classes as the omitted category. $Alpha_{it}$ measures the fund's annualized alpha from the Fama–French three-factor model. The control vector X_{it} includes log total net assets, log fund age (years since inception), log minimum investment, and an index fund indicator. Style fixed effects ($\mathcal{S}_i$) and year fixed effects (δ_t) are included as indicated in the column headers. Columns (1)-(3) include the full sample of mutual funds, ETFs, and SMAs. Columns (4)-(6) restrict the sample to specific mutual fund share class types. Columns (7)-(8) include only products that have a twin (same investment strategy) in another vehicle type, identified by StrategyId. Column (8) additionally absorbs strategy fixed effects. Standard errors are clustered by share class, which for ETFs and SMAs is synonymous with fund. All variables are winsorized at the 1st and 99th percentiles. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Source: Morningstar Direct

Table 4. Expense Ratios and Unrealized Gains

	Dependent variable: Expense Ratio (%)				
	(1)	(2)	(3)	(4)	(5)
Unrealized Gains Ratio	0.023	0.011	0.036**	0.040***	0.030**
	(0.015)	(0.014)	(0.015)	(0.015)	(0.015)
Alpha		0.017	0.060***	0.060***	0.067***
		(0.016)	(0.016)	(0.016)	(0.016)
Log TNA			-0.056***	-0.055***	-0.065***
			(0.004)	(0.004)	(0.004)
Log Age					0.108***
					(0.011)
Observations	6169	6034	6022	6022	6022
R^2	0.055	0.053	0.271	0.406	0.464
Style FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

Note: The table reports results from estimating Equation 3:

$$\text{Net Expense Ratio}_{it} = \beta_1 \,\text{UnrealizedGains}_{it} + \beta_2 \,\text{Performance}_{it} + \gamma' \mathbb{X}_{it} + \mathcal{S}_i + \delta_t + \varepsilon_{it},$$

where $\text{UnrealizedGains}_{it}$ is the unrealized gains ratio, measured as gross unrealized gains divided by total net assets. Performance_{it} is measured using either prior-year return or annualized Fama–French three-factor alpha. The control vector $\mathbb{X}_{it}$ includes log total net assets and log fund age. Style ($\mathcal{S}_i$) and year fixed effects (δ_t) are included and denoted as Style FE and Time FE in the table. Standard errors are clustered at the fund level. The sample covers U.S. equity mutual funds from 2019–2024. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Source: Morningstar Direct, SEC N-CSR, and SEC N-PORT.

TABLE 5. Fee Waivers and Unrealized Gains

	Dependent variable: Waiver Ratio (%)				
	(1)	(2)	(3)	(4)	(5)
Unrealized Gains Ratio	-0.051***	-0.050***	-0.032***	-0.032***	-0.020**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Alpha		0.027	0.050***	0.050***	0.045***
		(0.017)	(0.017)	(0.017)	(0.017)
Log TNA			-0.027***	-0.027***	-0.021***
			(0.003)	(0.003)	(0.003)
Log Age					-0.062***
					(0.008)
Observations	6055	5921	5921	5921	5921
R^2	0.026	0.026	0.077	0.080	0.130
Style FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes

Note: The table reports results from estimating Equation 4:

$$\text{Waiver Ratio}_{it} = \beta_1\, \text{UnrealizedGains}_{it} + \beta_2\, \text{Performance}_{it} + \gamma' \mathbb{X}_{it} + \mathcal{S}_i + \delta_t + \varepsilon_{it},$$

where Waiver Ratio_{it} is the fee waiver ratio for fund i in year t, defined as

$$\text{Waiver Ratio}_{it} = \frac{\text{Fee Waiver}_{it}}{\text{Gross Expense Ratio}_{it}}.$$

$\text{UnrealizedGains}_{it}$ is the unrealized gains ratio, measured as gross unrealized gains divided by total net assets. Performance_{it} is measured using either prior-year return or annualized alpha from the Fama–French three-factor model. The control vector $\mathbb{X}_{it}$ includes log total net assets and log fund age. Style ($\mathcal{S}_i$) and year fixed effects (δ_t) are included and denoted as Style FE and Time FE in the table. Standard errors are clustered at the fund level. The sample covers U.S. equity mutual funds from 2019–2024. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Source: Morningstar Direct, SEC N-CSR, and SEC N-PORT.

Table 6. Distribution Yields, Redemptions, and Outflows

	Dependent variable: Distribution Yield			
	(1)	(2)	(3)	(4)
Outflows × Unrealized Gains Ratio	0.48***	0.47***		
	(0.05)	(0.05)		
Outflows Ratio	0.00	-0.01		
	(0.01)	(0.01)		
Unrealized Gains Ratio	0.05***	0.04***		
	(0.01)	(0.01)		
High Outflows × Unrealized Gains Q3 (Top Tercile)			0.04***	0.04***
			(0.00)	(0.00)
High Outflows × Unrealized Gains Q2 (Middle Tercile)			0.02***	0.02***
			(0.00)	(0.00)
High Outflows (Top Tercile)			0.00	-0.00
			(0.00)	(0.00)
Unrealized Gains Q3 (Top Tercile)			0.02***	0.02***
			(0.00)	(0.00)
Unrealized Gains Q2 (Middle Tercile)			0.02***	0.02***
			(0.00)	(0.00)
Observations	5518	5507	5518	5507
R^2	0.156	0.328	0.226	0.302
Controls	No	Yes	No	Yes
Time FE	No	Yes	Yes	Yes
Style FE	No	Yes	No	Yes

Note: The Table reports the estimated coefficients from regression Equation 5 and 6:

$$\text{DistributionYield}_{it} = \beta_1 \,\text{Outflows}_{it} + \beta_2 \,\text{UnrealizedGains}_{it} + \beta_3 \left(\text{Outflows}_{it} \times \text{UnrealizedGains}_{it}\right) + \gamma' \mathbb{X}_{it} + \mathcal{S}_i + \delta_t + \varepsilon_{it},$$

and

$$\text{DistributionYield}_{it} = \beta_1 \,\text{HighOutflows}_{it} + \sum_{j=2}^{3} \beta_j \,\mathbb{I}(\text{UnrealizedGains}_{Qj,it}) + \sum_{j=2}^{3} \beta_{j+2} \left(\text{HighOutflows}_{it} \times \mathbb{I}(\text{UnrealizedGains}_{Qj,it})\right) + \gamma' \mathbb{X}_{it} + \mathcal{S}_i + \delta_t + \varepsilon_{it}$$

for mutual fund i in year t. The dependent variable $\text{DistributionYield}_{it}$ is defined as total annual distributions (short- and long-term capital gains) divided by the NAV before reinvestment. Outflows_{it} is the absolute value of annual outflows, computed as the sum of negative monthly flows scaled by lagged total net assets. $\text{UnrealizedGains}_{it}$ is the unrealized gains ratio, defined as gross unrealized gains divided by total net assets. HighOutflows_{it} is an indicator for outflows above the 66.7th percentile in a given year. $\mathbb{I}(\text{UnrealizedGains}_{Q2,it})$ and $\mathbb{I}(\text{UnrealizedGains}_{Q3,it})$ are indicators for funds in the second and third terciles of the unrealized gains distribution (the omitted baseline is Q1). The control vector $\mathbb{X}_{it}$ includes turnover ratio, log fund age, log total net assets, expense ratio, and cash holdings. Style ($\mathcal{S}_i$) and year fixed effects (δ_t) are included and denoted as Style FE and Time FE in the table. Standard errors are clustered at the fund level. The sample covers U.S. equity mutual funds from 2019–2024. All variables are winsorized at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Source: Morningstar Direct, SEC N-CSR, SEC N-PORT, and CRSP.

Table 7. Baseline Model Parameters

Parameter	**Symbol**	**Value**	**Source / Target Moment**
Capital gains tax rate	τ	0.20	U.S. long-term capital gains rate
Return on vehicle	R	0.12	Mean annual MF return
Liquidity shock probability	λ	0.05	Dávila and Goldstein (2023)
SMA Fee	f_{SMA}	0.0	Baseline
Investor heterogeneity	$\bar{\alpha}$	0.88	Moment matching (Table A.XI)

Note: The table reports the benchmark calibration assumes proportional liquidation of portfolio holdings, consistent with index fund practices (see Section C.IV). Uniform distribution for embedded gains chosen for tractability when an upper bound of 2.0 implies 50% fund-level unrealized gain ratio

Table 8. Equilibrium Outcomes

	α^*		MF AUM %		Dist Rate %		Opt. Fee %
	Safe	Run	Safe	Run	Safe	Run	
Baseline							
Uniform[0,0.88], proportional liquidation, $\lambda = 0.05$, $\tau = 0.20$							
	0.253	0.332	73.6	65.9	10.9	15.8	0.22
Unrealized Gains							
Proportional liquidation, $\lambda = 0.05$							
Uniform[0,0.88]	0.253	0.332	73.6	65.9	11.0	15.8	0.227
Uniform[0,2.0]	0.570	0.787	77.6	68.9	14.4	22.6	0.540
Uniform[0,3.0]	0.837	1.173	80.0	71.4	15.1	24.0	0.755
Uniform[0,4.0]	1.105	1.525	81.4	73.7	15.2	23.8	0.922
Uniform[0,5.0]	1.313	1.907	83.2	74.8	14.4	24.1	1.046

Note: The table reports equilibrium outcomes of the baseline model and alternative parameters. For each equilibrium, the table reports the investor threshold α^*, the share of assets under management remaining in the mutual fund (MF AUM %), and the implied distribution rate (Dist Rate %). "Low" refers to the stable equilibrium with limited redemptions, while "Run" refers to the fragile equilibrium with large redemptions. The final column reports the optimal mutual fund fee (Opt. Fee %). The table demonstrates multiple equilibria outcomes under varying parameter assumptions. Each panel varies one key parameter while holding others at baseline values. Additional baseline parameters: $R = 12\%$, $f_{SMA} = 0\%$.

TABLE 9. Liquidation Strategy Comparison Across Distribution Widths

	α^*		MF AUM %		Dist Rate %		Opt. Fee %
	Safe	Run	Safe	Run	Safe	Run	
Uniform[0,0.88]							
Proportional	0.253	0.332	73.6	65.9	11.0	15.8	0.227
Least-gains-first	0.266	0.369	72.4	62.2	5.1	11.2	0.386
Uniform[0,2.0]							
Proportional	0.570	0.787	77.6	68.9	14.4	22.6	0.540
Least-gains-first	0.549	0.734	78.4	71.2	7.6	14.9	0.679
Uniform[0,3.0]							
Proportional	0.837	1.173	80.0	71.4	15.1	24.0	0.755
Least-gains-first	0.759	1.004	81.7	75.9	8.5	16.0	0.846
Uniform[0,4.0]							
Proportional	1.105	1.525	81.4	73.7	15.2	23.8	0.922
Least-gains-first	0.960	1.223	83.7	79.4	9.3	15.9	0.970
Uniform[0,5.0]							
Proportional	1.313	1.907	83.2	74.8	14.4	24.1	1.046
Least Gains First	1.139	1.418	85.3	81.8	9.8	15.8	1.063

Note: The table reports equilibrium outcomes under different liquidation strategies across different levels of fund unrealized gains. For each equilibrium, the table reports the investor threshold α^*, the share of assets under management remaining in the mutual fund (MF AUM %), and the implied distribution rate (Dist Rate %). "Low" refers to the stable equilibrium with limited redemptions, while "Run" refers to the fragile equilibrium with large redemptions. The final column reports the optimal mutual fund fee (Opt. Fee %). The table demonstrates multiple equilibria outcomes under varying parameter assumptions. Each panel varies one key parameter while holding others at baseline values. Additional baseline parameters: $R = 12\%$, $f_{SMA} = 0\%$.

Appendix

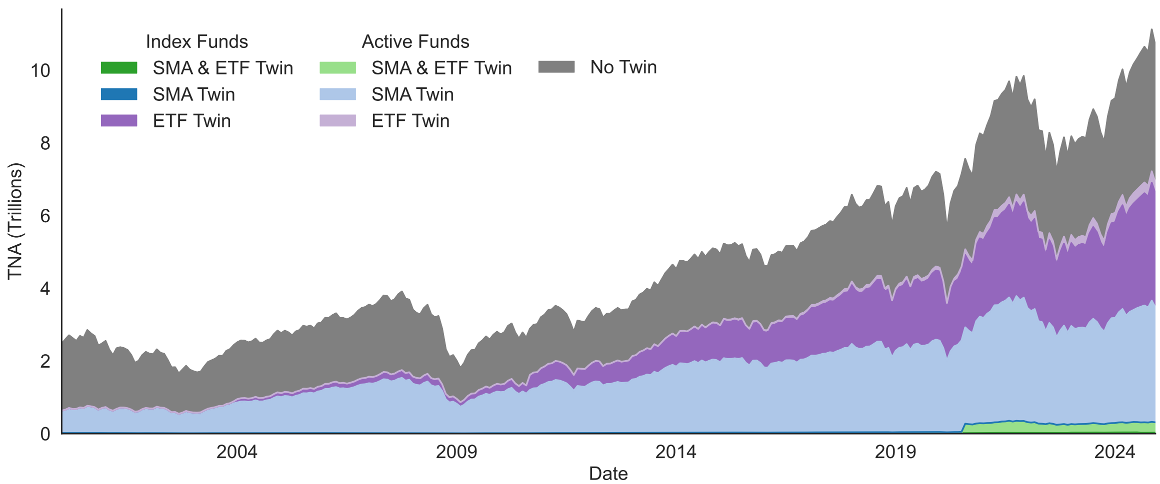
Jessica Goldenring

Contents

A Appendix Figures and Tables

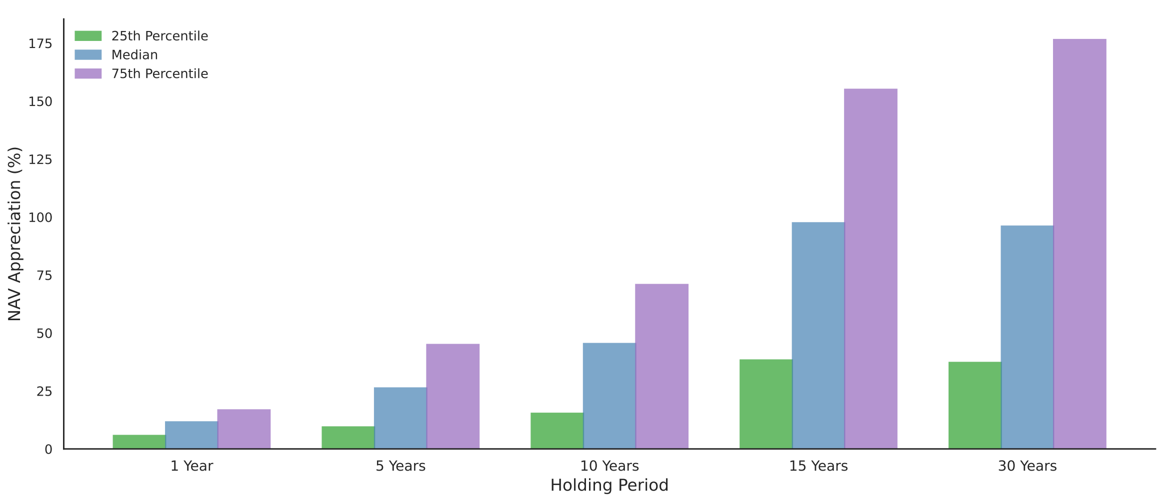
A.I Figures

FIGURE A.I. MFs with an ETF or SMA Twin by Activeness



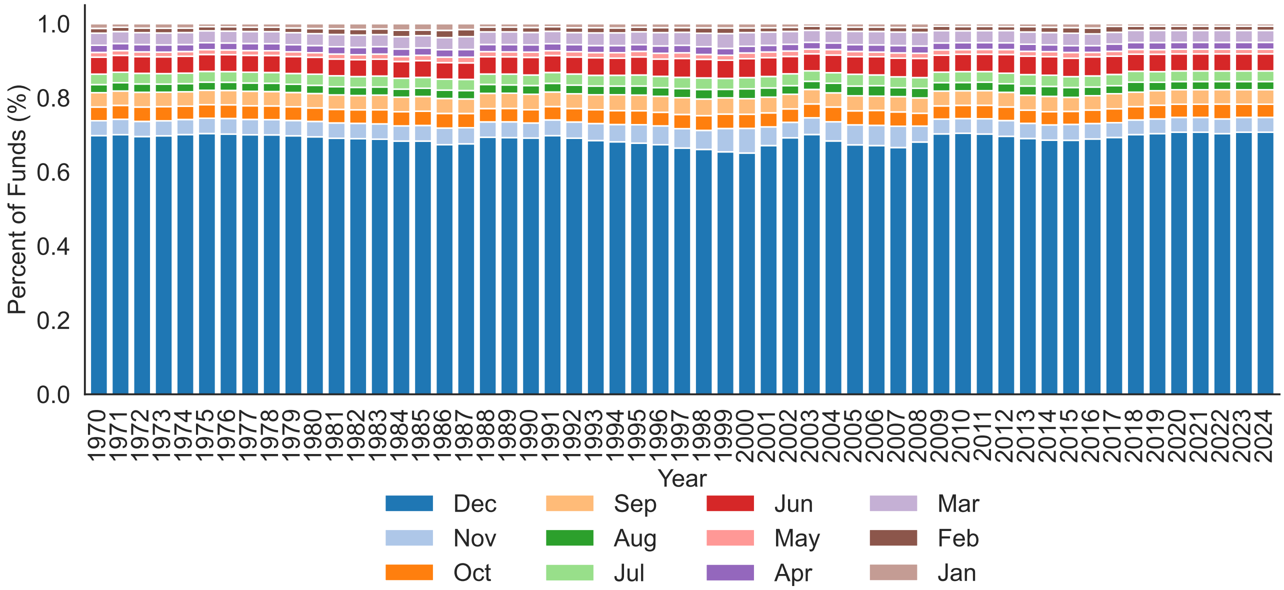
Note: The figure shows the assets of US Equity MFs by twins. Twins are defined by if the strategy is available as an SMAs or ETF. Strategy is defined by the StrategyId variable. Index Fund is defined by Morningstar Index Fund variables
Source: Morningstar Direct

FIGURE A.II. Appreciation of US Equity MFs After Different Holding Periods



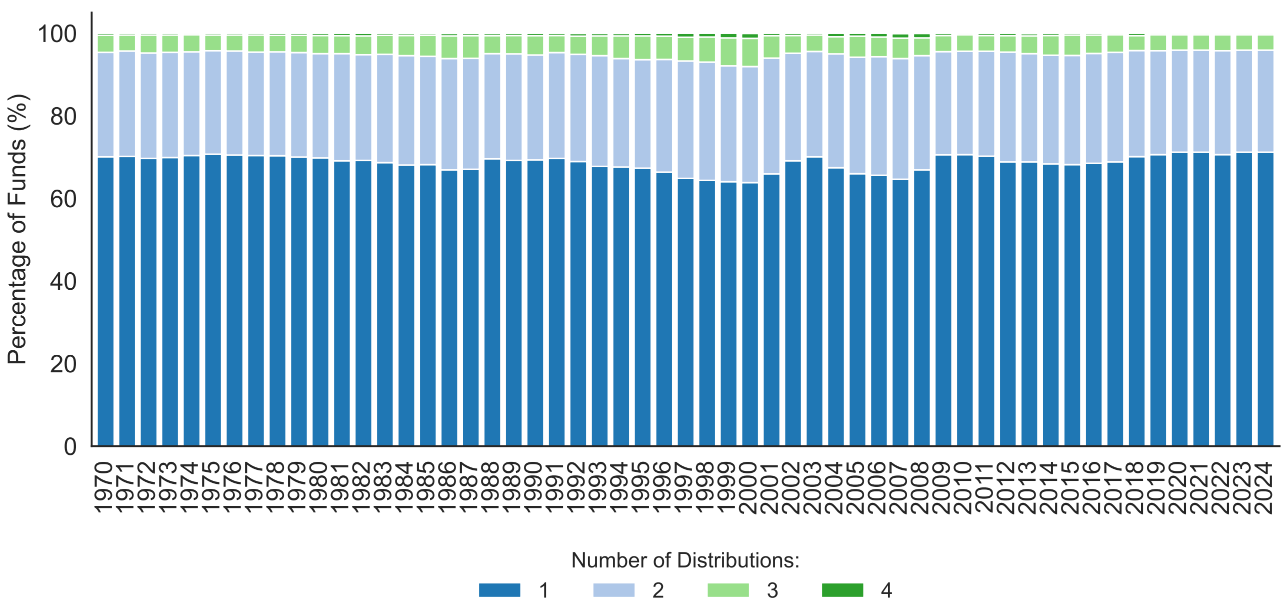
Note: The figure shows the the median 25th, Median, and 75th percentiles for the growth in NAV of MFs starting in December 2024 going back 1, 5, 10, 15, or 30 years for all US Equity MFs. For each holding period only MFs that have that long of history are included.
Source: Morningstar Direct

Figure A.III. Percent of Funds who Distribute



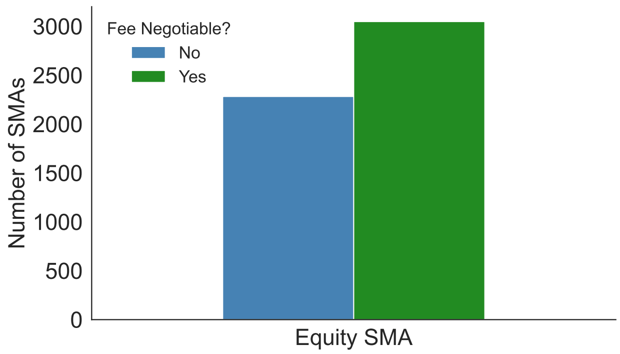
Note: The figure shows a stacked bar chart shows the percentage of equity MFs making capital gain distributions in each month from 1970 to 2024.
Source: CRSP

Figure A.IV. Number of Times a Year MFs Distribute

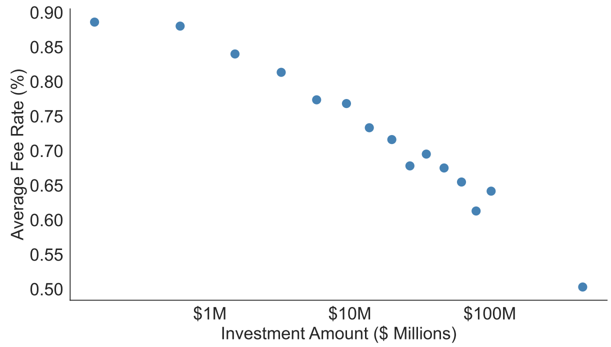


Note: This figure displays the percentage of equity funds by the number of capital gain distributions made per year from 1970 to 2024, showing the distribution patterns across different frequency categories.
Source: CRSP

Figure A.V. SMA Fees



(a) SMA Fee Negotiation



(b) SMA Fee Tier by Minimum Investment

Note: Panel (a) compares equity SMA composites with negotiable versus non-negotiable fee structures. Panel (b) shows the inverse relationship between minimum investment thresholds and fee rates across SMA composites.
Source: Morningstar Direct

A.II Tables

Table A.I. Unrealized Gains and Outflows: Alpha Specifications

	Dependent variable: Outflow Ratio						
	No Fund FE				Fund FE		
	Full Sample			Twin	Full Sample		Twin
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Unrealized Gain Ratio$_{t-1}$	-0.113***	-0.055***	-0.055***	-0.050**	-0.074***	-0.072***	-0.068**
	(0.013)	(0.016)	(0.017)	(0.020)	(0.026)	(0.026)	(0.033)
Alpha$_{t-1}$			-0.091***	-0.102***		-0.051	-0.052
			(0.031)	(0.037)		(0.032)	(0.037)
Turnover Ratio		0.069***	0.069***	0.067***	0.063***	0.062***	0.073**
		(0.013)	(0.013)	(0.016)	(0.021)	(0.021)	(0.030)
Log Age		-0.008	-0.008	-0.010	-0.072	-0.073	-0.043
		(0.006)	(0.006)	(0.008)	(0.059)	(0.059)	(0.082)
Log TNA		-0.004*	-0.004*	-0.001	0.038*	0.038*	0.053**
		(0.003)	(0.003)	(0.003)	(0.023)	(0.022)	(0.023)
Lag Cash		0.068	0.065	-0.345***	0.130		-0.017
		(0.177)	(0.177)	(0.124)	(0.272)		(0.214)
Expense Ratio		0.028*	0.026	0.036	0.208***	0.207***	0.176**
		(0.017)	(0.017)	(0.029)	(0.072)	(0.073)	(0.089)
Observations	4235	4235	4235	2558	4235	4235	2558
R^2	0.027	0.068	0.074	0.075	0.012	0.012	0.000
Fund FE	No	No	No	No	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Style FE	No	Yes	Yes	Yes	No	No	No

Note: The table reports results from estimating Equation 1:

$$\text{Outflows}_{it} = \beta_1\,\text{UnrealizedGains}_{i,t-1} + \beta_2\,\text{Alpha}_{i,t-1} + \mathbf{X}'_{it}\gamma + \mathcal{S}_i + \delta_t + \varepsilon_{it},$$

for mutual fund i in year t. The dependent variable Outflows_{it} is the absolute value of annual outflows, computed as the sum of negative monthly flows scaled by lagged total net assets. $\text{UnrealizedGains}_{i,t-1}$ is the lagged unrealized gains ratio, defined as gross unrealized gains divided by total net assets at the end of the prior year. $\text{Alpha}_{i,t-1}$ is the annualized Fama–French three-factor alpha from the prior 12 months. The control vector $\mathbf{X}_{it}$ includes turnover ratio, log fund age, log total net assets, expense ratio, and cash holdings. Column (1) reports a baseline specification with no fixed effects or controls. Columns (2)–(4) include style fixed effects ($\mathcal{S}_i$) and year fixed effects (δ_t). Columns (5)–(7) replace style fixed effects with fund fixed effects (μ_i) and year fixed effects (δ_t). Standard errors are clustered at the fund level. The sample covers U.S. equity mutual funds from 2019–2024. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Source: Morningstar Direct, SEC N-CSR, SEC N-PORT.

TABLE A.II. Unrealized Gains and Outflows: Return Specifications

	Dependent variable: Outflow Ratio						
	No Fund FE				Fund FE		
	Full Sample			Twin	Full Sample		Twin
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Unrealized Gain Ratio$_{t-1}$	-0.113***	-0.055***	-0.059***	-0.055***	-0.074***	-0.076***	-0.070**
	(0.013)	(0.016)	(0.017)	(0.020)	(0.026)	(0.026)	(0.033)
Return$_{t-1}$			-0.046***	-0.068***		-0.057***	-0.082***
			(0.016)	(0.019)		(0.021)	(0.020)
Turnover Ratio		0.069***	0.069***	0.067***	0.063***	0.060***	0.070**
		(0.013)	(0.013)	(0.016)	(0.021)	(0.021)	(0.030)
Log Age		-0.008	-0.009	-0.011	-0.072	-0.079	-0.053
		(0.006)	(0.006)	(0.008)	(0.059)	(0.059)	(0.082)
Log TNA		-0.004*	-0.004	-0.000	0.038*	0.045*	0.065***
		(0.003)	(0.003)	(0.003)	(0.023)	(0.024)	(0.024)
Lag Cash		0.068	0.071	-0.340***	0.130	0.113	-0.055
		(0.177)	(0.177)	(0.125)	(0.272)	(0.280)	(0.214)
Expense Ratio		0.028*	0.029*	0.040	0.208***	0.212***	0.182**
		(0.017)	(0.017)	(0.029)	(0.072)	(0.072)	(0.088)
Observations	4235	4235	4235	2558	4235	4235	2558
R^2	0.027	0.068	0.073	0.076	0.012	0.008	0.000
Fund FE	No	No	No	No	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Style FE	No	Yes	Yes	Yes	No	No	No

Note: The table reports results from estimating Equation 1:

$$\text{Outflows}_{it} = \beta_1 \, \text{UnrealizedGains}_{i,t-1} + \beta_2 \, \text{Return}_{i,t-1} + \mathbf{X}'_{it}\gamma + \mathcal{S}_i + \delta_t + \varepsilon_{it},$$

for mutual fund i in year t. The dependent variable Outflows_{it} is the absolute value of annual outflows, computed as the sum of negative monthly flows scaled by lagged total net assets. $\text{UnrealizedGains}_{i,t-1}$ is the lagged unrealized gains ratio, defined as gross unrealized gains divided by total net assets at the end of the prior year. $\text{Return}_{i,t-1}$ is the prior year's fund return. The control vector $\mathbf{X}_{it}$ includes turnover ratio, log fund age, log total net assets, expense ratio, and cash holdings. Column (1) reports a baseline specification with no fixed effects or controls. Columns (2)–(4) include style fixed effects ($\mathcal{S}_i$) and year fixed effects (δ_t). Columns (5)–(7) replace style fixed effects with fund fixed effects (μ_i) and year fixed effects (δ_t). Standard errors are clustered at the fund level. The sample covers U.S. equity mutual funds from 2019–2024. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Source: Morningstar Direct, SEC N-CSR, SEC N-PORT.

Table A.III. Net Expense Ratio and Unrealized Gains

	Dependent variable: Net Expense Ratio (%)									
	No Fund FE					Fund FE				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Unrealized Gains Ratio	0.023	0.011	0.036**	0.040***	0.030**	0.027*	0.016	0.027*	0.027*	0.023
	(0.015)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)
Alpha		0.017	0.060***	0.060***	0.067***		0.021	0.049***	0.049***	0.052***
		(0.016)	(0.016)	(0.016)	(0.016)		(0.016)	(0.016)	(0.016)	(0.016)
Log TNA			-0.056***	-0.055***	-0.065***			-0.036***	-0.036***	-0.040***
			(0.004)	(0.004)	(0.004)			(0.005)	(0.005)	(0.005)
Lag Cash			0.264***	0.246***	0.292***			0.089	0.089	0.104
			(0.080)	(0.082)	(0.081)			(0.076)	(0.076)	(0.073)
MF Index Fund				0.416***	0.385***					
				(0.036)	(0.035)					
Log Age					0.108***					0.089***
					(0.011)					(0.016)
Observations	6169	6034	6022	6022	6022	6169	6034	6022	6022	6022
R^2	0.055	0.053	0.271	0.406	0.464	0.000	0.000	0.251	0.251	0.338
Style FE	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table reports results from regressing net expense ratio on unrealized gains and controls. The dependent variable is the annual net expense ratio expressed in percent. UnrealizedGains$_{it}$ is the unrealized gains ratio, defined as gross unrealized gains divided by total net assets. Alpha$_{it}$ is the annualized Fama–French three-factor alpha from the prior 12 months. Controls are added progressively across columns and include log total net assets, an active fund indicator, log fund age, and cash holdings. Columns (1)–(5) include style fixed effects ($\mathcal{S}_i$) and year fixed effects (δ_t). Columns (6)–(10) replace style fixed effects with fund fixed effects (μ_i), which absorb all time-invariant fund characteristics including active fund status. Standard errors are clustered at the fund level. The sample covers U.S. equity mutual funds from 2019–2024. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Source: Morningstar Direct, SEC N-CSR, SEC N-PORT.

Table A.IV. Fee Waivers and Unrealized Gains: Alpha Specifications

	Dependent variable: Percent of Fee Waived (%)									
	No Fund FE					Fund FE				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Unrealized Gains Ratio	-0.051***	-0.050***	-0.032***	-0.034***	-0.022**	-0.033***	-0.030***	-0.020*	-0.021**	-0.018*
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.010)	(0.010)
Alpha		0.027	0.050***	0.050***	0.045***		0.027	0.053***	0.052***	0.049***
		(0.017)	(0.017)	(0.017)	(0.017)		(0.017)	(0.018)	(0.018)	(0.017)
Log TNA			-0.027***	-0.027***	-0.021***			-0.034***	-0.034***	-0.030***
			(0.003)	(0.003)	(0.003)			(0.007)	(0.007)	(0.007)
Lag Cash				-0.111	-0.133				-0.085	-0.097
				(0.083)	(0.083)				(0.093)	(0.091)
MF Index Fund					-0.008					
					(0.015)					
Log Age					-0.062***					-0.079***
					(0.008)					(0.022)
Observations	6055	5921	5921	5909	5909	6055	5921	5921	5909	5909
R^2	0.026	0.026	0.077	0.079	0.132	0.012	0.012	0.058	0.060	0.113
Fund FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Style FE	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table reports results from estimating Equation 4:

$$w_{it} = \beta_1 \, \text{UnrealizedGains}_{it} + \beta_2 \, \text{Alpha}_{it} + \gamma' \mathbf{X}_{it} + \mathcal{S}_i + \delta_t + \mu_i + \varepsilon_{it},$$

but with the addition of fund fixed effects. In the equation, w_{it} is the fee waiver percentage for fund i in year t, defined as

$$w_{it} = \frac{\text{Fee Waiver}_{it}}{\text{Gross Expense Ratio}_{it}}.$$

$\text{UnrealizedGains}_{it}$ is the unrealized gains ratio, measured as gross unrealized gains divided by total net assets. Alpha_{it} is the annualized Fama–French three-factor alpha from the prior 12 months. The control vector $\mathbf{X}_{it}$ includes log total net assets, cash holdings, an active fund indicator, and log fund age, added progressively across columns. Columns (1)–(5) include style fixed effects ($\mathcal{S}_i$) and year fixed effects (δ_t). Columns (6)–(10) replace style fixed effects with fund fixed effects (μ_i), which absorb all time-invariant fund characteristics including active fund status. Standard errors are clustered at the fund level. The sample covers U.S. equity mutual funds from 2019–2024. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. *Source:* Morningstar Direct, SEC N-CSR, SEC N-PORT.

Table A.V. Fee Waivers and Unrealized Gains: Return Specifications

	Dependent variable: Percent of Fee Waived (%)									
	No Fund FE					Fund FE				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Unrealized Gains Ratio	-0.051***	-0.051***	-0.033***	-0.035***	-0.023**	-0.033***	-0.032***	-0.022**	-0.023**	-0.019*
	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Return		0.019**	0.031***	0.031***	0.028***		0.018*	0.033***	0.032***	0.031***
		(0.009)	(0.009)	(0.009)	(0.009)		(0.009)	(0.010)	(0.010)	(0.010)
Log TNA			-0.027***	-0.027***	-0.021***			-0.033***	-0.033***	-0.030***
			(0.003)	(0.003)	(0.003)			(0.007)	(0.007)	(0.007)
Lag Cash				-0.108	-0.131				-0.075	-0.085
				(0.081)	(0.081)				(0.090)	(0.089)
MF Index Fund					-0.009					
					(0.015)					
Log Age					-0.062***					-0.078***
					(0.008)					(0.021)
Observations	6055	6055	6055	6043	6043	6055	6055	6055	6043	6043
R^2	0.026	0.026	0.079	0.081	0.133	0.012	0.013	0.060	0.062	0.115
Fund FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Style FE	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table reports results from estimating Equation 4:

$$w_{it} = \beta_1 \,\text{UnrealizedGains}_{it} + \beta_2 \,\text{Return}_{it} + \gamma' \mathbf{X}_{it} + \mathcal{S}_i + \delta_t + \mu_i + \varepsilon_{it},$$

but with the addition of fund fixed effects. In the equation, w_{it} is the fee waiver percentage for fund i in year t, defined as

$$w_{it} = \frac{\text{Fee Waiver}_{it}}{\text{Gross Expense Ratio}_{it}}.$$

$\text{UnrealizedGains}_{it}$ is the unrealized gains ratio, measured as gross unrealized gains divided by total net assets. Return_{it} is the prior year's fund return. The control vector $\mathbf{X}_{it}$ includes log total net assets, cash holdings, an active fund indicator, and log fund age, added progressively across columns. Columns (1)–(5) include style fixed effects ($\mathcal{S}_i$) and year fixed effects (δ_t). Columns (6)–(10) replace style fixed effects with fund fixed effects (μ_i), which absorb all time-invariant fund characteristics including active fund status. Standard errors are clustered at the fund level. The sample covers U.S. equity mutual funds from 2019–2024. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Source: Morningstar Direct, SEC N-CSR, SEC N-PORT.

Table A.VI. Distribution Yields, Redemptions, and Outflows

	Dependent variable: Distribution Yield							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outflows × Unrealized Gains Ratio	0.48***	0.38***	0.47***	0.36***				
	(0.05)	(0.05)	(0.05)	(0.05)				
Outflows Ratio	0.00	0.01	-0.01	0.00				
	(0.01)	(0.01)	(0.01)	(0.01)				
Unrealized Gains Ratio	0.05***	0.09***	0.04***	0.10***				
	(0.01)	(0.01)	(0.01)	(0.01)				
High Outflows × UG Ratio Q3 (Top Tercile)					0.04***	0.04***	0.04***	0.04***
					(0.00)	(0.00)	(0.00)	(0.00)
High Outflows × UG Ratio Q2 (Middle Tercile)					0.02***	0.02***	0.02***	0.02***
					(0.00)	(0.00)	(0.00)	(0.00)
High Outflows (Top Tercile)					0.00	-0.00	-0.00	-0.00
					(0.00)	(0.00)	(0.00)	(0.00)
Unrealized Gains Ratio Q3 (Top Tercile)					0.02***	0.03***	0.02***	0.03***
					(0.00)	(0.00)	(0.00)	(0.00)
Unrealized Gains Ratio Q2 (Middle Tercile)					0.02***	0.02***	0.02***	0.02***
					(0.00)	(0.00)	(0.00)	(0.00)
Alpha			0.07***	0.09***			0.07***	0.08***
			(0.01)	(0.01)			(0.01)	(0.01)
Log Age			0.01***	0.01			0.01***	0.02**
			(0.00)	(0.01)			(0.00)	(0.01)
Log TNA			-0.00***	-0.01*			-0.00***	-0.01***
			(0.00)	(0.00)			(0.00)	(0.00)
Turnover Ratio			0.06***	0.06***			0.06***	0.06***
			(0.00)	(0.01)			(0.00)	(0.01)
Lag Cash			-0.06*	-0.13**			-0.08***	-0.16***
			(0.03)	(0.06)			(0.03)	(0.06)
Observations	5518	5518	5507	5507	5518	5518	5507	5507
R^2	0.156	0.211	0.328	0.285	0.226	0.213	0.302	0.250
Entity FE	No	Yes	No	Yes	No	Yes	No	Yes
Time FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Style FE	No	No	Yes	No	No	No	Yes	No

Note: The Table presents the estimated coefficients from regression Equation 5 and 6 The dependent variable $\text{DistributionYield}_{it}$ is defined as total annual distributions (short and long-term capital gains) divided by average net assets before reinvestment. Outflows_{it} is the absolute value of annual outflows, computed as the sum of negative monthly flows scaled by lagged net assets. $\text{UnrealizedGains}_{it}$ is the unrealized gains ratio, defined as gross unrealized gains divided by total net assets. HighOutflows_{it} is an indicator for outflows above the 75th percentile in a given year. $\mathbb{I}(\text{UnrealizedGains}_{Q2,it})$ and $\mathbb{I}(\text{UnrealizedGains}_{Q3,it})$ are indicators for funds in the second and third terciles of the unrealized gains distribution (the omitted baseline is Q1). The controls include turnover ratio, log fund age, log total net assets, expense ratio, and cash holdings. All regressions include style fixed effects ($\mathcal{S}_i$) and year fixed effects (δ_t). Standard errors are clustered at the fund level. All variables are winsorized at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Source: Morningstar Direct, SEC N-CSR, SEC N-PORT.

Table A.VII. Top MF, SMA, and ETF Firms

MF Firm Name	TNA (Billions)	% of TNA	SMA Firm Name	TNA (Billions)	% of TNA	ETF Firm Name	TNA (Billions)	% of TNA
Vanguard	3505.29	32.89	Capital Group	1261.12	18.24	Vanguard	1889.73	32.97
Fidelity	2096.67	19.67	T. Rowe Price	745.04	10.78	iShares	1638.89	28.60
Capital Group	971.81	9.12	Northern Trust	403.27	5.83	State Street	883.09	15.41
T. Rowe Price	413.41	3.88	Wellington Management	267.07	3.86	Invesco	529.65	9.24
JPMorgan	299.60	2.81	BlackRock	252.86	3.66	Charles Schwab	259.33	4.52
Franklin Templeton	213.57	2	MFS	232.90	3.37	Dimensional	100.62	1.76
Dimensional	189.90	1.78	JPMorgan	217.35	3.14	First Trust	58.69	1.02
Charles Schwab	182.58	1.71	Franklin Templeton	182.85	2.65	Pacer	40.26	0.70
MFS	179.79	1.69	Columbia Threadneedle	134.22	1.94	Fidelity	34.06	0.59
Nuveen	135.97	1.28	AllianceBernstein	126.66	1.83	American Century	32.71	0.57
Jackson	126.69	1.19	Jennison	121.73	1.76	WisdomTree	32.49	0.57
Invesco	116.54	1.09	American Century	118.94	1.72	Capital Group	25.88	0.45
Dodge and Cox	111.65	1.05	Janus Henderson	102.02	1.48	JPMorgan	24.49	0.43
Columbia Threadneedle	111.35	1.04	UBS	97.71	1.41	Goldman Sachs	20.51	0.36
BlackRock	106.51	1	Invesco	89.16	1.29	VanEck	15.76	0.28
December 2024 % of TNA		82.21			62.97			97.47

Note: Table reports the top MF and SMA firm names and their assets for December 2024 year-end. TNA is the total net assets for that firm as of December 2024. Bottom row includes share of total vehicle assets accounted for by the firms shown in the table.
Source: Morningstar Direct

Table A.VIII. Equity SMA and MF Holdings

	Top 10	Top 20	Overlap % SMA	Overlap % MF	Weight % SMA	Weight % MF
Mean	8.53	17.63	92.91	90.13	93.66	92.36
1%	0.00	0.00	7.37	7.22	8.45	8.32
5%	4.00	8.00	63.40	43.59	74.05	60.40
10%	6.00	14.00	87.50	75.59	89.56	84.62
25%	8.00	18.00	95.24	92.55	95.23	94.19
50%	9.00	19.00	97.67	96.72	97.78	97.37
75%	10.00	20.00	98.91	98.36	99.50	99.11
90%	10.00	20.00	100.00	100.00	100.00	100.00
95%	10.00	20.00	100.00	100.00	100.00	100.00
99%	10.00	20.00	100.00	100.00	100.01	100.04

Note: Table reports the overlap of the top 10 and 20 holdings for MFs with an SMA twin. Sample includes twin holdings from 2011 to 2022 at the quarterly level.
Source: Morningstar Direct

B Data Appendix

B.I Morningstar Data Appendix

Morningstar Direct provides detailed information on assets, returns, fees, and vehicle characteristics. Morningstar identifies investments using two key identifiers: the share-class identifier

(SecId) and the fund identifier (FundId). For ETFs and SMAs, the identifier uniquely represents the vehicle. For MFs, multiple share classes (A, B, C, institutional, etc.) map to a single FundId.

Morningstar's SMA data are obtained through the same platform as MF and ETF data, with two key differences. First, SMAs are reported at the composite level—representative strategy-level portfolios rather than individual client accounts. Second, Morningstar relies on direct reporting from asset managers. Unlike MFs and ETFs, which are SEC-registered pooled vehicles with standardized public filings, SMAs report voluntarily and with varying frequency.

Standard Variables Total net assets (TNA), gross returns, and net returns are obtained from standard Morningstar reporting fields (Net Assets Share Class, Monthly Gross Return, Monthly Return). MFs and ETFs report all variables at monthly frequency. SMAs report returns monthly but assets at varying frequency: as of 2024, approximately 80% report assets monthly, with the remainder reporting quarterly. For analyses requiring asset data, I use the available reporting frequency for each SMA (monthly when available, quarterly otherwise).

Flows Flows are taken from Morningstar's Estimated Share Class (Net Flows) variable at monthly frequency. For robustness, I verify these using the standard identity $F_{it} = TNA_{it} - TNA_{i,t-1}(1 + r_{it})$. Flows are scaled by lagged assets: $f_{it} = F_{it}/TNA_{i,t-1}$.

For annual analysis, outflows are measured as the sum of months with negative net flows:

$$\text{Outflows}_{i,t} = \frac{\sum_{m \in \mathcal{M}_{\text{year } t}} |F_{i,m}| \, \mathbf{1}\{F_{i,m} < 0\}}{TNA_{i,t-1}},$$

where $\mathcal{M}$ is the set of months in year t.

Index Fund Morningstar provides an Index Fund variable, which identifies funds that explicitly track an index with the objective of replicating its returns. This variable is used to distinguish between active and passive strategies throughout the analysis.

StrategyId Morningstar's StrategyId variable links investments that follow the same underlying investment process. According to Morningstar:

> The Morningstar identifier links investments that follow the same investment process. Often investment management companies subadvise more than one mutual fund and offer equivalent investment pools in separate accounts, collective investment

trusts, or other vehicles. Following industry convention, Morningstar groups these substantively identical pools into a single strategy.

This variable is central for identifying twin funds across MFs, ETFs, and SMAs that implement substantively identical strategies.

Share Class Type Morningstar provides a Share Class Type variable that categorizes share classes within open-end mutual funds. In this analysis, I define retail share classes as A, B, C, D, M, and T shares. Retirement is categorized by the retirement share class type, and likewise the institutional category is formed by the Institutional share class type. Other share class types in the sample include Advisor, No Load, or a residual "Other" category. No load shares include S and D share classes. To address cases where the Morningstar variable is missing, I additionally classify share classes based on the fund name, using the suffix convention commonly employed in the industry. For example, "Example Mutual Fund A" denotes an A share class.

Fees I use both the gross and net expense ratio for fees for the ETF and MFs. The gross expense ratio is calculated as total gross expenses (net expenses plus any waived fees). It reflects what a fund's costs would be in the absence of fee waivers. The net expense ratio is the percentage of fund assets used to pay for operating and management expenses, including 12b-1 fees, administrative fees, and other asset-based costs, but excluding brokerage costs.

SMA fees are not reported directly. A key characteristic of SMAs is that the fees are often negotiable between the manager and investor and vary considerably by the minimum investment of the investor (Figure A.V). Following the existing literature, I therefore calculate the fee of the SMA as the difference between gross and net returns (Elton, Gruber, and Blake, 2013).

MF fees are contractually set and generally cannot be increased without shareholder approval. Funds frequently use fee waivers to temporarily reduce costs without altering contracts. The relationship is:

$$\text{Net Expense Ratio} = \text{Gross Expense Ratio} - \text{Fee Waivers}$$

Fee waivers mechanically increase investor returns: each basis point waived increases net returns by approximately one basis point. This provides managers flexibility to adjust pricing without triggering shareholder votes. I construct the fee waiver percentage as:

$$\text{Waiver Ratio}_{it} = \frac{\text{Fee Waiver}_{it}}{\text{Gross Expense Ratio}_{it}},$$

which measures waivers as a share of gross expenses, enabling comparison across funds with different fee levels. In the past decade, about 60% of the equity MFs have issued fee waivers.

B.II CRSP Distributions

Capital gains distributions data come from CRSP mutual fund distributions (1970-2024). CRSP reports both short-term (dis_type: CB, CH, CS, CM) and long-term capital gains (CL, CP, CJ, C), which I combine into total dollar distributions. CRSP records the payment date and reports months without distributions as missing rather than zero. To distinguish actual zero distributions from missing data, I infer each fund's typical distribution pattern from the prior five years and construct a balanced panel with explicit zeros in expected distribution months.

Distribution yields are computed as:

$$\text{Distribution Yield}_{it} = \frac{\text{Distribution}_{it}}{NAV_{it}^{\text{reinvest}} + \text{Distribution}_{it}},$$

I match CRSP distributions to Morningstar and SEC identifiers using Pastor, Stambaugh, and Taylor (2015)'s methodology linking CRSP share codes to Morningstar FundIds, supplemented by CRSP's direct EDGAR mapping (CRSP-CIK form) to connect to SEC N-PORT and N-CSR filings.

B.III Unrealized Gains

This section describes the construction of fund-level unrealized gains from SEC filings. Funds report annual gross unrealized gains in N-CSR filings and monthly net changes in N-PORT filings. I combine these to construct a consistent monthly time series of gross unrealized gains from 2019-2024.

N-PORT Data

Form N-PORT is the Monthly Portfolio Investments Report. It is filed by registered management investment companies (such as ETFs, MFs) to the SEC. Funds are required to file quarterly no later than 60 days after the quarter-end. The SEC adopted Form N-PORT in October 2016 as part of the Investment Company Reporting Modernization Rules but compliance was phased in starting in 2019. The form covers entire fund-level information and therefore not at the share class level. This remains fine for this analysis as holdings are at the fund level.

This paper uses the following line items:

- **Net Assets** (Item B.1.c; Quarterly)

- **Net Realized Gains and Change in Unrealized Appreciation** (Item B.5.d; Monthly)

Key to the analysis, the N-PORT data reports the net change in unrealized appreciation or depreciation. Thus this measure can be considered as the flow of unrealized gains from one month to another. This data is provided by the SEC as a structured data set. The data does not require any manual scraping and is provided by the SEC in quarterly data downloads.[1] The interpretation of the variable is from one month to another unrealized gains increased or decreased by some amount of US dollars. In order to get a gross measure of the total accumulated unrealized gains I combine this data with the N-CSR data.

N-CSR Data

The Certified Shareholder Report of Registered Management Investment Companies (N-CSR) contains the fund's annual or semiannual shareholder report. In this reporting form, firms report their tax disclosures in the Notes to Financial Statements. All registered management investment companies under the Investment Company Act of 1940 are required to file this report. However, unlike the N-PORT data, these data are not structured. Funds will use different terminology and report the values in different formats including tables or footnotes. Key to this analysis, funds report a gross measure of the unrealized appreciation and depreciation.

Sample Construction

I manually compile the gross measure of unrealized gains for funds. To do this I begin by pulling N-CEN identifiers from Morningstar. The N-CEN is the annual report for registered investment vehicle companies but important for my analysis Morningstar has these IDs allowing me to link my sample to the N-PORT data. Specifically, I use the Registrant Central Index Key (CIK) and the Fund Legal Entity Identifier (LEI). The CIK is the unique SEC identifier for the registrant (fund complex or trust) that files the report. The LEI is at the fund level. With these variables that map the funds ids, I combine this with the same variables in the N-PORT data to have a mapping between the two. This mapping enables me to link Morningstar funds to both N-PORT (for monthly flows) and N-CSR (for annual levels), as both SEC filings use the same CIK and LEI identifiers. I map 2,178 funds from Morningstar to Edgar. This sample represents 70% of assets in 2020 but 98% of assets in 2024. The discrepancy is in the start of the N-NPORT data beginning in 2019 shown in Figure A.VI.

With this mapping, I next use the SEC's EDGAR JSON submission API to obtain all N-CSR filings for each fund's central index key (CIK). For each filing, the API provides a direct URL

[1]Data can be found here: https://www.sec.gov/data-research/sec-markets-data/form-n-port-data-sets

to the filing page, which I then download and parse with BeautifulSoup to identify the fund's series identifier and extract header metadata such as the filing date and reporting period. For robustness, I cross-check coverage using the WRDS EDGAR database to ensure completeness for each CIK. The result is a list of text files that contain the unrealized gains for funds in the Morningstar sample.

I extract fund-level unrealized gain disclosures from N-CSR text files using a two-stage pipeline. First, I scan the plain text of each filing with a regular expression that captures windows around the phrase "gross unrealized appreciation", which yields candidate disclosure chunks. These chunks are then passed to a large language model (gpt-4o, with gpt-4-turbo as a fallback) using a structured extraction prompt that requires specific variables, enforces strict column names, and outputs only CSV-formatted results. The main prompt used is the following:

> You are a financial document extraction assistant. Given text from a mutual fund regulatory filing, extract the following information for each fund mentioned. Required fields are the Gross Unrealized Appreciation, Gross Unrealized Depreciation, Net Unrealized Appreciation/Depreciation (all numeric with no commas), the Reporting Date or Period (wrapped in double quotes), and the Fund Name (look for capitalized multi-word names before "Fund(s)"). In addition, extract optional fields if available: Undistributed Long-Term Capital Gains, Undistributed Ordinary Income, Capital Loss Carryforwards, Net Unrealized Gains, and Tax Cost (all numeric), otherwise return "NA." If the disclosure specifies units, such as "Amount ($000)," add a Unit column. The final output should be in CSV format with the following columns in order: Fund Name, Gross Unrealized Appreciation, Gross Unrealized Depreciation, Net Unrealized Appreciation, Undistributed LT Cap Gains, Undistributed Ordinary Income, Capital Loss Carryforwards, Net Unrealized Gains, Tax Cost, Reporting Period, Unit.

The pipeline enforces SEC-compliant headers and polite rate limits, maintains JSON checkpoints of previously processed chunks, and reuses valid records already present in legacy outputs to avoid reprocessing. After extraction, I apply several validation checks before adding results to the dataset: a schema check to ensure all expected fields are present (filling missing fields with "NA" where necessary), a screen to rule out invalid fund names while requiring a minimum proportion of plausible names within each filing, and a numeric consistency test verifying that net unrealized appreciation equals gross appreciation minus gross depreciation once the depreciation is signed correctly. Records that fail these checks are excluded, while validated results added to the final dataset.

Matching the N-CSR and N-PORT Data

N-CSR does not list an unique identifier when report unrealized gains and instead usually lists the name of the fund, I link these observations back to the structured data by matching fund names across filings. I use the N-PORT filings, which contain both the series name and the official series ID, as the anchor. Within each filing, I first pair exact matches and then apply conservative fuzzy matching to align fund names between the N-CSR and N-PORT. This procedure allows me to assign the correct series identifier to each fund in the N-CSR data and integrate the textual notes with structured regulatory data.

Combining N-PORT and N-CSR data enables reconciliation and extension of unrealized gains measurement over time. N-CSR filings report annual levels of gross unrealized gains, while N-PORT filings report monthly flows of net unrealized appreciation. By anchoring each fund to its most recent N-CSR level and combining intervening N-PORT flows, I construct a consistent monthly series of gross unrealized gains. This approach ensures internal consistency: the gross value from N-CSR plus subsequent net flows equals the next reported gross value. This method provides a higher-frequency panel for empirical analysis.

Constructing Monthly Time Series

N-CSR reports annual levels of gross unrealized gains ($UG_{i,t}^{\text{NCSR}}$) while N-PORT reports monthly changes ($\Delta UG_{i,t}$). To construct a consistent monthly time series, I anchor to N-CSR levels and use N-PORT flows to fill in the months between annual filings.

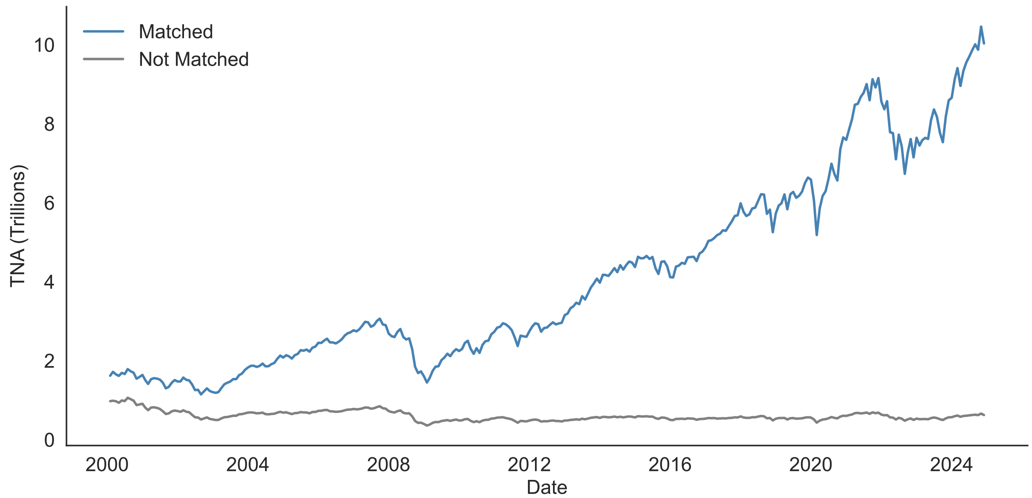
At each N-CSR filing date, I use the reported level directly. Between filings, I cumulate N-PORT monthly flows from the most recent N-CSR anchor:

$$UG_{i,t} = \begin{cases} UG_{i,t}^{\text{NCSR}} & \text{if } t \in \mathcal{N}_i, \\ UG_{i,n}^{\text{NCSR}} + \sum_{u=n+1}^{t} \Delta UG_{i,u}, & \text{where } n = \max\{s \in \mathcal{N}_i : s < t\}, \end{cases}$$

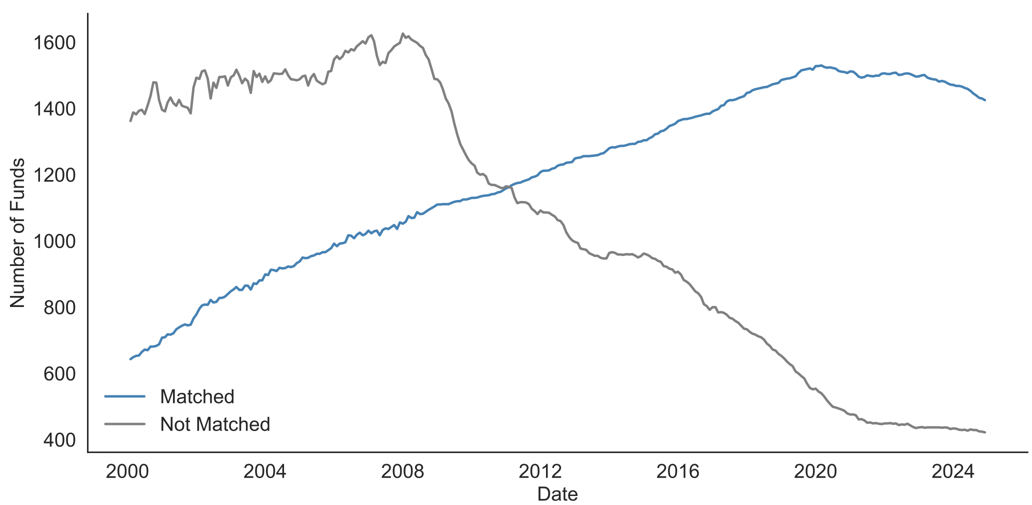
where $\mathcal{N}_i$ is the set of months with N-CSR filings for fund i. This approach ensures internal consistency: cumulative N-PORT flows between anchors exactly reconcile reported N-CSR levels.

FIGURE A.VI. Number of Funds from Morningstar Sample with Unrealized Gains Data

(a) Total Net Assets of Matched Sample



(b) Number of Funds of Matched Sample



Note: The figure shows the number of equity mutual funds and total net assets (TNA) over time, comparing funds that have matching records across Morningstar, N-CSR, and N-PORT databases versus funds that do not have complete matches across all three data sources.
Source: Morningstar Direct, SEC N-CSR, and SEC N-PORT

Validation

To validate this approach, I construct two alternative strategies. The first alternative (earliest anchor forward) projects forward from the first N-CSR filing, accumulating all subsequent N-PORT flows from that initial anchor. The second alternative strategy (latest anchor backwards) projects backward from the last N-CSR filing, working backward through the sample period. For each fund-strategy pair, I regress reported N-CSR values on constructed values:

For each fund-strategy pair, I regress reported N-CSR values on constructed values:

$$UG_{it}^{\text{NCSR}} = \alpha_i + \beta_i \times UG_{it}^{\text{strategy}} + \epsilon_{it}. \tag{A.1}$$

Perfect reconstruction yields $\beta_i = 1$ and $R^2 = 1$. Table A.IX shows that both alternative strategies achieve median $R^2 = 1$ and median slope $= 1$ across 1,253 funds, with 90% of funds achieving $R^2 \geq 1$. The equivalence of forward and backward strategies confirms internal consistency: N-PORT flows accurately reconcile N-CSR levels regardless of anchor choice.

Table A.IX. Validation of N-PORT and N-CSR

	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
R^2 Distributions											
Earliest Anchor Forward	0.78	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Latest Anchor Backwards	0.78	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Slope Distributions											
Earliest Anchor Forward	0.77	0.92	0.98	0.99	1.00	1.00	1.00	1.00	1.01	1.03	1.09
Latest Anchor Backwards	0.77	0.92	0.98	0.99	1.00	1.00	1.00	1.00	1.01	1.03	1.09

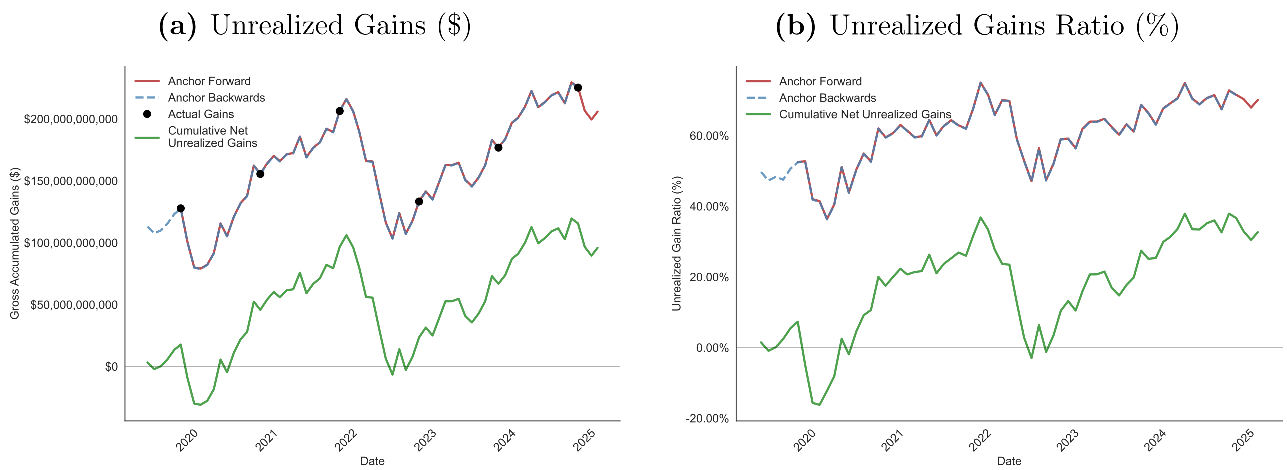
Note: Table reports percentiles of R^2 and regression slopes from strategy-specific regressions of constructed unrealized gains on reported values. Results are based on 1,253 unique funds.

A critical distinction for this analysis is the difference between using N-PORT data alone versus the constructed gross unrealized gains measure. For measuring investor embeddedness, what matters is the total accumulation of unrealized gains since fund inception, not merely the gains accrued within a single year. Figure A.VII illustrates this distinction using the Vanguard Institutional Index Fund as an example.

The figure demonstrates the data construction and validation process. The black dots represent gross unrealized gains reported in N-CSR filings, while the green line shows the cumulative sum of monthly N-PORT flows starting from zero. The red and dotted blue lines show the combined N-CSR and N-PORT measures using the earliest forward anchor and latest backward anchor strategies, respectively.

Two key insights emerge from this comparison. First, the alignment between N-CSR anchor points and the constructed series validates that N-PORT monthly flows can be accurately combined with N-CSR gross levels to create a consistent time series. Second, the figure reveals the substantial difference in magnitude between approaches: using only cumulative N-PORT data yields an unrealized gain ratio of approximately 20%, while the combined approach that incorporates the full historical accumulation shows Vanguard's true unrealized gain ratio of around 60%. This 40 percentage point difference underscores why the gross historical measure is essential for accurately capturing investor tax lock-in effects.

FIGURE A.VII. Vanguard Institutional Index Fund Unrealized Gains



Note: The figure shows the unrealized gains for Vanguard Index Fund. The left panel shows the connection between the SEC N-CSR data (black dots) and the SEC N-PORT (green line). The red and dotted blue line show the combined N-CSR and N-PORT data using forward and backwards accumulation. The right panel shows the unrealized gains as a ratio of a percent of net assets reported by N-PORT
Source: Morningstar, SEC N-PORT, and SEC N-CSR

Finally, I match these unrealized gains data to CRSP distributions using the CRSP-CIK mapping form. I manually verify and add seven additional FundId-CRSP mappings; 65 funds remain unmatched (primarily funds without tickers).

C Model Details

C.I Uniform Distribution

This section provides derivations for the uniform distribution case used as the baseline throughout the paper. Embedded gains α_i follow a uniform distribution on $[\underline{\alpha}, \bar{\alpha}]$ with probability density function:

$$p(\alpha) = \frac{1}{\bar{\alpha} - \underline{\alpha}} \text{ for } \alpha \in [\underline{\alpha}, \bar{\alpha}]$$

Distribution Rate Components

To compute the distribution rate $\delta(\alpha^*, \lambda)$, I first derive the underlying integrals for $L(\alpha^*, \lambda)$, who leaves the MF, and $R(\alpha^*, \lambda)$ who remains in the MF.

The value of assets held by investors with embedded gains below the threshold α^* is:

$$\begin{aligned} \int_{\underline{\alpha}}^{\alpha^*} (1+\alpha) \cdot p(\alpha)\, d\alpha &= \frac{1}{\bar{\alpha} - \underline{\alpha}} \int_{\underline{\alpha}}^{\alpha^*} (1+\alpha)\, d\alpha \\ &= \frac{1}{\bar{\alpha} - \underline{\alpha}} \left[(\alpha^* - \underline{\alpha}) + \frac{(\alpha^*)^2 - \underline{\alpha}^2}{2} \right]. \end{aligned}$$

The value of assets held by investors with embedded gains above the threshold is:

$$\int_{\alpha^*}^{\bar{\alpha}} (1+\alpha) \cdot p(\alpha)\, d\alpha = \frac{1}{\bar{\alpha} - \underline{\alpha}} \left[(\bar{\alpha} - \alpha^*) + \frac{\bar{\alpha}^2 - (\alpha^*)^2}{2} \right].$$

The total value of all fund assets is:

$$\int_{\underline{\alpha}}^{\bar{\alpha}} (1+\alpha) \cdot p(\alpha)\, d\alpha = \frac{1}{\bar{\alpha} - \underline{\alpha}} \left[(\bar{\alpha} - \underline{\alpha}) + \frac{\bar{\alpha}^2 - \underline{\alpha}^2}{2} \right].$$

The complete expressions for liquidations and remaining assets are:

$$\begin{aligned} L(\alpha^*, \lambda) = \lambda \int_{\underline{\alpha}}^{\bar{\alpha}} (1+\alpha) p(\alpha) d\alpha + (1-\lambda) \int_{\underline{\alpha}}^{\alpha^*} (1+\alpha) p(\alpha) d\alpha \\ = \frac{\lambda}{\bar{\alpha} - \underline{\alpha}} \left[(\bar{\alpha} - \underline{\alpha}) + \frac{\bar{\alpha}^2 - \underline{\alpha}^2}{2} \right] \\ + \frac{(1-\lambda)}{\bar{\alpha} - \underline{\alpha}} \left[(\alpha^* - \underline{\alpha}) + \frac{(\alpha^*)^2 - \underline{\alpha}^2}{2} \right]. \end{aligned}$$

$$
\begin{aligned}
R(\alpha^*, \lambda) &= (1-\lambda)\int_{\alpha^*}^{\bar{\alpha}} (1+\alpha)p(\alpha)d\alpha \\
&= \frac{(1-\lambda)}{\bar{\alpha}-\underline{\alpha}}\left[(\bar{\alpha}-\alpha^*) + \frac{\bar{\alpha}^2-(\alpha^*)^2}{2}\right].
\end{aligned}
$$

Liquidation Strategy

The embedded gains ratio under **proportional liquidation** equals the fund's weighted average:

$$
\begin{aligned}
\omega_{Prop} &= \frac{\int_{\underline{\alpha}}^{\bar{\alpha}} \alpha \cdot p(\alpha)\, d\alpha}{\int_{\underline{\alpha}}^{\bar{\alpha}} (1+\alpha) \cdot p(\alpha)\, d\alpha} \\
&= \frac{\frac{1}{\bar{\alpha}-\underline{\alpha}} \cdot \frac{\bar{\alpha}^2-\underline{\alpha}^2}{2}}{\frac{1}{\bar{\alpha}-\underline{\alpha}} \cdot \left[(\bar{\alpha}-\underline{\alpha}) + \frac{\bar{\alpha}^2-\underline{\alpha}^2}{2}\right]} \\
&= \frac{\bar{\alpha}+\underline{\alpha}}{2+\bar{\alpha}+\underline{\alpha}}.
\end{aligned}
$$

Under **least-gains first** (LGF) liquidation, only assets with $\alpha_i < \alpha^*$ are sold:

$$
\begin{aligned}
\omega_{LGF}(\alpha^*) &= \frac{\int_{\underline{\alpha}}^{\alpha^*} \alpha \cdot p(\alpha)\, d\alpha}{\int_{\underline{\alpha}}^{\alpha^*} (1+\alpha) \cdot p(\alpha)\, d\alpha} \\
&= \frac{\frac{(\alpha^*)^2-\underline{\alpha}^2}{2(\bar{\alpha}-\underline{\alpha})}}{\frac{1}{\bar{\alpha}-\underline{\alpha}}\left[(\alpha^*-\underline{\alpha}) + \frac{(\alpha^*)^2-\underline{\alpha}^2}{2}\right]} \\
&= \frac{\alpha^*+\underline{\alpha}}{2+\alpha^*+\underline{\alpha}}.
\end{aligned}
$$

Distribution Rates

The distribution rate under proportional liquidation is:

$$
\begin{aligned}
\delta^{\text{Prop}}(\alpha^*, \lambda) &= \omega_{Prop} \cdot \frac{L(\alpha^*, \lambda)}{R(\alpha^*, \lambda)} \\
&= \frac{\bar{\alpha}+\underline{\alpha}}{2+\bar{\alpha}+\underline{\alpha}} \cdot \frac{\lambda\left[(\bar{\alpha}-\underline{\alpha}) + \frac{\bar{\alpha}^2-\underline{\alpha}^2}{2}\right] + (1-\lambda)\left[(\alpha^*-\underline{\alpha}) + \frac{(\alpha^*)^2-\underline{\alpha}^2}{2}\right]}{(1-\lambda)\left[(\bar{\alpha}-\alpha^*) + \frac{\bar{\alpha}^2-(\alpha^*)^2}{2}\right]}.
\end{aligned}
$$

Under least-gains-first liquidation:

$$\delta^{\mathrm{LGF}}(\alpha^*,\lambda) = \omega_{LGF}(\alpha^*)\cdot\frac{L(\alpha^*,\lambda)}{R(\alpha^*,\lambda)}$$
$$= \frac{\alpha^*+\underline{\alpha}}{2+\alpha^*+\underline{\alpha}}\cdot\frac{\lambda\left[(\bar{\alpha}-\underline{\alpha})+\frac{\bar{\alpha}^2-\underline{\alpha}^2}{2}\right]+(1-\lambda)\left[(\alpha^*-\underline{\alpha})+\frac{(\alpha^*)^2-\underline{\alpha}^2}{2}\right]}{(1-\lambda)\left[(\bar{\alpha}-\alpha^*)+\frac{\bar{\alpha}^2-(\alpha^*)^2}{2}\right]}.$$

The two distribution rate expressions share an identical L/R ratio and the only difference is the ω term. Comparing the two liquidation strategies therefore reduces to comparing $\omega_{LGF}(\alpha^*)$ and ω_{Prop}:

$$\omega_{LGF}(\alpha^*) > \omega_{Prop} \iff \frac{\alpha^*+\underline{\alpha}}{2+\alpha^*+\underline{\alpha}} > \frac{\bar{\alpha}+\underline{\alpha}}{2+\bar{\alpha}+\underline{\alpha}}.$$

because this is strictly increasing, this holds if and only if $\alpha^* > \bar{\alpha}$, which is ruled out by definition since $\alpha^* \in [\underline{\alpha},\bar{\alpha}]$. Therefore:

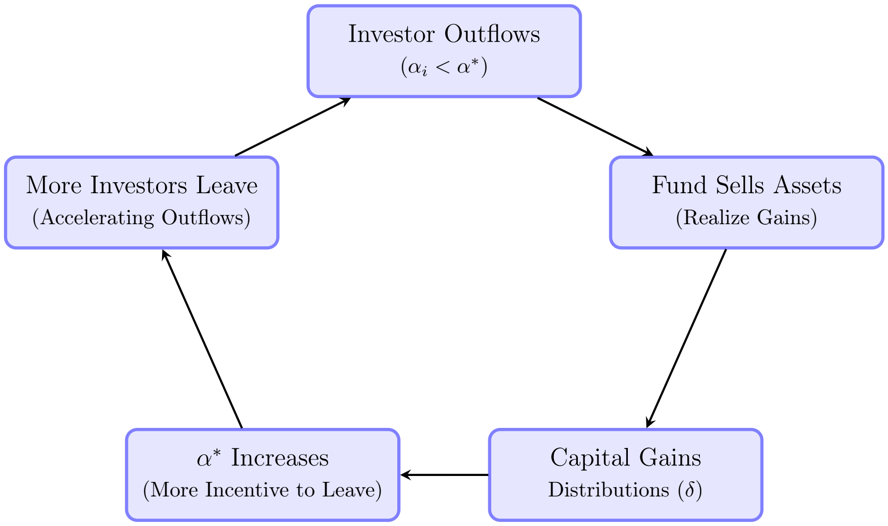
$$\delta^{\mathrm{LGF}}(\alpha^*,\lambda) \le \delta^{\mathrm{Prop}}(\alpha^*,\lambda) \quad \text{for all } \alpha^* \in [\underline{\alpha},\bar{\alpha}],$$

with equality only at $\alpha^* = \bar{\alpha}$, when no investors exit. Holding α^* fixed, least gains first is always weakly reduces distributions relative to proportional liquidation.

Distribution Feedback Loop

The figure below illustrates the mechanism generating multiple equilibria. Investor outflows force asset sales that realize gains. These distributions are passed through to remaining investors, raising α^* and inducing further exits. This feedback creates coordination risk where identical fundamentals can support either stable or fragile outcomes.

Figure A.VIII. Endogenous Distribution Feedback Loop



C.II Model Proofs

C.II1 Proof of Lemma 1

The result follows from the monotonicity of the wealth differential function $g(\alpha_i; \delta, f_{MF}, f_{SMA})$. We first establish that $\frac{\partial g}{\partial \alpha_i} > 0$ under reasonable parameter restrictions, then invoke continuity and the intermediate value theorem to show the existence of a unique threshold.

Terminal wealth in each vehicle is given by equations (10 and 9):

$$W^{SMA}_{i,t=2}(\alpha_i; f_{SMA}) = [1 + \alpha_i(1-\tau)][1 + (1-\tau)(R - f_{SMA})], \tag{A.2}$$

$$\begin{aligned} W^{MF}_{i,t=2}(\alpha_i; \delta, f_{MF}) = 1 &+ (1-\tau)[(1+\alpha_i)(1-\delta)(1+R-f_{MF}) - 1] \\ &+ (1-\tau)(1+\alpha_i)\delta[1 + (1-\tau)(R - f_{MF})]. \end{aligned} \tag{A.3}$$

Expanding and simplifying (A.3):

$$W^{MF}_{i,t=2} = 1 + (1-\tau)(1+\alpha_i)[(1+R-f_{MF})(1-\delta\tau)] - (1-\tau). \tag{A.4}$$

Taking derivatives with respect to α_i:

$$\begin{aligned} \frac{\partial W^{SMA}}{\partial \alpha_i} &= (1-\tau)[1 + (1-\tau)(R - f_{SMA})], \\ \frac{\partial W^{MF}}{\partial \alpha_i} &= (1-\tau)(1+R-f_{MF})(1-\delta\tau) \end{aligned}$$

The derivative of the switching value function $g = W^{MF} - W^{SMA}$ is:

$$\begin{aligned}\frac{\partial g}{\partial \alpha_i} &= (1-\tau)\left[(1+R-f_{MF})(1-\delta\tau) - 1 - (1-\tau)(R - f_{SMA})\right], \\ &= (1-\tau)\left[\tau R - f_{MF} + (1-\tau) f_{SMA} - \delta\tau(1+R-f_{MF})\right].\end{aligned}$$

For monotonicity, we require:

$$\tau R + (1-\tau) f_{SMA} > f_{MF} + \delta\tau(1+R-f_{MF}). \tag{A.5}$$

Under condition (A.5), $\frac{\partial g}{\partial \alpha_i} > 0$ for all $\alpha_i \in [\underline{\alpha}, \bar{\alpha}]$. Since $g(\alpha_i)$ is continuous on the compact interval $[\underline{\alpha}, \bar{\alpha}]$, the intermediate value theorem guarantees a unique threshold $\alpha^* \in [\underline{\alpha}, \bar{\alpha}]$ satisfying $g(\alpha^*) = 0$. Investors with $\alpha_i \geq \alpha^*$ remain in the MF; investors with $\alpha_i < \alpha^*$ switch to the SMA. Condition (A.5) holds when the tax rate τ and return R are sufficiently large relative to the fee differential and distribution rate.

C.II2 Proof of Lemma 2

The distribution rate function is:

$$\delta(\alpha^*, \lambda) = \omega(\alpha^*) \cdot \frac{L(\alpha^*, \lambda)}{R(\alpha^*, \lambda)},$$

where liquidations and remaining assets are:

$$\begin{aligned}L(\alpha^*, \lambda) &= \lambda \int_{\underline{\alpha}}^{\bar{\alpha}} (1+\alpha) p(\alpha)\, d\alpha + (1-\lambda) \int_{\underline{\alpha}}^{\alpha^*} (1+\alpha) p(\alpha)\, d\alpha, \\ R(\alpha^*, \lambda) &= (1-\lambda) \int_{\alpha^*}^{\bar{\alpha}} (1+\alpha) p(\alpha)\, d\alpha.\end{aligned}$$

Part 1: Strictly Increasing Applying the Leibniz integral rule:

$$\begin{aligned}\frac{\partial L}{\partial \alpha^*} &= (1-\lambda)(1+\alpha^*) p(\alpha^*) > 0, \\ \frac{\partial R}{\partial \alpha^*} &= -(1-\lambda)(1+\alpha^*) p(\alpha^*) < 0.\end{aligned}$$

By the quotient rule:

$$\frac{\partial}{\partial \alpha^*}\left(\frac{L}{R}\right) = \frac{(1-\lambda)(1+\alpha^*) p(\alpha^*)[R(\alpha^*) + L(\alpha^*)]}{R(\alpha^*)^2} > 0.$$

For proportional liquidation, ω^{Prop} is constant, so:

$$\frac{\partial \delta}{\partial \alpha^*} = \omega^{Prop} \cdot \frac{\partial}{\partial \alpha^*}\left(\frac{L}{R}\right) > 0.$$

For least-gains-first liquidation:

$$\omega^{LGF}(\alpha^*) = \frac{\int_{\underline{\alpha}}^{\alpha^*} \alpha \cdot p(\alpha)\, d\alpha}{\int_{\underline{\alpha}}^{\alpha^*} (1+\alpha) \cdot p(\alpha)\, d\alpha}.$$

Differentiating yields:

$$\frac{\partial \omega^{LGF}}{\partial \alpha^*} = \frac{p(\alpha^*)}{\left[\int_{\underline{\alpha}}^{\alpha^*} (1+\alpha) p(\alpha)\, d\alpha\right]^2} \int_{\underline{\alpha}}^{\alpha^*} (\alpha^* - \alpha) p(\alpha)\, d\alpha > 0.$$

Therefore:

$$\frac{\partial \delta}{\partial \alpha^*} = \frac{\partial \omega^{LGF}}{\partial \alpha^*} \cdot \frac{L}{R} + \omega^{LGF} \cdot \frac{\partial}{\partial \alpha^*}\left(\frac{L}{R}\right) > 0.$$

Part 2: Strictly Convex Let $A(\alpha^*) = (1-\lambda)(1+\alpha^*)p(\alpha^*)$ and $B(\alpha^*) = R(\alpha^*) + L(\alpha^*)$. From equation (C.II2):

$$\frac{\partial}{\partial \alpha^*}\left(\frac{L}{R}\right) = \frac{A(\alpha^*)B(\alpha^*)}{R(\alpha^*)^2}.$$

Taking the second derivative:

$$\frac{\partial^2}{\partial (\alpha^*)^2}\left(\frac{L}{R}\right) = \frac{A'BR^2 + 2A^2BR}{R^4},$$

where $A'(\alpha^*) = (1-\lambda)[p(\alpha^*) + (1+\alpha^*)p'(\alpha^*)]$ and $B'(\alpha^*) = 0$ (since $L' = -R'$). Both terms in the numerator are positive when the density satisfies:

$$(1+\alpha)p'(\alpha) \geq -p(\alpha) \quad \text{for all } \alpha \in [\underline{\alpha}, \bar{\alpha}]. \tag{A.6}$$

Under condition (A.6):

$$\frac{\partial^2}{\partial (\alpha^*)^2}\left(\frac{L}{R}\right) > 0.$$

For proportional liquidation:

$$\frac{\partial^2 \delta}{\partial (\alpha^*)^2} = \omega^{Prop} \cdot \frac{\partial^2}{\partial (\alpha^*)^2}\left(\frac{L}{R}\right) > 0.$$

For least-gains-first liquidation, the product rule gives:

$$\frac{\partial^2 \delta}{\partial(\alpha^*)^2} = \frac{\partial^2 \omega^{LGF}}{\partial(\alpha^*)^2} \cdot \frac{L}{R} + 2\frac{\partial \omega^{LGF}}{\partial \alpha^*} \cdot \frac{\partial}{\partial \alpha^*}\left(\frac{L}{R}\right) + \omega^{LGF} \cdot \frac{\partial^2}{\partial(\alpha^*)^2}\left(\frac{L}{R}\right) > 0.$$

Therefore:

$$\frac{\partial^2 \delta}{\partial(\alpha^*)^2} > 0$$

.

Both liquidation strategies yield $\frac{\partial^2 \delta}{\partial(\alpha^*)^2} > 0$, establishing strict convexity under the uniform distribution.

C.II3 Proof of Proposition 1

The proof proceeds by characterizing how the equilibrium condition responds to changes in the exit threshold α^*. Define the equilibrium condition implicitly as:

$$\Gamma(\alpha^*) := g(\alpha^*; f_{MF}, f_{SMA}, \delta(\alpha^*, \lambda, \omega)) = 0.$$

Equilibria are fixed points at zero of $\Gamma(\alpha^*)$. We characterize the shape of Γ using its total derivative:

$$\frac{d\Gamma}{d\alpha^*} = \left.\frac{\partial g}{\partial \alpha^*}\right|_{\delta \text{ fixed}} + \frac{\partial g}{\partial \delta} \cdot \frac{\partial \delta}{\partial \alpha^*}. \tag{A.7}$$

From Lemma 1, $\left.\frac{\partial g}{\partial \alpha^*}\right|_{\delta \text{ fixed}} > 0$ (monotonicity in embedded gains) and from Lemma 2, $\frac{\partial \delta}{\partial \alpha^*} > 0$ (more exits increase distributions). Thus the sign of $\frac{\partial g}{\partial \delta}$ determines whether the feedback is stabilizing or destabilizing. From the MF wealth function in equation (10):

$$\begin{aligned}\frac{\partial W^{MF}}{\partial \delta} &= -(1-\tau)[(1+\alpha^*)(1+R-f_{MF}) + (1+\alpha^*)(1+R-f_{MF})\tau] \\ &= -(1-\tau)(1+\alpha^*)(1+R-f_{MF})(1+\tau) < 0. \end{aligned} \tag{A.8}$$

Since W^{SMA} doesn't depend on δ:

$$\frac{\partial g}{\partial \delta} = \frac{\partial W^{MF}}{\partial \delta} < 0.$$

Substituting into equation (A.7):

$$\frac{d\Gamma}{d\alpha^*} = \underbrace{\left.\frac{\partial g}{\partial \alpha^*}\right|_{\delta \text{ fixed}}}_{>0} + \underbrace{\frac{\partial g}{\partial \delta}}_{<0} \cdot \underbrace{\frac{\partial \delta}{\partial \alpha^*}}_{>0}. \tag{A.9}$$

Thus the sign of $\frac{d\Gamma}{d\alpha^*}$ is ambiguous and depends on the relative magnitudes of the direct effect (positive) versus the feedback effect (negative).

Case 1: Outside Option Dominance (Δf large) When $f_{MF} - f_{SMA}$ is large, $W^{SMA} > W^{MF}$ for all α, so $g(\alpha) < 0$ for all $\alpha \in [\underline{\alpha}, \bar{\alpha}]$. Formally, from the wealth functions, the fee differential enters negatively in g:

$$\frac{\partial g}{\partial \Delta f} < 0.$$

Therefore, $\exists\ \overline{\Delta f}(\tau, \underline{\alpha}, \bar{\alpha})$ such that for $\Delta f > \overline{\Delta f}$:

$$g(\bar{\alpha}; f_{MF}, f_{SMA}, \delta(\bar{\alpha}, \lambda, \omega)) < 0.$$

By monotonicity (Lemma 1), $g(\alpha) < 0$ for all $\alpha \leq \bar{\alpha}$, so $\alpha^* = \underline{\alpha}$ is the unique equilibrium.

Case 2: MF Dominance ($\underline{\alpha}$ large, λ small) When embedded gains are uniformly high and liquidity shocks are rare, the exit tax dominates distribution concerns. For sufficiently large $\underline{\alpha}$:

$$g(\underline{\alpha}; f_{MF}, f_{SMA}, \delta(\underline{\alpha}, \lambda, \omega)) > 0.$$

By monotonicity, $g(\alpha) > 0$ for all $\alpha \geq \underline{\alpha}$, so $\alpha^* = \bar{\alpha}$ is the unique equilibrium.

Case 3: Multiple Equilibria (intermediate parameters) When parameters are intermediate such that:

$$g(\underline{\alpha}) < 0 < g(\bar{\alpha}).$$

the intermediate value theorem guarantees at least one interior zero. Multiple interior zeros arise when $\frac{d\Gamma}{d\alpha^*}$ changes sign.

Consider the second derivative:

$$\frac{d^2\Gamma}{d(\alpha^*)^2} = \left.\frac{\partial^2 g}{\partial (\alpha^*)^2}\right|_{\delta} + 2\frac{\partial^2 g}{\partial \alpha^* \partial \delta} \cdot \frac{\partial \delta}{\partial \alpha^*} + \frac{\partial g}{\partial \delta} \cdot \frac{\partial^2 \delta}{\partial (\alpha^*)^2}.$$

From Lemma 2, $\frac{\partial^2 \delta}{\partial (\alpha^*)^2} > 0$ (distribution convexity), and from equation A.8, $\frac{\partial g}{\partial \delta} < 0$. The convexity of δ creates the possibility that $\frac{d\Gamma}{d\alpha^*}$ changes from positive to negative and back to positive, generating an S-shaped curve with three zeros.

C.III Stability Analysis

The distribution feedback mechanism generates multiple equilibria with distinct stability properties. I analyze local stability, which captures robustness to small perturbations in the equilibrium cutoff.

Best-Response Characterization

The equilibrium can be equivalently characterized through the investor best-response function. Given that other investors use cutoff α', which determines the distribution rate $\delta(\alpha', \lambda, \omega)$, the best-response for an individual investor is the cutoff that solves:

$$BR(\alpha'; f_{MF}, f_{SMA}, \lambda, \omega) = \{\alpha \in [\underline{\alpha}, \bar{\alpha}] : g(\alpha; f_{MF}, f_{SMA}, \delta(\alpha', \lambda, \omega)) = 0\}.$$

This function maps the expected cutoff α' to the optimal individual response. An equilibrium cutoff α^* satisfies the fixed-point condition:

$$\alpha^* = BR(\alpha^*; f_{MF}, f_{SMA}, \lambda, \omega).$$

Graphically, equilibria correspond to intersections of the best-response function with the 45-degree line. This alternative characterization is useful for analyzing stability: the slope of $BR(\alpha')$ at the fixed point determines whether small perturbations amplify or dissipate.

Local Stability

Let α^* denote an equilibrium cutoff satisfying equation (23). Suppose the equilibrium threshold is perturbed by $\varepsilon > 0$, so that investors expect the cutoff to be $\alpha^* + \varepsilon$. Let α^p denote the new indifferent investor under this perturbed expectation, solving:

$$g(\alpha^p; f_{MF}, f_{SMA}, \delta(\alpha^* + \varepsilon, \lambda, \omega)) = 0.$$

or equivalently, $\alpha^p = BR(\alpha^* + \varepsilon)$.

Definition 2 (Local Stability) *An equilibrium cutoff* α^* *is* locally stable *if*

$$\lim_{\varepsilon \to 0} \frac{|\alpha^p - (\alpha^* + \varepsilon)|}{\varepsilon} < 1. \tag{A.10}$$

Intuitively, this requires the slope of $BR(\alpha')$ at α^* to be less than one: deviations shrink back toward equilibrium rather than amplify, following (Morris and Shin, 1998).

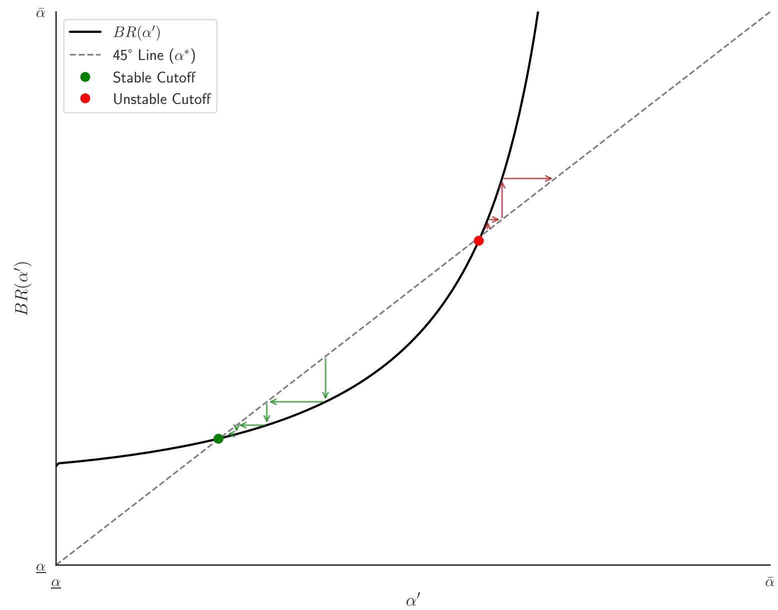
Graphical Interpretation

Figure A.IX illustrates the stability dynamics using the best-response characterization. The investor best-response function intersects the 45-degree line at two points, corresponding to the safe and run equilibria.

At the stable (safe) equilibrium, the best-response curve intersects the 45-degree line from below, with slope less than one. Perturbations generate adjustments that push back toward equilibrium. The horizontal arrows (representing the distance from perturbed belief to best response) are smaller than the corresponding vertical arrows (representing the initial perturbation), satisfying condition (A.10).

At the unstable (run) equilibrium, the best-response curve intersects from above, with slope greater than one. Perturbations amplify rather than dissipate: the horizontal adjustment exceeds the vertical perturbation, causing the system to move away from the unstable equilibrium toward either the safe equilibrium or complete exit.

FIGURE A.IX. Investor Best Response and Stability



Note: Illustration under baseline parameterization ($\tau = 0.20$, $f_{MF} = 0.01\%$, $f_{SMA} = 0\%$, $R = 12\%$, $\alpha \sim$ Uniform$[0, 5]$, $\lambda = 0.05$). The investor best-response function $BR(\alpha')$ intersects the 45-degree line at two points. At the stable low-switching equilibrium, the curve intersects from below (slope < 1); horizontal adjustments are smaller than vertical perturbations. At the unstable high-switching equilibrium, the curve intersects from above (slope > 1); perturbations amplify.

The instability of the run equilibrium arises from the convexity of the distribution rate function (Lemma 2). Near the run equilibrium, small increases in exits generate disproportionately large increases in distributions through two channels. First, the mechanical effect: as more investors exit, the per-capita distribution burden rises because the numerator (liquidations) grows while the denominator (remaining assets) shrinks. Second, under least-gains-first liquidation, the embedded gains ratio $\omega(\alpha^*)$ itself increases as low-basis assets are depleted, further amplifying distributions.

These accelerating distributions create a tipping point. Once exits push past the unstable equilibrium, the feedback loop becomes self-reinforcing: higher distributions induce more exits, which generate even higher distributions, leading to cascading redemptions. The unstable equilibrium thus acts as a threshold separating the basin of attraction for the safe equilibrium from trajectories leading to complete exit.

C.IV Empirical Motivation for Liquidation Strategies

The model considers two liquidation strategies: proportional and least-gain-first. This section provides empirical evidence that index funds sell approximately proportionally, while active funds deviate substantially, motivating both strategies.

Empirical Strategy

Index funds should approximate proportional liquidation to minimize tracking error, selling securities proportional to index weights. Active funds may deviate by targeting specific securities (lowest-gain-first) to reduce taxable distributions.

To test these predictions, I construct a measure of unrealized gains sold at the fund-month level. Let $\text{Outflows}_{i,t}$ denote the magnitude of net outflows for fund i in month t, and let $\text{UG}_{i,t}$ denote the fund's unrealized gains ratio. I define unrealized gains sold as:

$$ug_{it}^{sold} = \text{Outflows}_{it} \times \text{UG}_{it}.$$

Between two consecutive distribution dates, $(d_k, d_{k+1}]$, I aggregate the unrealized gains sold and scale by end-of-period net assets to obtain the theoretical proportional selling rate:

$$\text{Proportional Selling Rate}_{i,(d_k,d_{k+1})} = \frac{\sum_{t\in(d_k,\, d_{k+1}]} ug_{i,t}^{sold}}{NA_{i,d_{k+1}}}.$$

This measures the expected distribution yield under proportional selling. I then regress the proportional selling rate on the realized distribution yield:

$$\text{Proportional Selling Rate}_{it} = \beta_0 + \beta_1 \, \text{Distribution Yield}_{it} + \varepsilon_{it}$$

Under exact proportional liquidation, $\beta_1 = 1$. Values $\beta_1 < 1$ indicate the selective sale of lower-gain assets (consistent with least-gains-first); $\beta_1 > 1$ would indicate a disproportionate sale of higher-gain assets.

Table A.X reports results separately for index and active funds. Index funds exhibit $\beta_1 = 0.68$, close to proportional sales and consistent with minimizing tracking error that restricts asset selection. Active funds show $\beta_1 = 0.37$, indicating a substantial deviation consistent with tax-aware strategies prioritizing low-basis asset sales.

These findings motivate the model's focus on both liquidation strategies: proportional liquidation as a benchmark for index funds (which dominate taxable mutual fund assets), and least-gains-first as a stylized representation of the tax-efficient strategies available to active managers.

TABLE A.X. Testing Proportional Selling in Mutual Funds

	Index Funds	Active Funds
	(1)	(2)
Distribution Yield	0.681***	0.369***
	(0.032)	(0.006)
Observations	1,164	8,827
R-squared	0.284	0.323

Note: The dependent variable is the proportional selling rate defined in equation (C.IV). The unit of observation is a fund-distribution period. Standard errors are clustered at the fund level. Under exact proportional liquidation, the coefficient on distribution yield should equal one.
Source: CRSP, Morningstar Direct, SEC N-CSR, and SEC N-PORT

C.V Calibrating Investor Embeddedness from Fund-Level Data

This section explains how I map fund-level unrealized gains ratios observed in regulatory filings to the distribution of investor-level embedded gains α_i used in the model.

Fund-Level Unrealized Gains

For fund i at date t, regulatory filings (N-CSR and N-PORT) report gross unrealized gains and net assets. I define the fund-level unrealized gains ratio as:

$$\text{Unrealized Gains Ratio}_{it} = \frac{\text{Gross Unrealized Gain}_{it}}{\text{Net Assets}_{it}}.$$

This ratio measures the share of the fund's current value that reflects appreciation rather than cost basis. For example, a ratio of 0.50 means that half of the fund's value consists of unrealized gains.

Investor-Level Embedded Gains

At the investor level, an investor who purchased one dollar of the fund and now has embedded gain α holds a position worth $(1+\alpha)$ dollars. The investor's unrealized gains ratio is:

$$ug(\alpha) = \frac{\alpha}{1+\alpha}, \qquad \alpha > -1.$$

This function maps embedded gains to the unrealized gains ratio: $\alpha = 0$ corresponds to $ug = 0$ (no gains), $\alpha = 1$ corresponds to $ug = 0.5$ (gains equal to cost basis).

Aggregation

I model investor heterogeneity as uniform on $[0, \bar{\alpha}_{it}]$, choosing $\bar{\alpha}_{it}$ so the fund's observed ratio equals the value-weighted average across investors.

The fund-level unrealized gains ratio is the value-weighted average of individual ratios:

$$\text{Fund UG Ratio} = \frac{\int_0^{\bar{\alpha}} \frac{\alpha}{1+\alpha} \cdot (1+\alpha)\, p(\alpha)\, d\alpha}{\int_0^{\bar{\alpha}} (1+\alpha)\, p(\alpha)\, d\alpha}.$$

where $p(\alpha) = \frac{1}{\bar{\alpha}}$ is the density of the uniform distribution and $(1+\alpha)$ serves as the weight (each investor's wealth). The numerator simplifies to $\int_0^{\bar{\alpha}} \alpha\, p(\alpha)\, d\alpha$ because the $(1+\alpha)$ terms cancel.

For a uniform distribution on $[0, \bar{\alpha}]$, this evaluates to:

$$\begin{aligned} E\left[\frac{\alpha}{1+\alpha}\right] &= \frac{\int_0^{\bar{\alpha}} \alpha \cdot \frac{1}{\bar{\alpha}}\, d\alpha}{\int_0^{\bar{\alpha}} (1+\alpha) \cdot \frac{1}{\bar{\alpha}}\, d\alpha} \\ &= \frac{\bar{\alpha}^2/(2\bar{\alpha})}{(\bar{\alpha} + \bar{\alpha}^2/2)/\bar{\alpha}} \\ &= \frac{\bar{\alpha}/2}{1+\bar{\alpha}/2} = \frac{\bar{\alpha}}{2+\bar{\alpha}}. \end{aligned}$$

Setting this equal to the observed fund-level ratio and solving for $\bar{\alpha}$:

$$ug_{it} = \frac{\bar{\alpha}_{it}}{2+\bar{\alpha}_{it}} \quad \Longrightarrow \quad \bar{\alpha}_{it} = \frac{2 \cdot ug_{it}}{1 - ug_{it}}. \tag{A.11}$$

For an example, consider a mutual fund with an unrealized gains ratio of 0.50. Using equation (A.11):

$$\bar{\alpha} = \frac{2 \times 0.50}{1 - 0.50} = \frac{1.0}{0.5} = 2.0.$$

This fund is modeled as having investors with embedded gains uniformly distributed on $[0, 2.0]$. The value-weighted average unrealized gains ratio across these investors is 0.50, matching the observed fund-level ratio. Investors at the lower end of the distribution ($\alpha \approx 0$) face minimal switching costs, while those at the upper end ($\alpha \approx 2.0$) face substantial tax lock-in.

This calibration approach provides a transparent mapping from observable fund-level data to the investor heterogeneity that drives the model's dynamics. The uniform distribution assumption, while stylized, yields closed-form expressions and preserves the essential feature that funds with higher unrealized gains ratios contain investors with greater embedded gains and thus stronger lock-in effects.

Estimating $\bar{\alpha}$

The parameter $\bar{\alpha}$ is calibrated to match five empirical moments from Section 4: the mean and standard deviation of unrealized gains ratios (25.6% and 19.0%), the mean outflow rate (15.3%), the mean distribution yield (5.6%), and the mean twin expense ratio difference (0.30%). For each candidate value of $\bar{\alpha}$, I simulate 10,000 funds with heterogeneous investor bases. Each fund's investor distribution bound $\bar{\alpha}_i$ is drawn from a lognormal distribution with mean $\bar{\alpha}$ and standard deviation 1.2, chosen to match the observed cross-sectional dispersion in Figure 6a. Within each fund, investor embedded gains are drawn uniformly from $[0, \bar{\alpha}_i]$, and the fund-level unrealized gains ratio is computed as the value-weighted average across investors. The calibration minimizes the squared distance between simulated and empirical moments. The estimation yields $\bar{\alpha} = 0.88$. The table below reports the detailed model fit.

TABLE A.XI. Model Fit: Simulated vs. Empirical Moments

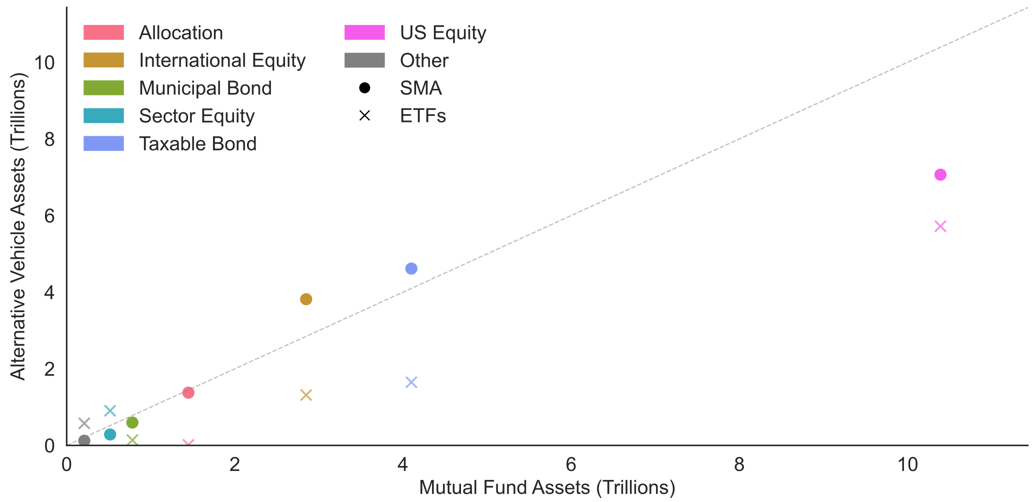
Moment	Data	Model	Diff.	|Error|
Mean UG ratio	0.256	0.269	−0.013	5.2%
Standard Deviation of UG ratio	0.190	0.145	0.046	24.0%
Mean outflow rate	0.153	0.133	0.019	12.6%
Mean distribution yield	0.056	0.055	0.001	1.7%
Mean expense ratio	0.003	0.002	0.006	33.3%

Note: Additional parameters include $\lambda = 0.05$, $\tau = 0.20$, $R = 0.12$, and $f_{SMA} = 0.0$.

D Investment Vehicle Universe

SMAs and ETFs have expanded to the point where they now rival MFs, making it important to benchmark all three vehicles against one another. Their assets have grown dramatically across categories, with SMAs and ETFs spanning a similarly broad range of asset classes as MFs. To place the three vehicles in broader context, Figure A.X presents a cross-sectional comparison of total net assets by US Category Group for MFs versus SMAs and ETFs as of December 2024. For SMAs, category-level assets consistently lie close to the 45-degree line, indicating that their scale is comparable to that of MFs. ETFs also approach parity with MFs in many categories, underscoring their growth.

FIGURE A.X. Investment Vehicle Assets by US Category Group



Note: This scatter plot displays the relationship between MF assets (x-axis) and SMA/ETF assets (y-axis) across different Morningstar US Category Groups as of December 2024, measured in trillions of dollars. Each category is represented by a distinct color, with dots indicating SMA assets and X markers showing ETF assets. The 45-degree reference line helps identify where asset values are equal across investment vehicle types. Lying on the 45-degree line means the two investment vehicle universes, for that category, are the same size. The "Other" category encompasses Miscellaneous, Nontraditional, Equity, Commodities, and Alternative categories from Morningstar's US Category Group classification system.
Source: Morningstar Direct

D.I Separately Managed Accounts

An SMA is a professionally managed investment account with direct investor ownership of underlying securities, unlike the pooled structure of MFs and ETFs. This paper focuses on SMAs in the Morningstar sample where managers self-report composite-level performance. These SMAs closely resemble MFs: managers often offer the same strategy in both formats as explicit twins with identical investment objectives and management.[2]

Asset managers emphasize three advantages of SMAs: customization, transparency, and tax benefits, all of which stem from their structural design. First, direct ownership allows for customization of portfolio management. Investors may request reasonable restrictions to a manager's strategy, such as limiting a certain stock or to align portfolios with social, political, or environmental values.[3] Second, in terms of transparency, SMAs go beyond both ETFs and MFs. ETFs provide more visibility than MFs, with many disclosing daily holdings, while MFs typically

[2]For example, Columbia Threadneedle offers its Large Cap Growth strategy as both SMA and MF with nearly identical language.
[3]From anecdotal evidence of speaking with managers, I find this is usually small changes to the portfolio.

report only quarterly and with a lag.[4] By contrast, many SMAs provide regular reporting and even real-time access to balances, transactions, fees, and cost bases. This ability to view cost bases is a distinctive feature of SMAs relates to the first tax advantage of SMAs.

Direct ownership of the underlying securities provides several tax advantages. First, individual ownership allows investors to control the timing of realized gains. In a MF, managers must apply a uniform cost-basis method across all investors. For example, index funds typically sell securities proportionally to maintain target weights, limiting the ability to dispose of only high-basis lots. As highlighted later in the model, a MF manager must weigh the tradeoff between minimizing current realized gains and preserving flexibility for future redemptions: selling lower-gain lots reduces immediate distributions but leaves the MF concentrated in high-gain securities, increasing the likelihood of larger taxable distributions later. By contrast, the reporting transparency in SMAs provides investors with full visibility into their own cost bases, which enables more targeted tax-management strategies such as loss harvesting and selective realization.

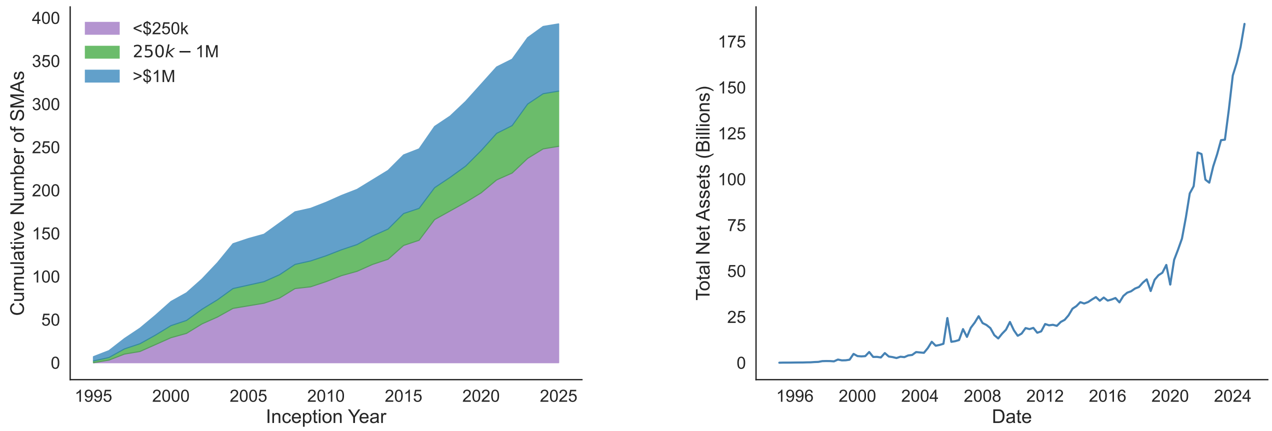
Moreover, the SMA structure creates the ability to conduct tax-loss harvesting. SMA managers can sell under performing assets to generate losses that offset gains both within the SMA portfolio and across the investor's broader holdings. If losses exceed gains, up to $3,000 may be used annually to offset ordinary income, with the remainder carried forward indefinitely. Managers can reinvest the proceeds into similar but not identical securities, subject to wash-sale rules prohibiting repurchases of the same security within 30 days. Advances in technology and lower transaction costs have made tax-loss harvesting scalable, and recent evidence shows it can significantly enhance after-tax returns (Chaudhuri, Burnham, and Lo, 2020; Khang, Paradise, and Dickson, 2021). This feature is unique to SMAs: while ETFs and MFs can carry forward losses internally, they cannot distribute losses to investors for use against other portfolio gains or income.

Figure A.XI illustrates rapid growth in explicitly tax-efficient SMA strategies. Panel (a) shows the number of tax-efficient composites by minimum investment level, while Panel (b) shows total assets have grown to $175 billion, with growth concentrated among strategies with lower minimum investments that increasingly compete with retail MF share classes.

Critically for this paper, direct ownership insulates SMA investors from capital gains distributions triggered by other investors' redemptions. In an SMA, redemptions by other investors have no tax implications for remaining investors. This contrast highlights the central tax friction examined in this paper: the distribution externality in pooled vehicles.

[4]Under SEC Rule 6c-11 of the Investment Company Act of 1940, ETFs must post daily portfolio holdings on their websites. Form N-PORT requires MFs to file monthly holdings with the SEC no later than 60 days after quarter-end, which are then made public at a quarterly frequency.

FIGURE A.XI. Tax-Efficient SMAs



(a) SMAs Composites with Tax Efficiency by Minimum Investment

(b) Total Net Assets of Tax Efficient SMAs

Note: Panel (a) shows the number of SMA composite that has a tax efficient strategy by minimum investment. Panel (b) show the total net assets for tax efficient strategies. A strategy is classified as tax-efficient if it meets any of the following: (i) "tax" appears in the Investment Philosophy or Strategy Investment Philosophy; (ii) the Name includes "tax," "Tx," or "Loss" (but not "Loss Averse" or "Stop/Loss"); (iii) the Name is a manual inclusion such as "GS TACS"; or (iv) any Tax Customized Separate Account field is marked "Proactive." Strategies with "Tax-Exempt" in the name are excluded.
Source: Morningstar Direct

E Vanguard Case

Until recently, Vanguard offered two parallel Target Retirement Funds (TRFs):

- **Investor Funds:** Low minimum investment ($1,000), higher expense ratio (0.14%), accessible to individuals and smaller retirement plans
- **Institutional Funds:** High minimum investment ($100 million), lower expense ratio (0.09%), designed for large retirement plans

Although both held similar portfolios, they were structured as distinct mutual funds rather than share classes. Investors switching between them had to redeem and repurchase, triggering capital gains for taxable accounts. Importantly, while designed as target retirement funds, these MFs were held in both tax-exempt retirement accounts and taxable brokerage accounts.

In December 2020, Vanguard reduced the Institutional TRF minimum from $100 million to $5 million, allowing smaller retirement plans to access lower fees. Vanguard had considered merging the fund series (which would have avoided tax consequences) but chose the minimum reduction approach, which preserved more fee revenue ($87 million difference).

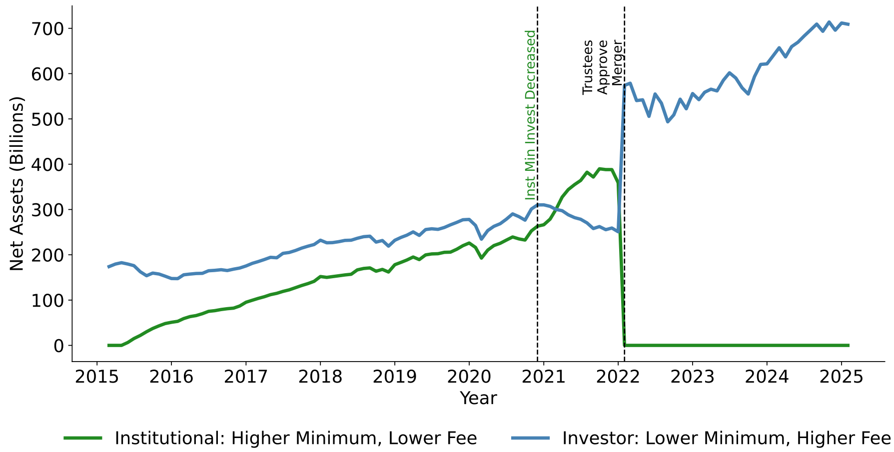
This triggered substantial redemptions from Investor TRFs. From December 2020 to January 2022, Investor TRF assets declined from $310 billion to $251 billion, while Institutional TRF assets increased from $263 billion to $359 billion (Figure A.XIIa). Redemptions forced asset sales that realized substantial gains, generating distributions reaching 10% of NAV in 2021 (Figure A.XIIb). These distributions created large unexpected tax liabilities for remaining taxable investors.

The SEC determined Vanguard failed to adequately disclose these potential tax consequences and brought an enforcement action in 2025 (SEC Release No. 33-11359). Vanguard agreed to pay $135 million in remediation to harmed taxable investors and merged the fund series in February 2022, eliminating the fee differential.

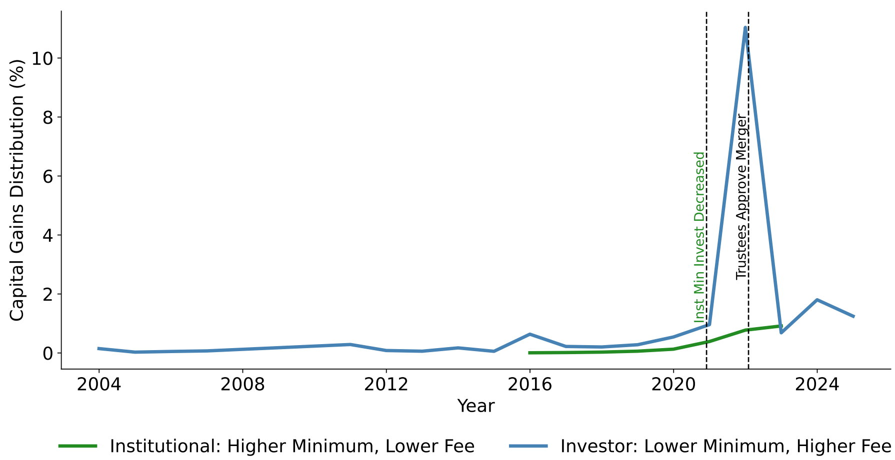
This episode illustrates the model's key dynamics: the minimum reduction created switching incentives, but redemptions by tax-exempt retirement plans (facing no switching costs) generated distributions imposed on remaining taxable investors. This demonstrates the strategic complementarity where exit decisions impose externalities on those who stay, with distribution burdens increasing with aggregate redemptions.

Figure A.XII. Net Assets and Capital Gains Distributions by Vanguard Target Retirement Funds

(a) Total Net Assets



(b) Capital Gains Distributions



Note: The figure shows the total net assets and capital gains distributions for Vanguard Target Date Funds. Funds included: 'VFFVX', 'VTTSX', 'VTOVX', 'VTINX', 'VTXVX', 'VTTVX', 'VFIFX', 'VFORX', 'VTHRX', 'VTWNX', 'VTTHX', 'VTIVX', 'VTENX', 'VSXFX', 'VLXVX', 'VITLX', 'VIRTX', 'VITVX', 'VITWX', 'VITRX', 'VRIVX', 'VTTWX', 'VITFX', 'VIRSX', 'VTRLX', 'VIVLX', 'VILVX', 'VSVNX'. Distributions include short term and long term capital gains.
Source: Morningstar, and CRSP