Norwegian School of Economics

Essays on Mutual Funds

André Wattø Sjuve

Dissertation submitted to the Department of Business and Management Science, Norwegian School of Economics, in partial fulfillment of the requirements for the degree of Philosophiae Doctor (PhD)

Bergen,
August 2022

Acknowledgments

Arnold Schwarzenegger said that there is no such thing as a self-made man and I am no exception. I would therefore like to express my appreciation to those who have gotten me to where I am today.

First of all, I would like to extend my greatest gratitude to my main supervisor Trond Døskeland for his support and guidance throughout the entire PhD programme. The time and resources given to me are comparable to none. I also want to thank Petter Bjerksund for his collaboration on one of the projects in this thesis, helpful feedback, and, in general, valuable discussions and conversations. Close cooperation with my fellow coauthor and office buddy Andreas Ørpetveit has been enriching and another source of motivation throughout. Our common research interests have led to numerous interesting discussions and potential future research projects. During my PhD, I have also been fortunate to become a part of a well-composed research group, from which I hope many papers will follow.

Furthermore, I would like to thank all my colleagues in the Department of Business and Management Science for a great and inspiring work environment, as well as with good help in everything administrative.

Last but certainly not least, I want to express my deepest appreciation to the people closest to me. Especially, I would like to thank my father, Kjell, for without whom I would not be here, not have considered this path, and much less gotten through. To my mother, Trine, and brother, Jørgen, extended family, and friends thank you for your patience, love, and support.

Bergen, August 2022
André Wattø Sjuve

Contents

Introduction **iii**

1 Are Sustainable Investors Fee Sensitive? **1**

2 Forced to be Active: Evidence From a Regulation Intervention **42**

3 Do Fees Matter? Investor's Sensitivity to Active Management Fees **100**

Introduction

The three papers of this thesis focus on actively managed equity mutual funds. Mutual funds are important as individuals are increasingly more responsible for securing their future financial welfare by acting as their own money managers. This individualized responsibility along with the complex nature of financial markets results in an increased demand for financial advice and products that can help, of which mutual funds hold a central position.

Fund managers in active mutual funds make investment decisions on behalf of their investors. Evaluating their performance is normally done by comparing realized returns against a predefined benchmark index (typically, a market index). Mutual funds, therefore, sell the potential for above-market returns. Investors are usually required to pay a premium over the index alternative for this potentiality.[1] If managers are sufficiently skilled to provide investors with a realistic opportunity for above-market returns, the premium they charge may be justified. Nevertheless, the literature is filled with studies that document how active funds, on average, struggle to deliver positive returns net of fees (see, e.g., Fama and French 2010; Ferreira et al. 2013; Leippold and Rueegg 2020).

Despite the rapid rise and popularity of passively managed low-cost index funds and ETFs (see, for example, Cremers et al. 2016), actively managed funds continue to manage trillions of dollars (Investment Company Institute, 2022). Their dubious track record implies that the average investor would have been better off by investing passively. This suggests that many investors are not knowledgeable enough to adequately evaluate the quality of their funds.[2] It is therefore important to reduce the information gap between fund managers and investors in order to ensure investor protection and continued participation in financial markets.

Traditionally, choosing a mutual fund has been done with the aim of finding the fund most likely to deliver the highest returns. However, increasing attention and interest in sustainability

[1] See, for example, Morningstar (2019)

[2] While passive funds are a part of the financial landscape, investing in them does not require the same evaluation of quality since the primary objective of index funds is to replicate the returns of a market index.

has extended the basis of decision-making for investors beyond this, and sustainable funds have been shown to attract investor capital (Hartzmark and Sussman, 2019). Thus, mutual funds no longer compete for flows solely on the basis of performance. This additional decision variable in the investment process further adds to the complexity of the investment decision. Moreover, despite the popularity of the term "sustainability", it lacks a clear and consistent definition. This shortcoming, in combination with investor demand for sustainability, has given rise to the problem of greenwashing in the fund industry.[3] For capacity-constrained consumers it can be difficult to evaluate whether the fund's investment policies align with what they claim to do.

Apart from greenwashing, another real-life example of how information asymmetry can induce opportunistic behavior by asset management providers is the existence of closet index funds. These funds are marketed and sold as active despite closely following the benchmark index (see, for example, Cremers and Petajisto (2009) for a discussion on U.S. closet index funds and Cremers et al. (2016) for a discussion of international closet index funds). Consequently, their investors only receive part of the service they pay for. In recent years, financial supervisory authorities (FSAs) around the world have made this matter a priority. Among the first to intercede were the Scandinavian countries.[4] The consequences of this intervention in Norway was a first-of-its-kind lawsuit against one of its largest domestic funds. Ultimately, the Norwegian Supreme Court ruled that the fund had not been managed in a way that gave the unitholders the financial risk exposure – and thus the possibility of outperformance of the benchmark index – for which they had paid. As a result, the fund had to return 0.8% of the annual management fee from 2010 to 2014 to their investors.[5] In the U.S. lawsuits are in progress (see, for example, Financial Times, Nov. 2021).

Being financially literate is necessary to be able to properly evaluate investment choices and navigate financial markets. Knowledgeable investors face a smaller information gap in interactions with finance professionals, and they are less likely affected by market frictions. Improving investors' information set is, therefore, an important part of ensuring their confidence and participation in financial markets. There has been an ongoing discussion of late on whether investors use a full asset pricing model Barber et al. (2016); Berk and van Binsbergen (2016) or react to easily available signals (see, e.g., Ben-David et al. (2022) and Evans and Sun (2021)

[3] The term greenwashing is defined as "expressions of environmentalist concerns especially as a cover for products, policies, or activities", see Fidelity

[4] See Financial Supervisory Authority of Denmark (2013); Norwegian Ministry of Finance (2015); Financial Supervisory Authority of Sweden (2015) for details about the interventions in Scandinavian countries

[5] For more information about the verdict see Lovdata HR-2020-475-A

for evidence on how Morningstar ratings drive fund flows) when allocating capital to mutual funds.[6]

This thesis is organized into three self-contained empirical papers and has a particular focus on issues in the active mutual fund industry related to the role of fees, the extent of financial sophistication for mutual fund investors, as well as the conflicts that can arise when the gap between provider and customer is pronounced. I briefly summarize them in the following paragraphs.

Paper 1. Are Sustainable Investors Fee Sensitive?

This paper analyzes whether sustainable investors are sensitive to fees and whether investors in funds with a higher sustainability rating are more sensitive to fees than investors in non-sustainable funds. Using an international sample of equity mutual funds, I find that they in general are more sensitive to fees and that this sensitivity increases with Morningstar's globe ratings. Moreover, investors in funds without an explicit sustainability objective are more sensitive to fees than investors in funds with such an objective. The results indicate that sustainable investors elicit a stronger response to fees than other investors, but that those investing in sustainable funds without a sustainability mandate respond the most.

[6] Morningstar ratings are purely quantitative and mostly based on past performance. Consequently, findings of investors responding to such ratings might simply be an indication of them being performance-chasers (see, for example, Sirri and Tufano (1998) and Ferreira et al. (2012)

Paper 2. Forced to be Active: Evidence From a Regulation Intervention

with Petter Bjerksund, Trond Døskeland and Andreas Ørpetveit

Closet indexers are low-activity mutual funds that are sold and marketed as active. Their investors therefore only receive part of the service they pay for. Supervisory authorities all over the world are now considering how these funds should be regulated. We examine evidence from interventions carried out by Scandinavian regulators. The impact is identified by comparing scrutinized Scandinavian closet index funds with similar unaffected European funds. Given the choice between reducing fees or increasing activity, the scrutinized funds opt for the latter. Although this results in a more actively managed fund, performance deteriorates. Thus, regulation leads to the worst outcome.

Paper 3. Do Fees Matter? Investor's Sensitivity to Active Management Fees

with Trond Døskeland and Andreas Ørpetveit

The active mutual fund equilibrium model developed by Berk and Green (2004) predicts that fees should not matter for investors' mutual fund choices. However, there is often a difference between what a rational model indicates and what investors actually do. In this study, we look at how fees influence demand for active mutual funds by analyzing time variation in active funds' fees. Since investors should not pay "alpha fees" for "beta performance", we measure fees both as an excess fee above the passive alternative and as the unit price of active management, where active share represents the activity level. Using international data, we identify a negative time-series relation between fee measures and fund flows. The results also hold after controlling for Morningstar ratings.

References

Barber, B. M., Huang, X., and Odean, T. (2016). Which Factors Matter to Investors? Evidence from Mutual Fund Flows. *The Review of Financial Studies*, 29(10):2600–2642.

Ben-David, I., Li, J., Rossi, A., and Song, Y. (2022). What do mutual fund investors really care about? *The Review of Financial Studies*, 35(4):1723–1774.

Berk, J. B. and van Binsbergen, J. H. (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, 119(1):1–23.

Cremers, K. M. and Petajisto, A. (2009). How active is your fund manager? a new measure that predicts performance. *The review of financial studies*, 22(9):3329–3365.

Cremers, M., Ferreira, M. A., Matos, P., and Starks, L. (2016). Indexing and active fund management: International evidence. *Journal of Financial Economics*, 120(3):539–560.

Evans, R. B. and Sun, Y. (2021). Models or Stars: The Role of Asset Pricing Models and Heuristics in Investor Risk Adjustment. *The Review of Financial Studies*, 34(1):67–107.

Fama, E. F. and French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance*, 65(5):1915–1947.

Ferreira, M. A., Keswani, A., Miguel, A. F., and Ramos, S. B. (2012). The flow-performance relationship around the world. *Journal of Banking & Finance*, 36(6):1759–1780.

Ferreira, M. A., Keswani, A., Miguel, A. F., and Ramos, S. B. (2013). The determinants of mutual fund performance: A cross-country study. *Review of Finance*, 17(2):483–525.

Financial Supervisory Authority of Denmark (2013). Market developments 2013. Available at `https://www.dfsa.dk/~/media/Tal-og-fakta/2014/Markedsudvikling/Market-developments-2013-Investment-associations.pdf?la=en`, Last accessed on 2022-06-13.

Financial Supervisory Authority of Sweden (2015). Consumer protection on the financial market. Available at `https://www.fi.se/contentassets/9791b37d1d074ed4a3fc193786d449f6/konsumentrapp_2015engny.pdf`, Last accessed on 2022-06-13.

Hartzmark, S. M. and Sussman, A. B. (2019). Do investors value sustainability? a natural experiment examining ranking and fund flows. *The Journal of Finance*, 74(6):2789–2837.

Investment Company Institute (2022). A Review of Trends and Activities in the Investment Company Industry. Available at `https://www.icifactbook.org/`, Last accessed on 2022-06-13.

Leippold, M. and Rueegg, R. (2020). How rational and competitive is the market for mutual funds? *Review of Finance*, 24(3):579–613.

Morningstar (2019). U.S. Fund Fee Study. Available at `https://www.morningstar.com/articles/925303/2018-morningstar-fee-study-finds-that-fund-prices-continue-to-decline`.

Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *The journal of finance*, 53(5):1589–1622.

Are Sustainable Investors Fee Sensitive?

André Wattø Sjuve

Abstract

This paper analyzes whether sustainable investors are sensitive to fees and whether investors in funds with a higher sustainability rating are more sensitive to fees than investors in non-sustainable funds. Using an international sample of equity mutual funds, I find that they in general are more sensitive to fees and that this sensitivity increases with Morningstar's globe ratings. Moreover, investors in funds without an explicit sustainability objective are more sensitive to fees than investors in funds with such an objective. The results indicate that sustainable investors elicit a stronger response to fees than other investors, but that those investing in sustainable funds without a sustainability mandate respond the most.

Keywords: Sustainability; ESG; Fund Fees; Mutual Funds; Fund Flows; Ratings

JEL Classification D14; G11; G24

1.1 Introduction

Sustainability has grown exponentially in recent years, and policymakers have advocated for more transparency in the sustainability of mutual fund portfolios to facilitate for larger flows to sustainable investments. As a result, it is now a salient feature of investor information.[1] A prominent example is Morningstar's globe ratings that rank the sustainability of a fund's portfolio.[2] After its introduction, Hartzmark and Sussman (2019) showed that U.S. mutual fund investors allocated more money to funds with the highest rating and less to those with the lowest rating.[3] Along with the increased focus on sustainability, 2021 saw money held in sustainable mutual funds and ESG-focused exchange-traded funds globally reaching 2.7 USD trillion, representing a 53% growth from 2020 (Bloomberg, 2022).

As the title suggests, the aim of this paper is to answer the question of whether investors' demand for sustainability is sensitive to fees. There are different approaches to sustainable investing, but in most cases the investment opportunity set becomes constrained.[4] It is argued in standard portfolio choice theory that constraining the investment universe cannot improve performance. For investors in sustainable funds who at the same time seek to obtain an adequate financial return on those investments, fees should therefore matter more.

Whether or not sustainability actually inhibits financial performance is debated. A fund's performance could be improved by focusing on sustainability if sustainability is positively related to the stocks' future returns and most market participants do not take it into account (Pedersen et al., 2021). However, the increasing demand for sustainable assets in recent years could potentially have increased their valuations and reduced future expected returns (Heinkel et al., 2001; Pástor et al., 2021; Pedersen et al., 2021).[5] A better understanding of how investors relate to fees in the area of sustainable investing could help to improve our understanding of how they deal with the conflicting evidence and confusion when it comes to the potential for financial outperformance.

Focusing on demand sensitivity to changes in investors' economic conditions, Pástor and Vorsatz (2020) and Döttling and Kim (2021) use the COVID-19 induced market crash in 2020 to

[1] For an illustration, visit the Morningstar page of any given mutual fund such as Fidelity Funds - Sustainable US Equity Fund A-USD to see how much of the front page is dedicated to communicating the sustainability information of the fund.
[2] The globe rating is a discrete rating scheme that assigns from one to five globes.
[3] See also Ammann et al. (2019) for international evidence.
[4] See Sustainable investing explained for examples.
[5] The confusion surrounding the effects of ESG on financial performance leads to frequent discussions in the press. See, for example, ESG outperformance narrative "is flawed", new research shows, Financial Times, May 3, 2021.

study how sensitive demand is to a negative income shock. Pástor and Vorsatz (2020) find that investors remained focused on sustainability, while Döttling and Kim (2021) document fragile demand from investors when faced with the economic stress induced by the pandemic. Consequently, investors view sustainability as a positive attribute, but how sensitive their demand is to changes in economic conditions remains an ongoing inquiry. This paper thus contributes to the literature by focusing on demand sensitivity through investigating how investors respond to changes in mutual fund fees rather than how they respond to a negative income shock.

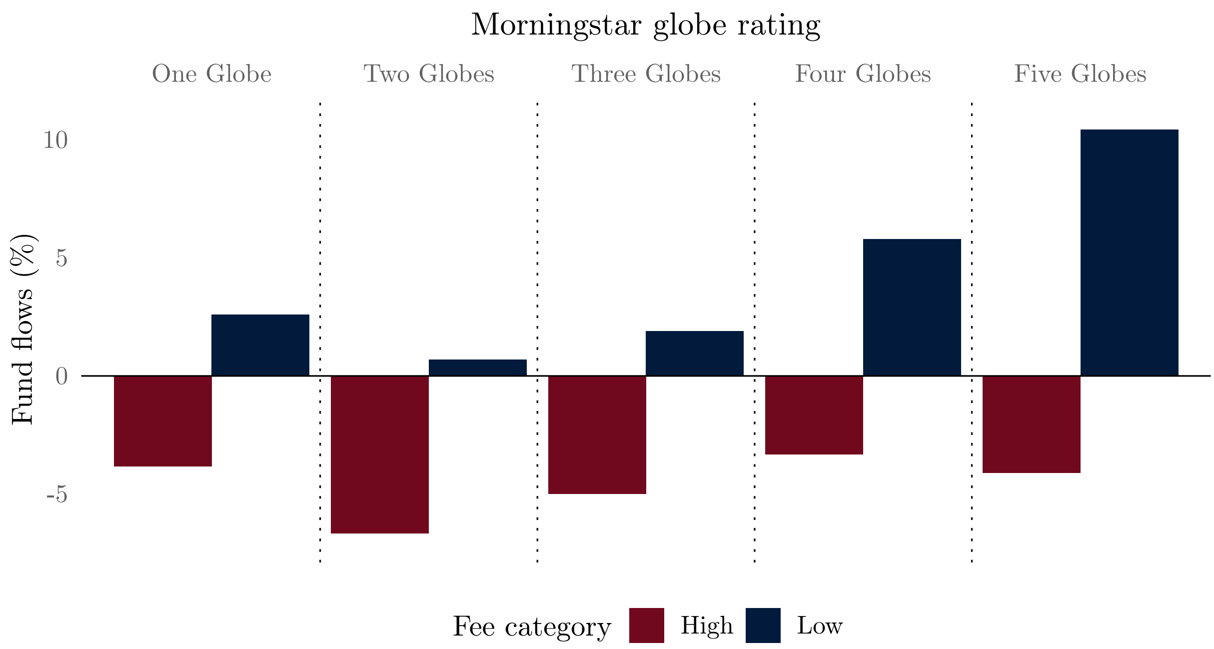
Sustainable funds do not all have the same scope. Some have an explicit sustainability objective while others do not. A fund with a five-globe rating from Morningstar is not required have to have an explicit sustainability goal to receive a top rating. Investors in these funds do not need to be steered by the same goals. To that end, and to better understand the demand for sustainability, recent studies have focused on incorporating heterogeneous investors in their models (see, Pástor et al. 2021; Pedersen et al. 2021; Oehmke and Opp 2020; Humphrey et al. 2021; Goldstein et al. 2022). Additionally, Riedl and Smeets (2017) document that social preferences weigh heavier than financial gain among investors who hold socially responsible mutual funds. The attention given to a heterogeneous mix of investors in the literature and the empirical findings on explicit sustainable funds motivates the question of whether investors who opt into a fund with an explicit sustainability objective respond differently than an investor in a fund without one.

In this study, I utilize an international sample of actively managed equity mutual funds and Morningstar's sustainability (globe) ratings as a proxy for sustainability to examine whether sustainable investors are sensitive to fees, whether they are more sensitive to fees than investors in non-sustainable funds, and whether investors in different types of sustainable funds respond differently. The results show that fund flows vary significantly in response to changes in fees when conditioning on sustainability. Distinguishing between explicitly sustainable funds and regular funds separately also reveals that it is the latter group of funds that makes up the predominant drivers of this variation. Given that mutual fund shares are issued and redeemed at a fixed price, essentially no matter the quantity, it is possible to study the behavior of investors without considering the counterparty to the trade, the mutual fund. As such, the laboratory of mutual funds represents a rare opportunity to study demand without having to consider its impact on price.

Figure 1.1 illustrates the main finding of this paper: investors in sustainable funds are sensitive to fees. Moreover, investors in the most sustainable mutual funds exhibit a higher sensitivity than investors in lower rated mutual funds. This is in line with the findings of Döttling and Kim (2021). Figure 1.1 plots the globe ratings on the x-axis and the average annualized monthly net flows on the y-axis. The bars distinguish between funds with high and low fees.[6] Independent of the globe rating, high-fee funds have an average outflow of capital and low-fee funds an inflow of capital. The difference between the two groups becomes more pronounced as the globe rating increases. As a result, the difference-in-differences estimate between the five-globe and one-globe funds is positive and indicative of investors in more sustainable funds eliciting a stronger response to fees compared to the lowest rated funds.

Figure 1.1.
Fund flows by globe ratings and fees

This figure shows how capital flows to and from actively managed mutual funds vary when sorting on globe ratings and fund fees (one globe is assigned to the funds in the bottom 10%, and five globes to the funds in the top 10% of eligible funds). In each domicile-month (time t), funds that have been assigned a globe rating are sorted into deciles based on fees independently of their globe rating. After removing year-by-month fixed effects, annualized average monthly net flows in the following month ($t+1$) are plotted, for each globe rating. The fee category labeled "Low" contains the funds ranked as the least expensive (rank 4 and 5) and the "High" category contains the funds ranked the most expensive (rank 1 and 2).



The effect size related to Figure 1.1 is both statistically and economically robust to different model specifications. Panel regressions analyzing the interaction effect between fees and globe

[6] In each domicile-month (time t), funds that have been assigned a Morningstar globe rating are sorted into deciles based on fees independently of their globe rating. Fund flows are adjusted for year-month fixed effects.

ratings indicate that investors in funds with a five-globe rating are significantly more sensitive to fees than investors in average rated funds. For example, a one standard deviation increase in fees is associated with a decrease in net flows of -0.43% ($= -0.70 \cdot 0.62$) for the five-globe funds.[7] This number exceeds the overall average net flow. In other words, conditioning flows on fees and sustainability ratings implies fluctuations in the conditional expected net flow that are of first-order economic importance for the highest rated funds. For the one-globe funds, a one standard deviation increase in fees is associated with a decrease in net flows of -0.11% ($= -0.179 \cdot 0.62$). Testing the difference in interaction coefficients between five-globe and one-globe funds also produces statistically significant results.

When dividing the sample into funds with and without an explicit sustainability objective, similar differences in fee-sensitivity are revealed between the most sustainable funds and the average rated funds. The biggest difference lies in comparing the extreme ranks, where there is only a statistically significant difference for the group of funds without an explicit sustainability objective. These results have implications for how we understand heterogeneity in the demand for sustainable investing.

The rest of the paper is organized as follows: Section 1.2 details the sample and presents summary statistics, the main results are presented in Section 1.3, additional analyses are detailed in Section 1.4 and concluding remarks can be found in Section 1.5.

1.2 Measures and sample

1.2.1 Morningstar sustainability ratings

In 2016, Morningstar published a rating scheme that ranked the sustainability of funds' portfolios with the aim of enabling investors to evaluate how different funds meet environmental, social, and governance standards. Referred to as the globe ratings, they range from one (low) to five (high) globes based on the funds' portfolio sustainability scores. The score is an asset-weighted average of the ESG risk score of the underlying securities held in the portfolio. To be scored, 67% of the fund's portfolio companies must be associated with an ESG risk rating from Sustainalytics.[8] Until late 2019, a higher portfolio sustainability score was better, as it indicated that the portfolio companies were better at addressing ESG issues. In a 2019 revision,

[7] The standard deviation is measured as the average standard deviation within each fixed effects group. See Table 1.4 for details.

[8] Sustainalytics is a company specializing in ESG ratings of companies and was bought by Morningstar in 2020, see Morningstar, Inc. Completes Acquisition of Sustainalytics, 2020 for details.

Morningstar incorporated Sustainalytics' new ESG risk rating, where the portfolio sustainability score indicates the degree to which the enterprise value of the fund's underlying holdings is at risk. Up until 2018, a higher score was considered better, whereas after 2019, a lower score is better as it means less ESG risk in the portfolio. Regardless of these changes, the portfolio sustainability score takes on a value between 0 and 100.

A fund's globe rating is based on the percentile rank of its portfolio sustainability score relative to other funds in the same Morningstar global category (MSGC), meaning that systematic differences in the investment opportunities' ESG scores of funds with different specializations (e.g., growth vs. value) do not affect the globe ratings.[9] Only funds belonging to categories with at least ten funds are ranked. These ratings and their methodology were unknown to managers and investors before their introduction in March 2016. Correspondingly, institutional and retail flows increased (decreased) for funds with the top (bottom) globe rating after March 2016 (Hartzmark and Sussman, 2019). The attention-grabbing and easy-to-process nature of the globe ratings make them appealing to both retail and institutional mutual fund investors.

1.2.2 Empirical strategy

Fund flows comprise the main variable of interest. According to, Chevalier and Ellison (1997) and Sirri and Tufano (1998), this is defined as the net growth rate in total net assets (TNA) that is not due to dividends and capital gains on the assets under management, but to new external money. Net flow ($\mathrm{NF}_{i,t}$) for fund i at time t is calculated as:

$$\text{Net flow}_{i,t} = \frac{\mathrm{TNA}_{i,t}}{\mathrm{TNA}_{i,t-1}} - (1 + r_{i,t}), \tag{1.1}$$

where $\mathrm{TNA}_{i,t}$ is USD total net assets for fund i at the end of period t and $r_{i,t}$ is fund i's total return from $t-1$ to t. As no information on the timing of new investments is available, Equation 1.1 assumes that flows occur at the end of each period. Flows are noisy and may vary systematically based on characteristics such as size. To address this, a normalized flow measure is made, as demonstrated by Hartzmark and Sussman (2019), to mitigate concerns that the results are driven by such properties. The variable is constructed by splitting funds into deciles based on size (within each country) and assigning each fund to percentiles based on flows within each size decile each month. A complete reference of the variables used, as well as their sources,

[9] Before the methodology revision, the peer groups were those belonging to the same Morningstar category.

can be found in Appendix 1.A, Table 1.A.1.

To empirically test the questions outlined in the introduction, the following baseline regression is specified:

$$\text{Net flow}_{i,t} = \varphi + \beta\text{Fees}_{i,t-1} + \sum_g \gamma_g \text{Globe}_{g,i,t-1} + \sum_g \mu_g \text{Fees}_{i,t-1} \times \text{Globe}_{g,i,t-1} + \theta\mathbf{X}_{i,t-1} + \varepsilon_{i,t} \tag{1.2}$$

where Net flow$_{i,t}$ follows Equation 1.1 and is the percentage fund flows for fund i in month t. Globe$_{g,i,t-1}$ is an indicator variable that equals one for funds in each globe rating category, and zero otherwise.[10] X$_{i,t-1}$ is a vector of additional controls associated with variation in fund flows: lagged returns, log of fund size, log of family size, log of fund age, and performance ratings (Hartzmark and Sussman, 2019; Pástor and Vorsatz, 2020). In addition to these, the log of each country's industry size is controlled for. Given that the globes are assigned monthly within each Morningstar global category, the main source of variation used to estimate the parameters in Equation 1.2 comes from the monthly cross-section of funds within the same category. This entails including year × Morningstar global category × month fixed effects to isolate the relevant variation, which also controls for time variation by category.

Detecting this fee-flow relationship demands variation within the individual Morningstar global categories in each time period. Therefore, if there are global category-months where all funds have the same fee level, they will not contribute towards the overall results. If there are many global category-months where this is the case, the results could come from only a subset of the data. In this setting, however, the probability of this occurring is low, and when checked, no groups were found to have zero variation in fees.

1.2.3 Data

Data on living and non-surviving long-only open-ended equity mutual funds domiciled in the countries listed in Table 1.B.1 was collected from Morningstar Direct for the period between January 2015 and September 2021.[11] Only funds with a managed portfolio have been used, thus excluding pure and enhanced index funds, ETFs, and fund-of-funds. Beyond cross-sectional information, USD converted time series on gross and net returns, net assets, and fund size for all

[10] $g \in \{\text{One Globe}, \text{Two Globes}, \text{Four Globes}, \text{Five Globes}\}$

[11] Mutual fund studies with an international scope often use countries of domicile to place funds (see, e.g., Cremers and Curtis (2016); Demirci et al. (2020)).

share classes for the entire sample period were gathered, in addition to ratings data. CPI data was downloaded from the World Bank data library to adjust all dollar-denominated variables to the September 2021 dollar level.

Share classes are treated as individual observations by Morningstar. Variables reported at this level, such as total net assets (TNA), are therefore aggregated to the fund level. Where available, the oldest share class is used as the fund's main share class, and if unavailable, the share class with the earliest inception date is used. Benchmark assignment is based on Morningstar's own categorization. Their MPT (Modern Portfolio Theory) indexes are advantageous in that they are based on the funds' investment universes rather than their self-declared strategies. This effectively addresses the problem of the funds strategically choosing their benchmarks (Cremers et al., 2016). Funds with less than two years of monthly observations are excluded in order to make meaningful inference. Funds without data on age, fees, TNA, flows, and Morningstar global category are also excluded. Lastly, funds lacking data on sustainability scores or ratings are left out. All continuous variables are winsorized at the 1st and 99th percentiles to mitigate the potential impact of outliers.

In 2018, Morningstar updated their sustainability rating methodology. Firstly, ratings went from being based on a point-in-time to a trailing 12-month weighted average portfolio score. Secondly, peer group assignment shifted from being based on the Morningstar Category to the Morningstar Global Category system. Lastly, the coverage threshold for a portfolio to receive a sustainability score was raised from 50% to 67% of assets under management (AUM).[12] Consequently, data on globe ratings are unavailable for the period preceding the methodology change, and the sample period is effectively from August 2018 through the time of the data request in September 2021. The final sample is comprised of 16,327 funds managing 8.3 USDtn as of September 2021.

1.2.4 Summary statistics

Panel A in Table 1.1, reports summary statistics for the sample funds. The average fund sees a 0.37% net monthly outflow, charges 1.4% p.a. for active management, is 14 years old with 558 USDm in assets and delivers a positive gross alpha of 0.09% (9bps) per month. The "SD ratio" column indicates how much variation is lost when using year × Morningstar global category × month fixed effects by displaying the ratio between the average standard deviation within

[12] For details see Interpreting the Morningstar Sustainability Rating Changes (2018)

each fixed effects group ("Within SD") and the standard deviation across the entire sample.[13] As is evident, most of the variation is retained. This alleviates any concern that the choice of identifying variation is economically insignificant (Mummolo and Peterson, 2018).

Panel B in Table 1.1 presents summary statistics for the sample of funds split by Morningstar's globe ratings. Funds with the highest ratings also receive the largest inflows, with a monthly average of 20bps. The difference compared to one-globe funds is stark. Moreover, funds with a five-globe rating have 14bps lower annual fees, in addition to being larger in size, older, and having a higher average alpha. Interestingly, there is a monotonic relationship between globe rating and fees, star rating and gross alpha. This echoes Pástor and Vorsatz (2020) who find a similar relationship between globe rating and benchmark-adjusted performance, as well as Döttling and Kim (2021) who find the same for returns and star ratings.

[13]For example, the average variation within each year $\times$ Morningstar global category $\times$ month is 5.5%, which amounts to 93% of the total variation in net flows (5.89%)

Table 1.1.
Summary statistics

This table presents summary statistics of monthly mutual fund and country characteristics. For each variable, the following information is given: number of fund-month (N) observations, observed minimum, median, mean, and max values, as well as the pooled sample standard deviation (SD) and the average within-year × Morningstar global category × month standard deviation, which reflects the variation used in the main regression analyses. Lastly, the ratio between the pooled sample standard deviation and the within standard deviation is given to provide information about the amount of variation lost in the main fixed effects specifications. All calculations are derived from the full set of monthly sample observations.

	Panel A: Main variables							
Variable	N	Min	Median	Mean	Max	SD	Within SD	SD ratio
Net flow (%)	602,013	-21	-0.71	-0.37	32	5.89	5.50	0.93
Fund fee (%)	598,722	0.10	1.30	1.40	3.90	0.69	0.62	0.91
TNA (USDm)	603,515	2.32	103	558	10,185	1,436	928	0.65
Fund return (%)	607,742	-18	1.51	1.09	17	5.99	2.32	0.39
Gross alpha (%)	588,740	-6.80	0.05	0.09	7.30	2.23	2.11	0.94
Fund age (Years)	609,266	0.47	12	14	51	10	9.26	0.91
Family TNA (USDbn)	603,515	0.00	4.52	30	808	100	88	0.89
Family age (Years)	609,266	3.26	32	40	121	26	23	0.88
Industry size (USDtn)	609,266	0.63	32	177	894	286	193	0.68
Ratings								
Globe rating	500,720	1	3	3.05	5	1.11	1.08	0.97
Star rating	522,893	1	3	3.15	5	1.09	1.07	0.98
Δ Globe rating	480,002	-4	0	0.00	4	0.31	0.26	0.83
Δ Star rating	507,750	-4	0	0.00	4	0.39	0.38	0.97
Globe upgrade	480,002	0	0	0.04	1	0.19	0.16	0.84
Globe downgrade	480,002	0	0	0.04	1	0.18	0.15	0.81
Star upgrade	507,750	0	0	0.07	1	0.26	0.25	0.97
Star downgrade	507,750	0	0	0.07	1	0.26	0.25	0.96

	Panel B: Mean values by globe rating				
Variable	One Globe	Two Globes	Three Globes	Four Globes	Five Globes
Net flow (%)	-0.49	-0.59	-0.50	-0.21	0.02
Normalized flow (%)	49	48	49	52	54
Fee (%)	1.45	1.39	1.36	1.34	1.32
Star rating	2.90	3.05	3.16	3.28	3.37
Fund size (USDm)	487	580	616	661	646
Fund age (Years)	13	14	15	15	15
Gross alpha (%)	0.05	0.06	0.07	0.10	0.13

Table 1.2 presents the probabilities of moving to a different globe or star category. In general, for a fund ranked a given number of globes there is a high probability of roughly 93% of staying in that category the next month. Correspondingly, the probability of staying within the same star rating from one month to the next is a little lower at around 86%. When comparing the extreme moves within the respective ratings, there seems to be more short-term mobility in the globe ratings than in the star ratings. For instance, there are 27 (0.05%) funds that go

from one to five globes in a month, while only 6 (0.01%) move from one to five stars. Although the numbers are small, there are five times as many fund-month observations moving from the bottom to the top of the globe ratings as for the star ratings. Likewise, the same relationship persists in the opposite direction, with 16 (0.04%) funds being downgraded from five globes to one globe in a month and only 4 (0.01%) dropping as much on the star rating in the same amount of time.

Table 1.2.
Ratings transition matrix

This table presents the transition matrix for Morningstar's globe (Panel A) and star ratings (Panel B). For each rating, the probability of moving between ratings from one month to the next is shown.

Panel A: Transition probability globe ratings

Current month rating	Next month rating				
	One Globe	Two Globes	Three Globes	Four Globes	Five Globes
One Globe	39,390 (94.07%)	2,319 (2.21%)	436 (0.26%)	138 (0.12%)	27 (0.05%)
Two Globes	2,028 (4.84%)	96,583 (92.05%)	5,829 (3.45%)	536 (0.47%)	128 (0.25%)
Three Globes	371 (0.89%)	5,403 (5.15%)	156,781 (92.76%)	5,902 (5.18%)	419 (0.83%)
Four Globes	68 (0.16%)	562 (0.54%)	5,590 (3.31%)	104,774 (91.99%)	2,514 (5.00%)
Five Globes	16 (0.04%)	58 (0.06%)	379 (0.22%)	2,543 (2.23%)	47,208 (93.86%)

Panel B: Transition probability star ratings

Current month rating	Next month rating				
	One Star	Two Stars	Three Stars	Four Stars	Five Stars
One Star	30,459 (86.9%)	4,179 (4.01%)	112 (0.06%)	24 (0.02%)	6 (0.01%)
Two Stars	4,451 (12.70%)	88,182 (84.57%)	11,102 (6.24%)	170 (0.13%)	36 (0.06%)
Three Stars	122 (0.35%)	11,724 (11.24%)	153,100 (86.02%)	13,132 (10.16%)	204 (0.33%)
Four Stars	16 (0.05%)	157 (0.15%)	13,445 (7.55%)	108,839 (84.17%)	7,118 (11.64%)
Five Stars	4 (0.01%)	23 (0.02%)	215 (0.12%)	7,145 (5.53%)	53,785 (87.96%)

1.3 Are Sustainable Investors Fee Sensitive?

This section tests and reports findings on whether sustainable investors are sensitive to fees. In Section 1.3.1, results pertaining to the first hypothesis are presented and discussed. To address the second hypothesis, Section 1.3.2 then presents the results from tests of funds with different mandates.

1.3.1 Controlling for sustainability

Do investors in funds with a high globe rating respond more to fees than other investors? To address this question, I examine how mutual fund flows vary as a function of fees within the different globe ratings. If changes to the fee level make the fund less desirable, money will flow out and the fund will shrink, and vice versa. This stands in contrast to individual stocks, which have a fixed supply in the short run and therefore do not provide such a direct measure of investor responses.

Table 1.3 is the equivalent of Figure 1.1 and double-sorts funds based on Morningstar globe ratings and how expensive they are. To determine how expensive a fund is, they are sorted within each domicile-month into deciles based on their reported fees. Rank assignment follows the same classification mechanism Morningstar uses when assigning their ratings. This means that the top 10% receive the best rating, the next 22.5% are assigned the next best rating, the middle 35% receive an average rating, the subsequent 22.5% are below average rating, while the bottom 10% gets the lowest rating. As investors who rely on ratings tend to focus on extreme outcomes of discrete measures (Hartzmark, 2015; Feenberg et al., 2017; Hartzmark and Sussman, 2019), the annualized average monthly net flow is jointly tabulated for the funds with the two highest (least expensive) and lowest (most expensive) fee ratings.

The columns in Table 1.3 correspond to the globe ratings and the rows to the two fee classifications. The value of −4.15 in the top-left corner indicate that, on average, funds with the lowest sustainability ratings and the highest fees had an annualized monthly net flow of −4.15%. At the other end of the scale, the funds with the highest sustainability ratings and the lowest fees had an annualized monthly net flow of 10.35%. The bottom row tabulates the difference in average net flows for funds with the same globe but different fee ratings. In the first column, the statistically significant difference of 6.57 implies that the difference in flows to the funds with low and high fees is 6.57 percentage points (pp), meaning that there is significant variation in flows within the funds with the lowest sustainability rating when sorting on fees.

The ΔGlobes column shows the difference between funds with the highest and lowest globe rating. The statistically insignificant value of -0.21 in the top row indicates that funds with a five-globe rating had a 0.21 percentage point lower average net flow than the lowest rated funds. This implies that capital flows to and from funds with high fees are largely invariant to globe ratings. In sum, Table 1.3 shows how investors reallocate capital out of the most expensive and into the least expensive funds. For funds with one, two, or three globes, the withdrawal of capital from the most expensive funds is larger than the inflow of capital to the less expensive funds.

The difference in net flows between the low and high fee funds is largest for the funds with the highest sustainability rating. The value in the bottom right corner, 8.14, is the difference-in-differences estimate between the five-globe and one-globe funds. This number is statistically significant and signals that investors in the most sustainable funds are more responsive to fees than investors in one-globe funds.

Table 1.3.
Fund flows sorted by globe ratings and fees

This table shows how flows to and from actively managed mutual funds vary when sorting on globe ratings and fund fees (one globe is assigned to the funds in the bottom 10%, and five globes to the funds in the top 10% of eligible funds). In each domicile-month (time t), funds that have been assigned a Morningstar globe rating are sorted into deciles based on fees independently of their globe rating. After removing year-by-month fixed effects, annualized average monthly net flows in the following month ($t+1$) are tabulated for each globe rating. The "High" row corresponds to the funds ranked most expensive (ranks 1 and 2), while the "Low" row corresponds to the funds ranked least expensive (ranks 4 and 5). Inference on differences between groups of funds is based on Newey and West (1987) standard errors. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

Rating	One Globe	Two Globes	Three Globes	Four Globes	Five Globes	ΔGlobes
High fee	-4.15	-6.89	-5.24	-3.66	-4.36	-0.21
Low fee	2.42	0.49	1.58	5.64	10.35	7.93***
Low − High	6.57***	7.38***	6.82***	9.30***	14.71***	8.14***

Table 1.4 presents the results from panel regressions following the setup from Equation 1.2. Net flow is the dependent variable in all columns. In the first two columns, the variable "High sustainability" is a dummy variable equal to 0 if the globe rating is below 4 and equal to 1 if the globe rating is 4 or 5. In columns 3 and 4, a dummy variable coding for whether a globe rating of 1, 2, 4 or 5 has been given is used. The three-globe funds are the omitted category, and results for the two-globe and four-globe funds are not reported for brevity. Overall, flow responds negatively to fees. Columns 1 and 2 show that funds with a high sustainability rating

(four or five globes) on average receive significantly more money than lower rated funds. The interaction term between the high sustainability dummy and fees in column 1 also shows that flows are more sensitive to changes in fees than for low sustainability funds. This result also stands when additional control variables are included.

The five-globes coefficients in columns 3 and 4 confirm previous findings in the literature that funds with a high sustainability rating receive a larger inflow of funds than lower rated funds (Hartzmark and Sussman, 2019; Ammann et al., 2019). Moreover, the interaction term between fees and the highest sustainability rating enters negatively and is significant for all specifications, indicating that flows in the most sustainable funds are significantly more sensitive to fees than average rated funds. Nor does this change when additional control variables are added. For this group of funds, the estimated flow-fees slope coefficient is (from column 3) -0.70 ($= -0.233 - 0.467$). The average within year×MSGC×month standard deviation for fees is 0.62. Thus, a one standard deviation increase in fees is associated with a decrease in net flows of -0.43% ($= -0.70 \cdot 0.62$). This number is substantial in that it exceeds the overall average net flow of -0.37%. In other words, conditioning flows on fees implies fluctuations in the conditional expected net flow that are of first-order economic importance. Even with the inclusion of control variables in column 4, the interaction coefficient is economically significant in that a one standard deviation increase in fees is associated with a decrease in net flows of -0.303% representing a shift in expected net flows of 82% ($= -0.303/-0.37$).

Columns 3 and 4 also test whether the difference between the interaction terms for the one-globe and five-globe funds are equal. In column 3, the difference in coefficient size is -0.521 and the p-value from testing whether the difference in coefficients is zero is lower than 1%. The difference in coefficients is statistically significant at the 5% level when control variables are added in column 4. The reduced difference between the interaction terms when moving from column 3 to 4 is largely due to the decrease in the interaction term for the one-globe funds. Largely, these results also hold when using normalized flow as the dependent variable instead; see Table 1.C.1 in Appendix 1.C for details. These findings are also robust to alternative fixed effects specifications; see Table 1.C.2 and 1.C.3 in Appendix 1.C for details.

This section documents how investors differ in their response to fees across different globe ratings. Consistent with the literature, there is a clear tendency for flows to leave funds with a low globe rating and to enter funds with a high rating. However, when also sorting fees within globe ratings, we see a more nuanced picture where investors clearly respond to fees independently of globe ratings. This is also significantly more pronounced for the highest-rated

Table 1.4.
Sustainability and flow - fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how net flows vary with sustainability measures and fees. In columns 1 and 2, the measure of sustainability is a single dummy variable equal to 0 if the globe rating is below 4 and equal to 1 if the globe rating is 4 or 5. Columns 3 and 4 use dummy variables coding for whether a globe rating of 1, 2, 4, or 5 has been given to a fund. Even-numbered columns also include additional controls. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow (t)			
	(1)	(2)	(3)	(4)
Fees	−0.233***	−0.215***	−0.233***	−0.179***
	(0.052)	(0.048)	(0.059)	(0.056)
High sustainability	0.645***	0.434***		
	(0.096)	(0.071)		
High sustainability x Fees	−0.197***	−0.101**		
	(0.054)	(0.048)		
One Globe			−0.020	0.137
			(0.134)	(0.117)
Five Globes			1.16***	0.775***
			(0.149)	(0.116)
One Globe x Fees			0.054	−0.090
			(0.077)	(0.070)
Five Globes x Fees			−0.467***	−0.309***
			(0.087)	(0.074)
(Five − One) Globe x Fees			−0.521	−0.219
p-value			< 0.01	0.014
MSGC × YM FE	×	×	×	×
Other controls		×		×
R^2	0.029	0.060	0.029	0.060
Observations	490,468	461,679	490,468	461,679

funds. Taken together, it is clear that fees matter in general, but they matter more to investors who opt for funds with a high sustainability rating.

1.3.2 Controlling for fund type

The results in the previous section suggest that investors in funds with a high sustainability rating are more sensitive to fees than investors in funds with a low sustainability rating. This contrast is strongest when comparing the extreme globe ratings of one and five globes. However, the analyses in Section 1.3.1 do not distinguish between different types of sustainable funds. For example, some funds have an explicit sustainability mandate and others are focused on having portfolios aligned with individual ESG pillars. The heterogeneity in types of sustainable funds allows investors to allocate funds in accordance with their environmental and social preferences. Those with the strongest preferences can opt for the funds that are explicitly dedicated to sustainable investing while those with a particular desire to have an environmentally friendly portfolio can choose a fund that invests to limit our carbon footprint. Riedl and Smeets (2017) find that those who hold funds with an explicit sustainability objective are willing to forego financial returns and pay higher fees. Therefore, there is potentially a heterogeneity among sustainable investors, which in turn could entail different demand sensitivities with respect to fees.

To investigate this, the analyses from Section 1.3.1 are applied to different subsets of funds: regular, green, and low carbon. The first group is comprised of the funds that have been assigned a Morningstar globe rating, but that do not have an explicit sustainability objective. This means that they are sustainable by choice not by mandate. The green group consists of funds with an explicit sustainability objective, flagged by Morningstar, or if the fund name contains a keyword that is typically associated with sustainability.[14] Lastly, the low-carbon group consists of funds that have been assigned a low-carbon label by Morningstar, which is a Morningstar label that has been shown to impact flows (Ceccarelli et al., 2021).[15]

Table 1.5 reports the results from panel regressions similar to those shown in columns 1 and 2 in Table 1.4, but divided across different fund types. Net flow is the dependent variable in all specifications and high sustainability is a dummy variable taking the value of one if the globe rating is four or five and zero otherwise. To start, the coefficient on fees is negative and statistically significant with and without additional control variables in columns 5 and 6 only. Across specifications, high sustainability funds receive significantly more capital than low sustainability funds. Looking at the interaction term between the high sustainability dummy and fees, there is no consistent evidence of any difference in fee sensitivity between high and low sustainability funds for the green and low-carbon funds. For regular funds, the interaction

[14] Fund names with the following keywords are labeled as green: sustainable and its derivations, SRI, ethical, green, climate, positive change, environment, and its derivations, ESG, circular, impact, carbon, transition, clean, social and its derivations, responsible, governance, engagement, and renewable.

[15] Low-carbon mutual funds allow investors to purchase lower exposure to climate change risk at the cost of lower sectorial diversification.

term enters significantly in both specifications.

Table 1.5.
Fund type and flow - fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how net flows vary with sustainability measures and fees for different subsamples of funds, i.e., funds with and without an explicit sustainability objective. Columns 1 and 2 consist of the subset of funds labeled as green by Morningstar or that indicate sustainability in the fund name. Columns 3 and 4 analyze the subset of funds given the low-carbon label by Morningstar. Lastly, columns 5 and 6 consist of the funds with a globe rating, but without an explicit sustainability objective. All columns use a single dummy variable equal to 0 if the globe rating is below 4 and equal to 1 if the globe rating is 4 or 5. Even-numbered columns also include additional controls. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow (t)					
Fund type:	Green		Low carbon		Regular	
	(1)	(2)	(3)	(4)	(5)	(6)
Fees	-0.421^{***}	-0.107	0.015	-0.065	-0.260^{***}	-0.266^{***}
	(0.113)	(0.093)	(0.097)	(0.088)	(0.049)	(0.049)
High sustainability	0.644^{***}	0.735^{***}	0.360^{**}	0.324^{**}	0.429^{***}	0.352^{***}
	(0.226)	(0.205)	(0.158)	(0.136)	(0.107)	(0.087)
High sustainability x Fees	-0.185	-0.255^{*}	-0.201^{*}	-0.120	-0.183^{**}	-0.133^{**}
	(0.135)	(0.127)	(0.106)	(0.098)	(0.071)	(0.061)
MSGC × YM FE	×	×	×	×	×	×
Other controls		×		×		×
R^2	0.054	0.094	0.036	0.075	0.034	0.057
Observations	73,309	65,835	107,593	103,850	309,566	291,994

In Table 1.6, the analysis is the same as in columns 3 and 4 in Table 1.4 but split across different fund types. Net flow is the dependent variable in all specifications and each globe rating enters with a distinct dummy variable. By first focusing on funds with either an explicit sustainability orientation (columns 1-2) or a clear environmental profile (3-4), there is evidence of flows responding more to fees within the funds with the highest sustainability rating. However, the difference in coefficients between the interaction terms for the highest and lowest rated funds within these two subsets is not statistically significant at any of the conventional levels. Overall, there is evidence that funds with the highest sustainability ratings receive larger flows than middle-rated funds but little evidence of large and significant differences in fee sensitivity when comparing the extreme rankings.

Table 1.6.
Fund type, globe ratings and flow - fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how net flows vary with globe ratings and fees for different subsamples of funds, i.e., funds with and without an explicit sustainability objective. Columns 1 and 2 consist of the subset of funds labeled as green by Morningstar or that indicate sustainability in the fund name. Columns 3 and 4 analyze the subset of funds given the low-carbon label by Morningstar. Lastly, columns 5 and 6 consist of the funds with a globe rating, but without an explicit sustainability objective. All columns use dummy variables coding for whether a globe rating of 1, 2, 4, or 5 has been given to a fund. Even-numbered columns also include additional controls. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow (t)					
Fund type:	Green		Low carbon		Regular	
	(1)	(2)	(3)	(4)	(5)	(6)
Fees	−0.330**	−0.089	0.012	−0.049	−0.289***	−0.243***
	(0.132)	(0.113)	(0.100)	(0.093)	(0.058)	(0.061)
One Globe	0.839*	0.243	0.467	0.078	−0.123	0.114
	(0.473)	(0.441)	(0.332)	(0.286)	(0.149)	(0.137)
Five Globes	1.20***	1.07***	0.913***	0.659***	0.751***	0.541***
	(0.320)	(0.276)	(0.224)	(0.200)	(0.188)	(0.164)
One Globe x Fees	−0.217	0.008	−0.143	−0.223	0.130	−0.050
	(0.276)	(0.246)	(0.211)	(0.181)	(0.091)	(0.084)
Five Globes x Fees	−0.434**	−0.384**	−0.486***	−0.310**	−0.423***	−0.346***
	(0.196)	(0.174)	(0.142)	(0.134)	(0.118)	(0.101)
(Five − One) Globe x Fees	−0.218	−0.392	−0.343	−0.087	−0.552	−0.296
p-value	0.473	0.162	0.156	0.679	< 0.01	0.014
MSGC × YM FE	×	×	×	×	×	×
Other controls		×		×		×
R^2	0.055	0.095	0.036	0.075	0.034	0.058
Observations	73,309	65,835	107,593	103,850	309,566	291,994

Columns 5 and 6 focus on the subset of funds with a sustainability rating, but without any explicit sustainability objective. Here, five-globe funds receive on average significantly more flows than average rated funds, but the coefficient size is smaller than those in columns 1 and 2 pertaining to the explicitly green funds. The interaction term between five globes and fees is statistically significant at the 1% level, also when adding additional control variables. Furthermore, the coefficient does not change much, indicating that little of the variation in fee sensitivity is soaked up by including these control variables for the highest rated funds. The estimate in column 5 of -0.423 is close in size to the same specification in column 4, Table 1.4. This estimate implies that for the highest rated regular funds, the estimated flow-fees slope coefficient is -0.71 $(= -0.289 - 0.423)$. This indicates that a one standard deviation increase in fees is associated with a decrease in net flows of -0.44% $(= -0.71 \cdot 0.62)$. This is only marginally different from the conditional estimate when using the entire fund sample in Table 1.4. Furthermore, the difference in interaction term coefficients between the one-globe and five-globe funds is significant at the 1% and 5% levels in columns 5 and 6, respectively. These results also hold when using normalized flow as the dependent variable; see Table 1.C.5 in Appendix 1.C for details.

In sum, distinguishing between subsets of funds based on different types of sustainability reveals that the largest differences in fee sensitivity primarily lie within the group of funds with a high sustainability rating but without an explicit sustainability objective.

1.4 Further analysis

This section presents the results of additional analyses performed to provide grater understanding of the results from the previous section. Firstly, I examine how flows vary depending on sustainability rating when using the unit cost of sustainability as a measure of fee instead of total fees. Secondly, I examine whether the estimates and findings in Table 1.4 persist when the choice of rating changes.

1.4.1 Controlling for unit cost

The nature of the globe ratings is such that they are distributed within each Morningstar global category. Consequently, if you wanted to invest in the most sustainable fund, you would look at the underlying portfolio sustainability score and not the globe rating itself. To investigate whether investors in funds with a similar globe rating incorporate this information, I capture sustainability and fee in one variable by calculating the unit cost of sustainability (ESG fee). Here, it is defined as the ratio between fees and the fund's sustainability score.[16] To account for the changes made to Morningstar's portfolio sustainability scores in 2019, the variable is rescaled to make the levels before and after 2019 comparable. Lastly, the values are negated (i.e., the denominator is $100 - \text{Sustainability score}_{i,t}$) such that a higher value indicates a more

[16]Similarly, the Sharpe ratio summarizes risk and return.

sustainable portfolio.

$$\text{ESG fee}_{i,t} = \frac{\text{Fee}_{i,t}}{100 - \text{Sustainability score}_{i,t}} \tag{1.3}$$

Table 1.7 follows the structure of Table 1.3, but the funds are now divided into groups based on whether their unit cost of sustainability is high or not. Effectively, the "High ESG fee" group consists of the funds that have been assigned the two lowest fee rating levels based on ESG fee instead of total fees, and vice versa for the "Low ESG fee" group. Again, the columns represent the different globe ratings. Compared to Table 1.3, the intra-globe differences are slightly higher, and the difference-in-differences estimate somewhat lower. In general, the results are in line with those in Table 1.3.

Table 1.7.
Fund flows sorted by globe ratings and ESG fees

This table presents how flows to and from actively managed mutual funds vary when sorting on globe ratings and ESG fees (one globe is assigned to the funds in the bottom 10%, and five globes to the funds in the top 10% of eligible funds). In each domicile-month (time t), funds that have been assigned a Morningstar globe rating are sorted into deciles based on ESG fees independently of their globe rating. After removing year-by-month fixed effects, annualized average monthly net flows in the following month ($t+1$) are tabulated for each globe rating. The "High" row corresponds to the funds ranked most expensive (ranks 1 and 2), while the "Low" row corresponds to the funds ranked least expensive (ranks 4 and 5). Inference on differences between groups of funds is based on Newey and West (1987) standard errors. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

Rating	One Globe	Two Globes	Three Globes	Four Globes	Five Globes	ΔGlobes
High ESG fee	−4.81	−7.05	−5.28	−3.68	−5.09	−0.28
Low ESG fee	4.95	2.81	2.53	6.39	10.12	5.17***
Low − High	9.76***	9.86***	7.81***	10.07***	15.21***	5.45***

Table 1.8 presents the results from panel regressions following the setup from Equation 1.2. Net flow is the dependent variable in all columns. The differing characteristic between Table 1.4 and Table 1.8 is that in the latter, the unit cost of sustainability is used instead of the fund fees. Overall, flow responds negatively to ESG fee. The interaction term between the high sustainability dummy and ESG fee in column 1 shows that flows are more sensitive to changes in unit cost than is the case for low sustainability funds. This result stands when additional control variables are included, but the effect is halved.

In columns 3 and 4, the interaction term between ESG fee and the highest sustainability rating enters negatively and significantly for all specifications, indicating that flows in the most sustainable funds are significantly more sensitive to unit cost than average-rated funds. Furthermore, this does not change when adding controls even though the coefficient size is reduced. Columns 3 and 4 also tests whether the difference between the interaction terms for the one-globe and five-globe funds are equal. In column 3, the difference in coefficient size is -0.366 which is significant at the 1% level. The difference in coefficients is statistically significant at the 5% level when control variables are added in column 4. The reduced difference between the interaction terms when moving from column 3 to 4 is largely due to the decrease in the interaction term for the one-globe funds.

Table 1.8.
Sustainability and flow-ESG fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how net flows vary with sustainability measures and ESG fees. In columns 1 and 2, the measure of sustainability is a single dummy variable equal to 0 if the globe rating is below 4 and equal to 1 if the globe rating is 4 or 5. Columns 3 and 4 use dummy variables coding for whether a globe rating of 1, 2, 4, or 5 has been given to a fund. Even-numbered columns also include additional controls. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow (t)			
	(1)	(2)	(3)	(4)
ESG Fee	−0.178***	−0.168***	−0.171***	−0.135***
	(0.038)	(0.034)	(0.043)	(0.041)
High sustainability	0.636***	0.418***		
	(0.093)	(0.069)		
High sustainability x ESG Fee	−0.152***	−0.074**		
	(0.040)	(0.036)		
One Globe			0.053	0.185
			(0.133)	(0.114)
Five Globes			1.14***	0.758***
			(0.144)	(0.111)
One Globe x ESG Fee			0.010	−0.082
			(0.056)	(0.050)
Five Globes x ESG Fee			−0.356***	−0.233***
			(0.065)	(0.054)
(Five − One) Globe x Fees			−0.366	−0.152
p-value			< 0.01	0.020
MSGC × YM FE	×	×	×	×
Other controls		×		×
R^2	0.029	0.060	0.029	0.060
Observations	490,182	461,479	490,182	461,479

1.4.2 Controlling for rating

Section 1.3.1 showed that investors in funds with a higher sustainability rating are more sensitive to fees than investors in lower-rated funds. As such, a natural question to consider is whether similar results would appear regardless of rating or whether the results in Section 1.3.1 are specific to the globe ratings. Consequently, in this section I analyze to what extent investors in high rated funds are responsive to fees when using Morningstar's star ratings. As a determinant of fund flows, the star ratings have been proven to be an important factor, explaining flows above and beyond historical performance (Evans and Sun, 2021; Ben-David et al., 2022; Del Guercio and Tkac, 2008). As a robustness check, I also rerun the main analyses adding the star rating as a control variable. For details on the star rating, see Appendix 1.E.

Following the same layout as Table 1.3, Table 1.9 double-sorts net flows based on star rating (in columns) and fees (in rows). Again, the bottom row tests the difference between net flows in the most and least expensive funds within the same star rating, and the last column tests the difference between net flows for funds across star ratings within the same fee group. Figure 1.E.1 provides a visual presentation of Table 1.9. What is immediately obvious from looking at Table 1.9, and consistent with the literature is that there are significant differences in net flows between the lowest and highest rated funds regardless of fee level, and this difference is orders of magnitude larger than the difference between the highest and lowest globe ratings.[17]

Looking at the bottom row, investors within all star rating levels are sensitive to fees, except for the highest rated funds. The difference in net flows for the five-star funds is not statistically significant. Compared to funds with lower star ratings, investors opting into the five-star funds do not appear to select funds based on fees. Furthermore, it is only the five-star funds with high fees that see an average net inflow of capital. The difference-in-differences estimate of -2.27 is statistically insignificant, but the sign indicates that in stark opposition to the corresponding estimate in Table 1.3, investors in one-star funds are more fee-sensitive than those in five-star funds. Taken together, the results in Table 1.9 suggest that what is observed when double sorting on globe and fee ratings in Section 1.3.1 is not due to an underlying mechanism that is present independently of the rating scheme.

To further investigate this, Table 1.10 shows the results from panel regressions following the setup from Equation 1.2. Columns 1 and 2 add the star rating as a control variable to the regression specifications from columns 2 and 4 in Table 1.4, and columns 3 and 4 use the star rating as the interacting variable with fees instead of the globe rating. Net flow is the dependent variable in all columns.

In columns 1 and 2, the star rating is added as an explicit control, and it enters with a positive sign and is statistically significant at the 1% level in both specifications. In contrast to column 2 in Table 1.4, when adding the star rating as a control, the interaction term between the high sustainability dummy and fees becomes statistically insignificant, meaning that the star rating absorbs a great deal of the variation. Focusing on each globe rating individually, adding the

[17]See, for example Evans and Sun (2021).

Table 1.9.
Fund flows sorted by star ratings and fees

This table presents how flows to and from actively managed mutual funds vary when sorting on star ratings and fund fees (one star is assigned to the funds in the bottom 10%, and five stars to the funds in the top 10% of eligible funds). In each domicile-month (time t), funds that have been assigned a Morningstar star rating are sorted into deciles based on fees independently of their star rating. After removing year-by-month fixed effects, annualized average monthly net flows in the following month ($t + 1$) are tabulated for each star rating. The "High fee" row corresponds to the funds ranked most expensive (ranks 1 and 2), while the "Low fee" row corresponds to the funds ranked least expensive (ranks 4 and 5). Inference on differences between groups of funds is based on Newey and West (1987) standard errors. Asterisks denote statistical significance: $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

Rating	One Star	Two Stars	Three Stars	Four Stars	Five Stars	ΔStars
High fee	-13.85	-11.7	-9.07	-2.58	14.67	28.52^{***}
Low fee	-10.37	-9.45	-5.09	1.38	15.88	26.25^{***}
Low - High fee	3.48^{**}	2.25^{**}	3.98^{***}	3.96^{***}	1.21	-2.27

star rating as a control in column 2 does not change the results from column 4 in Table 1.4 by any mentionable degree. The interaction term between the five-globe rating and fees remains negative and statistically significant at the 1% level. The coefficient size also remains largely the same. The most evident change is the interaction term between the one-globe rating and fees, which now becomes larger in absolute value and statistically significant at the 10% level. This further reduces the difference between the interaction terms for the one-globe and five-globe ratings such that it is only significant at the 10% level. Thus, controlling for past performance on the entire sample of funds somewhat weakens the relationships described in Section 1.3.1.

In columns 3 and 4, the star rating is interacted with fees. As is evident, the funds with five stars receive significantly larger inflows than average-rated funds. Likewise, one-star funds receive on average significantly less capital. Without any additional control variables, neither of the interaction terms are statistically significant, which indicates that investors across star rating levels do not respond differently to fees. Furthermore, the interaction terms for the one-star and five-star ratings are not significantly different from one another. Lastly, by adding additional control variables in column 4, the interaction term between one star and fees becomes significant at the 5% level, but the coefficient has a positive sign, meaning they are significantly less sensitive to fees than investors in average-rated funds. The resulting fee-sensitivity estimate for the one-star funds is therefore 0.019 ($= -0.142 + 0.161$). This entails that a one standard deviation increase in fees is associated with an increase in net flow of 0.01% ($= 0.019 \cdot 0.62$), which is of little economic significance. The difference in interaction terms, however, is statistically

Table 1.10.
Star ratings and flow - fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how net flows vary with star ratings and fees. Columns 1 and 2 add star ratings as an explicit control variable to the regressions run in Table 1.4, columns 2 and 4. Columns 3 and 4 use dummy variables coding for whether a star rating of 1, 2, 4, or 5 has been given to a fund. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow (t)			
	(1)	(2)	(3)	(4)
Fee	-0.095^{*}	-0.040	0.016	-0.142^{**}
	(0.049)	(0.057)	(0.048)	(0.054)
High sustainability	0.313^{***}			
	(0.068)			
High sustainability x Fees	-0.071			
	(0.047)			
Star rating	0.342^{***}	0.343^{***}		
	(0.022)	(0.022)		
One Globe		0.261^{**}		
		(0.116)		
Five Globes		0.688^{***}		
		(0.118)		
One Globe x Fees		-0.139^{*}		
		(0.068)		
Five Globes x Fees		-0.306^{***}		
		(0.077)		
One star			-0.686^{***}	-0.364^{**}
			(0.135)	(0.135)
Five stars			2.00^{***}	1.48^{***}
			(0.140)	(0.124)
One star x Fee			0.112	0.161^{**}
			(0.075)	(0.074)
Five stars x Fee			-0.014	-0.116
			(0.088)	(0.083)
Diff coefficient		-0.168	-0.125	-0.277
p-value		0.062	0.260	< 0.01
MSGC × YM FE	×	×	×	×
Other controls	×	×		×
R^2	0.061	0.061	0.046	0.062
Observations	428,149	428,149	430,514	428,149

significant, indicating that investors at the extreme ends of the ratings respond differently from one another when it comes to fees.

Lastly, in Table 1.11, the star rating is added as a control to the panel regressions from Tables 1.5 and 1.6, where the odd-numbered columns correspond to the odd-numbered columns in Table 1.5 and the even-numbered columns to the even-numbered columns in Table 1.6. Across the subsamples of funds, the star rating enters with a positive sign and is statistically significant, while the interaction between the high sustainability dummy and fees becomes statistically insignificant. Focusing on the individual globe ratings, adding star ratings as a control slightly reduces the coefficient size in absolute terms for the interaction between fees and the five-globe rating for the green and low-carbon funds, while making the negative relationship between fee sensitivity and subsequent flows stronger for the regular funds.Looking at the difference between the interaction terms in the even-numbered columns, the difference is only statistically significant for the subsample of regular funds. In sum, adding star rating as an explicit control does not materially change the conclusions from Section 1.3.2

Section 1.3.1 and 1.3.2 concluded that, overall, investors in the funds with the highest sustainability ratings are more fee sensitive than those in lower rated funds, and especially in relation to the funds with the highest sustainability rating but without an explicit sustainability objective. This section has investigated whether similar results emerge when the globe rating is replaced by the star rating, and has demonstrated that they do not. Furthermore, the main results also hold for the most part when conditioning on star ratings.

Table 1.11.
Fund type and flow - fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how net flows vary with sustainability measures and fees for different subsamples of funds, i.e., funds with and without an explicit sustainability objective when star rating is added as a control variable. Columns 1 and 2 consist of the subset of funds labeled as green by Morningstar or that indicate sustainability in the fund name. Columns 3 and 4 analyzes the subset of funds given the low carbon label by Morningstar. Lastly, columns 5 and 6 consist of the funds with a globe rating, but without an explicit sustainability objective. Odd-numbered columns use a single dummy variable that equals 0 if the globe rating is below 4 and 1 if the globe rating is 4 or 5. Even-numbered columns use dummy variables coding for whether a globe rating of 1, 2, 4, or 5 has been given to a fund. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow (t)					
Fund type:	Green		Low carbon		Regular	
	(1)	(2)	(3)	(4)	(5)	(6)
Fees	−0.020	0.005	0.118	0.146	−0.161***	−0.115*
	(0.098)	(0.126)	(0.084)	(0.091)	(0.052)	(0.063)
High sustainability	0.600***		0.232*		0.243***	
	(0.204)		(0.133)		(0.086)	
High sustainability x Fees	−0.193		−0.122		−0.096	
	(0.126)		(0.097)		(0.060)	
Star rating	0.307***	0.307***	0.433***	0.433***	0.313***	0.315***
	(0.050)	(0.050)	(0.031)	(0.031)	(0.027)	(0.028)
One Globe		0.267		0.274		0.231*
		(0.451)		(0.289)		(0.134)
Five Globes		0.919***		0.488**		0.566***
		(0.267)		(0.199)		(0.168)
One Globe x Fees		−0.018		−0.279		−0.103
		(0.252)		(0.185)		(0.082)
Five Globes x Fees		−0.336*		−0.291**		−0.380***
		(0.173)		(0.136)		(0.104)
(Five − One) Globe x Fees		−0.318		−0.011		−0.277
p-value		0.261		0.958		0.019
MSGC × YM FE	×	×	×	×	×	×
Other controls	×	×	×	×	×	×
R^2	0.090	0.090	0.077	0.077	0.060	0.060
Observations	60,263	60,263	97,302	97,302	270,584	270,584

1.5 Conclusion

This paper analyzes demand sensitivity for sustainability among mutual fund investors by focusing on how they respond to fees. This is, to the best of my knowledge, the first paper that focuses on how sensitive the demand for sustainability is with respect to fees.

I find that investors in mutual funds with a five-globe rating are significantly more sensitive to fees than both average rated and one-globe funds. By addressing the fact that not all sustainable funds are equal in that some have explicit sustainability objectives and others not, I find that it is largely investors in high rated funds without an explicit sustainability objective that elicit the strongest demand response to changes in fees, while investors in explicit sustainable funds are less sensitive to fees. These results have implications for how we understand the demand for sustainable investing. The particulars of why one group of investors exhibits a more pronounced response than other groups should be the topic of future research.

References

Ammann, M., Bauer, C., Fischer, S., and Müller, P. (2019). The impact of the morningstar sustainability rating on mutual fund flows. *European Financial Management*, 25(3):520–553.

Ben-David, I., Li, J., Rossi, A., and Song, Y. (2022). What do mutual fund investors really care about? *The Review of Financial Studies*, 35(4):1723–1774.

Ceccarelli, M., Ramelli, S., and Wagner, A. F. (2021). Low-carbon mutual funds. *Swiss Finance Institute Research Paper*, 0(19-13).

Chevalier, J. and Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of political economy*, 105(6):1167–1200.

Cremers, K. and Curtis, Q. (2016). Do mutual fund investors get what they pay for: Securities law and closet index funds. *Va. L. & Bus. Rev.*, 11:31.

Cremers, M., Ferreira, M. A., Matos, P., and Starks, L. (2016). Indexing and active fund management: International evidence. *Journal of Financial Economics*, 120(3):539–560.

Del Guercio, D. and Tkac, P. A. (2008). Star power: The effect of monrningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis*, 43(4):907–936.

Demirci, I., Ferreira, M. A., Matos, P., and Sialm, C. (2020). How global is your mutual fund? international diversification from multinationals. Technical report, National Bureau of Economic Research.

Döttling, R. and Kim, S. (2021). Sustainability preferences under stress: Evidence from mutual fund flows during covid-19. *Available at SSRN 3656756*.

Evans, R. B. and Sun, Y. (2021). Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. *The Review of Financial Studies*, 34(1):67–107.

Feenberg, D., Ganguli, I., Gaule, P., and Gruber, J. (2017). It's good to be first: Order bias in reading and citing nber working papers. *Review of Economics and Statistics*, 99(1):32–39.

Goldstein, I., Kopytov, A., Shen, L., and Xiang, H. (2022). On esg investing: Heterogeneous preferences, information, and asset prices. Technical report, National Bureau of Economic Research.

Hartzmark, S. M. (2015). The worst, the best, ignoring all the rest: The rank effect and trading behavior. *The Review of Financial Studies*, 28(4):1024–1059.

Hartzmark, S. M. and Sussman, A. B. (2019). Do investors value sustainability? a natural experiment examining ranking and fund flows. *The Journal of Finance*, 74(6):2789–2837.

Heinkel, R., Kraus, A., and Zechner, J. (2001). The effect of green investment on corporate behavior. *Journal of financial and quantitative analysis*, 36(4):431–449.

Humphrey, J., Kogan, S., Sagi, J., and Starks, L. (2021). The asymmetry in responsible investing preferences. Technical report, National Bureau of Economic Research.

Mummolo, J. and Peterson, E. (2018). Improving the interpretation of fixed effects regression results. *Political Science Research and Methods*, 6(4):829–835.

Oehmke, M. and Opp, M. M. (2020). A theory of socially responsible investment. *Swedish House of Finance Research Paper*, 1(20-2).

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2):550–571.

Pástor, L. and Vorsatz, M. B. (2020). Mutual fund performance and flows during the covid-19 crisis. *The Review of Asset Pricing Studies*, 10(4):791–833.

Pedersen, L. H., Fitzgibbons, S., and Pomorski, L. (2021). Responsible investing: The esg-efficient frontier. *Journal of Financial Economics*, 142(2):572–597.

Riedl, A. and Smeets, P. (2017). Why do investors hold socially responsible mutual funds? *The Journal of Finance*, 72(6):2505–2550.

Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *The journal of finance*, 53(5):1589–1622.

1.A Variables

Table 1.A.1.
Variable definitions

Variable	Definition
Main variables	
Net flow	Percentage growth in TNA (in USD), net of internal growth (assuming reinvestment of dividends and distributions). See Section 1.2, Equation (1.1) for details.
Normalized flow	Based on net flows, the funds are split into deciles based on size and each fund is assigned percentiles based on flows within each size decile.
Expenses	Annualized monthly expense ratio.
Portfolio sustainability score	The Morningstar Portfolio Sustainability Score is an asset-weighted average of Sustainalytics' company level ESG risk score. This measures the degree to which a company's economic value may be at risk driven by ESG factors.
Other controls	
Morningstar rating	A purely quantitative, backward-looking measure of a fund's past performance, measured from one to five stars. Star ratings are calculated at the end of every month. See Morningstar Ratings for more information.
Fund size	CPI adjusted to equal September 2021 USD TNA (USDm) in period t. Used in log form in panel regressions.
Rolling returns	Rolling 12/24-month prior returns. Calculated from t = -13 or -25 to t = -1.
Family size	Log of 1+ the sum of the TNA of all funds in the same management company as fund j in period t excluding fund j itself. Based on the CPI adjusted fund size variable in million USD.
Family age	Natural logarithm of the number of years since the inception of the fund family's first fund.
Fund industry size	The natural logarithm of the sum of total net assets in the industry of open-end equity mutual funds (in million USD), CPI-adjusted to the dollar-level in September 2021.

1.B Breakdown of funds by country

Table 1.B.1.
Distribution by country

This table presents the number of funds, the number of fund-month observations and the percentage share of fund-month observations per country for the sample of actively managed equity mutual funds.

Country	Funds	N fund-months	Percent share
Australia	670	25,089	4.1%
Austria	158	5,908	1.0%
Belgium	130	4,814	0.8%
Brazil	374	13,984	2.3%
Canada	1,277	47,633	7.8%
Chile	113	4,181	0.7%
China	296	10,942	1.8%
Denmark	234	8,595	1.4%
Finland	157	5,859	1.0%
France	883	33,028	5.4%
Germany	390	14,405	2.4%
Hong Kong	98	3,653	0.6%
India	9	334	0.1%
Indonesia	2	76	0.0%
Ireland	557	20,680	3.4%
Italy	103	3,875	0.6%
Japan	1,074	40,234	6.6%
Liechtenstein	82	3,085	0.5%
Luxembourg	2,114	77,849	13%
Malaysia	262	9,850	1.6%
Mexico	81	3,073	0.5%
Netherlands	186	6,891	1.1%
New Zealand	73	2,728	0.4%
Norway	119	4,405	0.7%
Portugal	33	1,218	0.2%
Singapore	95	3,560	0.6%
South Africa	306	11,401	1.9%
South Korea	564	21,253	3.5%
Spain	275	10,227	1.7%
Sweden	259	9,654	1.6%
Switzerland	288	10,758	1.8%
Taiwan	320	12,003	2.0%
Thailand	400	14,915	2.4%
United Kingdom	1,188	44,142	7.2%
United States	3,157	118,964	20.0%
Total	**16,327**	**609,266**	**100%**

1.C Main results

This section contains additional analyses and robustness checks for the main results presented in Section 1.3.

Table 1.C.1.
Sustainability and normalized flow - fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how normalized flows vary with sustainability measures and fees. In columns 1 and 2, the measure of sustainability is a single dummy variable equal to 0 if the globe rating is below 4 and equal to 1 if the globe rating is 4 or 5. Columns 3 and 4 use dummy variables coding for whether a globe rating of 1, 2, 4, or 5 has been given to a fund. Even-numbered columns also include additional controls. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Normalized flow (t)			
	(1)	(2)	(3)	(4)
Fees	−3.48***	−3.63***	−3.50***	−3.46***
	(0.416)	(0.355)	(0.445)	(0.395)
High sustainability	4.49***	3.11***		
	(0.600)	(0.484)		
High sustainability x Fees	−0.866**	−0.175		
	(0.328)	(0.302)		
One Globe			−0.751	−0.144
			(0.799)	(0.758)
Five Globes			7.40***	5.00***
			(0.889)	(0.786)
One Globe x Fees			0.468	−0.330
			(0.484)	(0.457)
Five Globes x Fees			−2.22***	−1.24**
			(0.539)	(0.497)
(Five − One) Globe x Fees			−2.69	−0.911
p-value			< 0.01	0.139
MSGC × YM FE	×	×	×	×
Other controls		×		×
R^2	0.051	0.102	0.051	0.102
Observations	490,468	461,679	490,468	461,679

Table 1.C.2.
Fixed effects: High vs. low sustainability and flow - fee sensitivity
This table shows the estimated slope coefficients from panel regressions of how net flows vary with sustainability measures and fees when using alternative fixed effects groupings. In all columns, the measure of sustainability is a single dummy variable equal to 0 if the globe rating is below 4 and equal to 1 if the globe rating is 4 or 5. Even-numbered columns also include additional controls. The sample period is from August 2018 to September 2021. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow (t)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fees	-0.228^{***}	-0.170^{***}	-0.227^{***}	-0.165^{***}	-0.228^{***}	-0.204^{***}	-0.227^{***}	-0.199^{***}
	(0.055)	(0.056)	(0.053)	(0.050)	(0.055)	(0.053)	(0.052)	(0.048)
High sust	0.711^{***}	0.573^{***}	0.655^{***}	0.528^{***}	0.710^{***}	0.529^{***}	0.655^{***}	0.458^{***}
	(0.100)	(0.081)	(0.096)	(0.080)	(0.100)	(0.075)	(0.096)	(0.072)
High sust x Fees	-0.234^{***}	-0.167^{***}	-0.205^{***}	-0.128^{**}	-0.234^{***}	-0.169^{***}	-0.206^{***}	-0.116^{**}
	(0.053)	(0.047)	(0.053)	(0.048)	(0.054)	(0.047)	(0.053)	(0.047)
Year-month FE					×	×	×	×
MSGC FE			×	×			×	×
Other controls		×		×		×		×
R^2	0.003	0.019	0.010	0.025	0.007	0.032	0.014	0.041
Observations	490,468	461,679	490,468	461,679	490,468	461,679	490,468	461,679

1.D Unit cost of sustainability

1.E Morningstar Performance Rating

This section provides additional details accompanying the results presented in Section 1.4.2. The Morningstar star ratings were first presented in 1985 and are a backward-looking and monthly rating based on quantitative fund performance. The rating ranges from one (low) to five (high) stars, which are assigned based on a peer ranking within Morningstar style categories. Therefore, the ratings are not influenced by virtue of systematic performance differences between funds relying on different investment styles (for example, growth vs. value). Morningstar measures a fund's performance using its own Risk-Adjusted Return and varying duration (three-, five- and ten-year performance). They then arrive at the final rating by taking the weighted average of all time-period specific ratings. To be eligible for rating, a fund must have been active and disclosed returns for a minimum of 36 months.[18]

[18]For a general overview of the star ratings and specifics on procedures and the particulars that go into calculating them, see The Morningstar Rating™ for Funds.

Table 1.C.3.
Fixed effects: Globe ratings and flow - fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how net flows vary with sustainability measures and fees when using alternative fixed effects groupings. All columns use dummy variables coding for whether a globe rating of 1, 2, 4, or 5 has been given to a fund. Even-numbered columns also include additional controls. The sample period is from August 2018 to September 2021. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow (t)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fees	−0.233***	−0.146**	−0.224***	−0.125**	−0.232***	−0.181***	−0.224***	−0.157***
	(0.063)	(0.064)	(0.059)	(0.059)	(0.063)	(0.062)	(0.059)	(0.056)
One Globe	−0.099	0.031	−0.010	0.096	−0.098	0.050	−0.009	0.135
	(0.139)	(0.126)	(0.134)	(0.122)	(0.138)	(0.123)	(0.134)	(0.119)
Five Globes	1.21***	0.965***	1.15***	0.908***	1.21***	0.897***	1.15***	0.799***
	(0.153)	(0.127)	(0.150)	(0.124)	(0.153)	(0.124)	(0.150)	(0.116)
One Globe x Fees	0.089	−0.038	0.036	−0.096	0.088	−0.031	0.036	−0.102
	(0.083)	(0.075)	(0.078)	(0.071)	(0.082)	(0.074)	(0.078)	(0.071)
Five Globes x Fees	−0.523***	−0.400***	−0.470***	−0.349***	−0.524***	−0.391***	−0.471***	−0.326***
	(0.089)	(0.077)	(0.088)	(0.076)	(0.090)	(0.077)	(0.088)	(0.074)
Diff coefficient	−0.612	−0.362	−0.506	−0.253	−0.612	−0.359	−0.507	−0.224
p-value	< 0.01	< 0.01	< 0.01	0.005	< 0.01	< 0.01	< 0.01	0.013
Year-month FE					×	×	×	×
MSGC FE			×	×			×	×
Other controls		×		×		×		×
R^2	0.003	0.019	0.010	0.025	0.007	0.032	0.014	0.041
Observations	490,468	461,679	490,468	461,679	490,468	461,679	490,468	461,679

Table 1.C.4.
High vs. low sustainability, fund type and normalized flow-fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how normalized flows vary with sustainability measures and fees for different subsamples of funds, i.e., funds with and without an explicit sustainability objective. Columns 1 and 2 consist of the subset of funds labeled as green by Morningstar or that indicate sustainability in the fund name. Columns 3 and 4 analyze the subset of funds given the low carbon label by Morningstar. Lastly, columns 5 and 6 consist of the funds with a globe rating, but without an explicit sustainability objective. All columns use a single dummy variable equal to 0 if the globe rating is below 4 and equal to 1 if the globe rating is 4 or 5. Even-numbered columns also include additional controls. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Normalized flow (t)					
Fund type:	Green		Low carbon		Regular	
	(1)	(2)	(3)	(4)	(5)	(6)
Fees	-5.63^{***}	-3.97^{***}	-0.574	-1.48^{**}	-3.90^{***}	-4.22^{***}
	(0.722)	(0.601)	(0.617)	(0.548)	(0.395)	(0.358)
High sustainability	2.54^{**}	2.90^{**}	2.91^{***}	3.08^{***}	3.69^{***}	3.04^{***}
	(1.17)	(1.12)	(1.04)	(0.879)	(0.700)	(0.599)
High sustainability x Fees	0.894	0.684	-1.17^{*}	-0.930	-1.38^{***}	-0.870^{**}
	(0.782)	(0.748)	(0.678)	(0.611)	(0.420)	(0.379)
MSGC × YM FE	×	×	×	×	×	×
Other controls		×		×		×
R^2	0.082	0.137	0.045	0.113	0.054	0.094
Observations	73,309	65,835	107,593	103,850	309,566	291,994

Table 1.C.5.
Globe ratings, fund type and normalized flow-fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how normalized flows vary with sustainability measures and fees for different subsamples of funds, i.e., funds with and without an explicit sustainability objective. Columns 1 and 2 consist of the subset of funds labeled as green by Morningstar or that indicate sustainability in the fund name. Columns 3 and 4 analyze the subset of funds given the low carbon label by Morningstar. Lastly, columns 5 and 6 consist of the funds with a globe rating, but without an explicit sustainability objective. All columns use dummy variables coding for whether a globe rating of 1, 2, 4, or 5 has been given to a fund. Even-numbered columns also include additional controls. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Normalized flow (t)					
Fund type:	Green		Low carbon		Regular	
	(1)	(2)	(3)	(4)	(5)	(6)
Fees	-5.08^{***}	-3.76^{***}	-0.556	-1.57^{**}	-4.22^{***}	-4.20^{***}
	(0.788)	(0.703)	(0.644)	(0.592)	(0.431)	(0.414)
One Globe	3.38	0.760	3.59^{*}	-0.667	-1.63^{*}	-0.443
	(2.44)	(2.31)	(2.08)	(1.82)	(0.896)	(0.852)
Five Globes	5.20^{***}	4.79^{***}	6.73^{***}	5.13^{***}	5.42^{***}	3.87^{***}
	(1.65)	(1.58)	(1.56)	(1.41)	(1.17)	(1.08)
One Globe x Fees	-1.88	-1.23	-1.27	-0.951	1.36^{**}	0.269
	(1.67)	(1.45)	(1.36)	(1.22)	(0.539)	(0.526)
Five Globes x Fees	-0.045	0.043	-3.08^{***}	-2.11^{**}	-2.56^{***}	-1.95^{***}
	(1.14)	(1.08)	(0.996)	(0.950)	(0.729)	(0.673)
(Five − One) Globe x Fees	1.83	1.27	-1.81	-1.16	-3.92	-2.22
p-value	0.294	0.415	0.248	0.420	< 0.01	0.006
MSGC × YM FE	×	×	×	×	×	×
Other controls		×		×		×
R^2	0.083	0.138	0.046	0.114	0.054	0.094
Observations	73,309	65,835	107,593	103,850	309,566	291,994

Figure 1.D.1.
Fund flows by globe ratings and ESG fee

This figure shows how flows to and from actively managed mutual funds vary when sorting on globe ratings and ESG fees (one globe is assigned to the funds in the bottom 10%, and five globes to the funds in the top 10% of eligible funds). In each domicile-month (time t), funds that have been assigned a Morningstar globe rating are sorted into deciles based on ESG fees independently of their globe rating. After removing year-by-month fixed effects, annualized average monthly net flows in the following month ($t+1$) are plotted for each globe rating. The ESG fee category labeled "Low" contains the funds ranked as the least expensive (ranks 4 and 5) and the "High" category contains the funds ranked the most expensive (ranks 1 and 2).

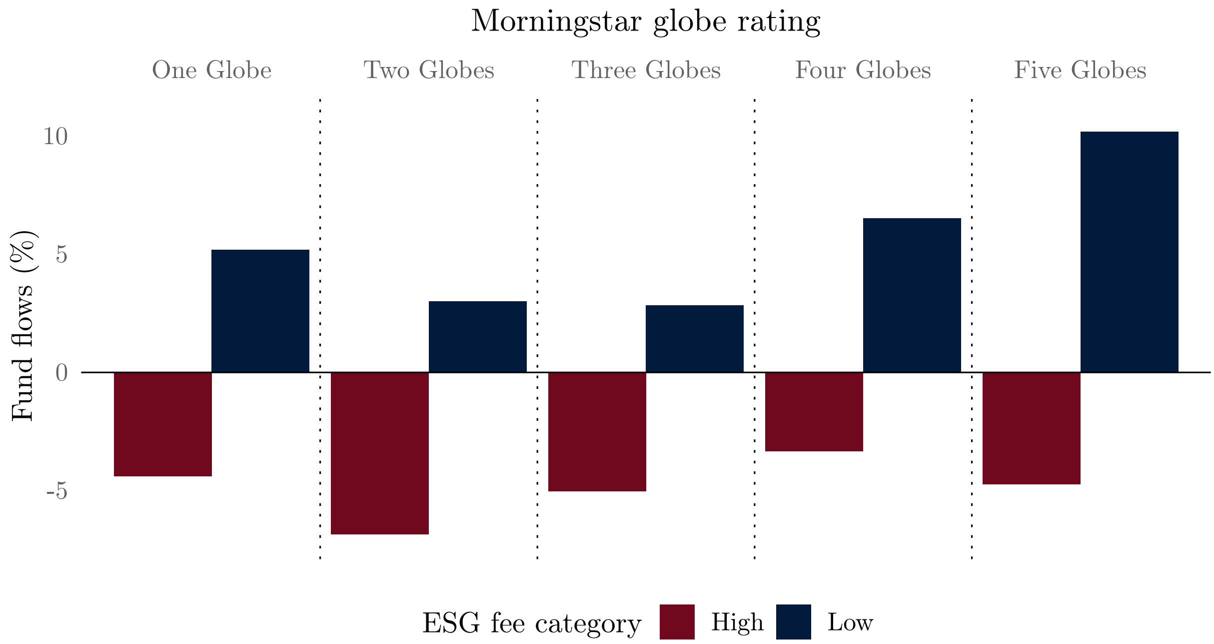


Figure 1.E.1.
Fund flows sorted by star ratings and fees

This figure shows how flows to and from actively managed mutual funds vary when sorting on star ratings and fund fees (one star is assigned to the funds in the bottom 10%, and five stars to the funds in the top 10% of eligible funds). In each domicile-month (time t), funds that have been assigned a Morningstar star rating are sorted into deciles based on fees independently of their star rating. After removing year-by-month fixed effects, annualized average monthly net flows in the following month ($t+1$) are plotted for each star rating. The fee category labeled "Low" contains the funds ranked as the least expensive (ranks 4 and 5) and the "High" category contains the funds ranked the most expensive (ranks 1 and 2).

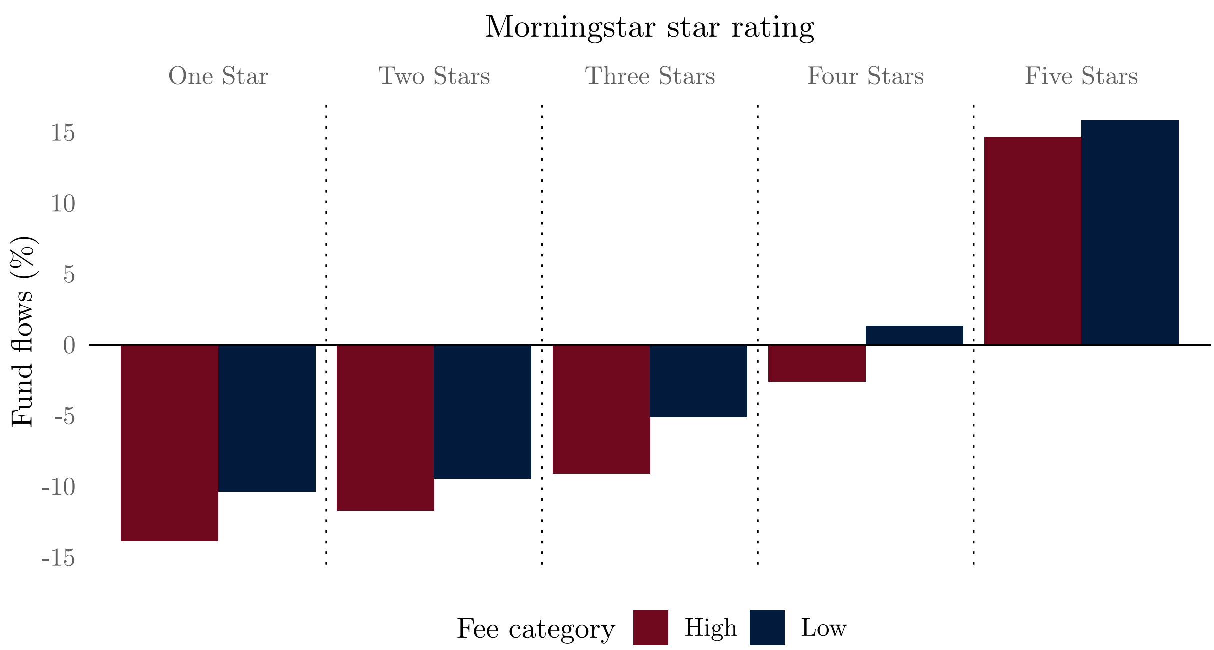


Table 1.E.1.
Star ratings and normalized flow-fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how normalized flows vary with star ratings and fees. Columns 1 and 2 add star ratings as an explicit control variable to the regressions run in Table 1.4, columns 2 and 4. Columns 3 and 4 use dummy variables coding for whether a star rating of 1, 2, 4, or 5 has been given to a fund. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Normalized flow (t)			
	(1)	(2)	(3)	(4)
Fee	−2.78***	−2.46***	−1.45***	−2.83***
	(0.366)	(0.403)	(0.404)	(0.406)
High sustainability	2.43***			
	(0.491)			
Star rating	2.97***	2.97***		
	(0.160)	(0.160)		
High sustainability x Fees	−0.053			
	(0.311)			
One Globe		1.20		
		(0.758)		
Five Globes		4.65***		
		(0.817)		
One Globe x Fees		−0.833*		
		(0.466)		
Five Globes x Fees		−1.31**		
		(0.522)		
One star			−4.62***	−3.13***
			(0.927)	(0.896)
Five stars			14.2***	10.6***
			(0.887)	(0.801)
One star x Fee			0.323	0.791
			(0.535)	(0.517)
Five stars x Fee			0.007	−0.681
			(0.524)	(0.508)
Diff coefficient		−0.473	−0.317	−1.47
p-value		0.455	0.650	0.026
MSGC × YM FE	×	×	×	×
Other controls	×	×		×
R^2	0.108	0.108	0.081	0.109
Observations	428,149	428,149	430,514	428,149

Table 1.E.2.
Fund type and normalized flow-fee sensitivity

This table shows the estimated slope coefficients from panel regressions of how normalized flows vary with sustainability measures and fees for different subsamples of funds, i.e., funds with and without an explicit sustainability objective when star rating is added as a control variable. Columns 1 and 2 consist of the subset of funds labeled as green by Morningstar or that indicate sustainability in the fund name. Columns 3 and 4 analyze the subset of funds given the low carbon label by Morningstar. Lastly, columns 5 and 6 consist of the funds with a globe rating, but without an explicit sustainability objective. Odd-numbered columns use a single dummy variable that equals 0 if the globe rating is below 4 and 1 if the globe rating is 4 or 5. Even-numbered columns use dummy variables coding for whether a globe rating of 1, 2, 4, or 5 has been given to a fund. The sample period is from August 2018 to September 2021. All columns include year × Morningstar global category × month fixed effects. Analysis is at the fund level. Standard errors clustered by fund and month are in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Normalized flow (t)					
Fund type:	Green		Low carbon		Regular	
	(1)	(2)	(3)	(4)	(5)	(6)
Fees	−3.35***	−2.97***	−0.125	−0.071	−3.51***	−3.34***
	(0.623)	(0.743)	(0.560)	(0.599)	(0.375)	(0.430)
High sustainability	2.66**		2.58***		2.31***	
	(1.20)		(0.879)		(0.612)	
High sustainability x Fees	0.694		−1.03		−0.614	
	(0.771)		(0.622)		(0.391)	
Star rating	1.98***	1.95***	3.64***	3.63***	2.83***	2.84***
	(0.303)	(0.302)	(0.235)	(0.235)	(0.193)	(0.193)
One Globe		1.45		0.413		0.959
		(2.44)		(1.89)		(0.864)
Five Globes		4.86***		4.12***		4.20***
		(1.66)		(1.43)		(1.11)
One Globe x Fees		−1.71		−1.18		−0.281
		(1.53)		(1.34)		(0.540)
Five Globes x Fees		−0.217		−2.09**		−2.18***
		(1.12)		(0.988)		(0.686)
(Five − One) Globe x Fees		−0.318		−0.011		−0.277
p-value		0.261		0.958		0.019
MSGC × YM FE	×	×	×	×	×	×
Other controls	×	×	×	×	×	×
R^2	0.136	0.136	0.125	0.126	0.102	0.102
Observations	60,263	60,263	97,302	97,302	270,584	270,584

Forced to be Active:
Evidence From a Regulation Intervention*

Petter Bjerksund, Trond Døskeland, André Wattø Sjuve, and Andreas Ørpetveit

Abstract

Closet indexers are low-activity mutual funds that are sold and marketed as active. Their investors therefore only receive part of the service they pay for. Supervisory authorities all over the world are now considering how these funds should be regulated. We examine evidence from interventions carried out by Scandinavian regulators. The impact is identified by comparing scrutinized Scandinavian closet index funds with similar unaffected European funds. Given the choice between reducing fees or increasing activity, the scrutinized funds opt for the latter. Although this results in a more actively managed fund, performance deteriorates. Thus, regulation leads to the worst outcome.

Keywords: Asset management; Active portfolio management; Regulation of financial services
JEL Classification D14; G11; K12

*Bjerksund and Døskeland have been expert witnesses for the Norwegian Consumer Council in the lawsuit described in this paper between the Consumer Council and DNB. Corresponding author is Trond Døskeland, email: trond.doskeland@nhh.no and address: Helleveien 30, 5045 Bergen, Norway.

2.1 Introduction

Actively managed mutual funds sell the potential to beat their benchmark (usually a market index). Investors who choose these types of funds are typically looking for an opportunity to outperform the market. To that end, they pay a premium over index funds for the service of dedicated fund managers who try to beat the market.[1] The fund managers' efforts may justify this extra cost if it creates an opportunity to make excess returns by deviating from the fund's benchmark. However, several papers have identified funds with relatively high fees, which, at the same time, have a low degree of active management (see, e.g., Cremers and Petajisto (2009), Petajisto (2013), or Cremers et al. (2016)). These funds are labeled "closet indexers". Thus, investors in these funds do not receive the active portfolio management service they pay for.

Financial supervisory authorities around the world have addressed the problem of closet index funds.[2] The Scandinavian countries (Denmark, Norway, and Sweden) were early to put pressure on potential closet indexers, and carried out extensive investigations in 2014 and 2015.[3] The recommendations directed at closet indexers were either to update investor information and reduce the fee or increase the fund's activity. This paper examines the impact of policy scrutiny by comparing Scandinavian closet index funds exposed to scrutiny with unaffected European closet index funds.

We hypothesize that funds under scrutiny will opt to increase activity and maintain their fee level. Cutting fees is expensive, while increasing activity is cheap by comparison. Rephrased, the funds are "forced" to increase activity. Furthermore, we investigate the consequences of increased activity for investors. One constraint much discussed in the active management literature is diseconomies of scale, i.e., that the quality of marginal investment opportunities declines as active assets under management (AUM) increase. When not subject to an anti-closet-indexing constraint, managers will choose a (subjective) threshold alpha and closet-index assets when opportunities fall below that point. Gross investor returns are maximized when the threshold alpha is zero (Berk and Green, 2004). However, managers may use different threshold alphas. If they are risk- or effort-averse, they will choose a positive threshold alpha, while if they are overconfident, risk-seeking (e.g., due to convex incentives), or desire to signal skill via activity,

[1] See, for example, Morningstar (2019).
[2] See, for example, ESMA (2016), New York Office of the Attorney General (2018), Financial Conduct Authority (2019), and ESMA (2020).
[3] See, for example, Reuters article (2017) or Kjørven (2019).

they will choose a negative threshold alpha.[4]

Using regulation to induce additional active management may benefit investors if the threshold is positive, but not if it is negative. Our empirical work addresses this by analyzing the effect on investor returns of inducing additional active management. If managers have already exploited their best investment ideas, more activity will lead to lower excess returns. Conversely, a positive threshold alpha implies that investors will benefit from an increased activity level.

Before we report our findings, we set out the history of closet indexing in Scandinavia, focusing in particular on the world's first closet index fund to be ordered by the courts to repay its investors.[5] Closet indexing has been a concern in all the Scandinavian countries, and the region's different national level Financial Supervisory Authorities (FSAs) have identified questionable practices (Kjørven, 2019). As an example, we describe the Norwegian mutual fund, DNB Norge, managed by the largest banking group in Norway, DNB. Investor information and an annual fee of 1.8% suggested that investors should expect active management. Figure 2.1.1 plots the development of DNB Norge's active share from the start of 2010 until May 2018. The Norwegian FSA investigated DNB Norge from 2010 until 2014. A quantitative analysis of the degree of active management identified a tracking error of 1.28% per year and an average active share of 12.25%. Based on the discrepancy between what the investors had reason to expect and the actual degree of active management, the FSA concluded that DNB Norge was a closet index fund.

The Norwegian FSA imposed a corrective order on DNB and gave the fund two alternatives.[6] They could bring the fund's management in line with the active management characteristics reflected by the fee and the fund's prospectus or adjust the fee to a level in line with the applied investment strategy. DNB decided to implement a combination of the two. They reduced the fee from 1.8% to 1.4% and more than doubled the fund's active share. Figure 2.1.1 shows how the active share more than doubled by the end of 2014.

Based on the corrective order, the Norwegian Consumer Council (NCC) filed a class-action lawsuit on behalf of the fund's 180,000 investors claiming that DNB had misled and charged its investors excessive management fees.[7] In the first instance, Oslo District Court, DNB won. In

[4] What is optimal for the investor is not necessarily optimal for the fund manager. There are many reasons for conflict of interest between manager and investor. For example, they may have different skills or information sets (see, e.g., Sirri and Tufano (1998) and Agarwal et al. (2014)), risk preferences (see, e.g., Huang et al. (2011)), or incentives (see, e.g., Sensoy (2009) and Ferreira et al. (2018)).

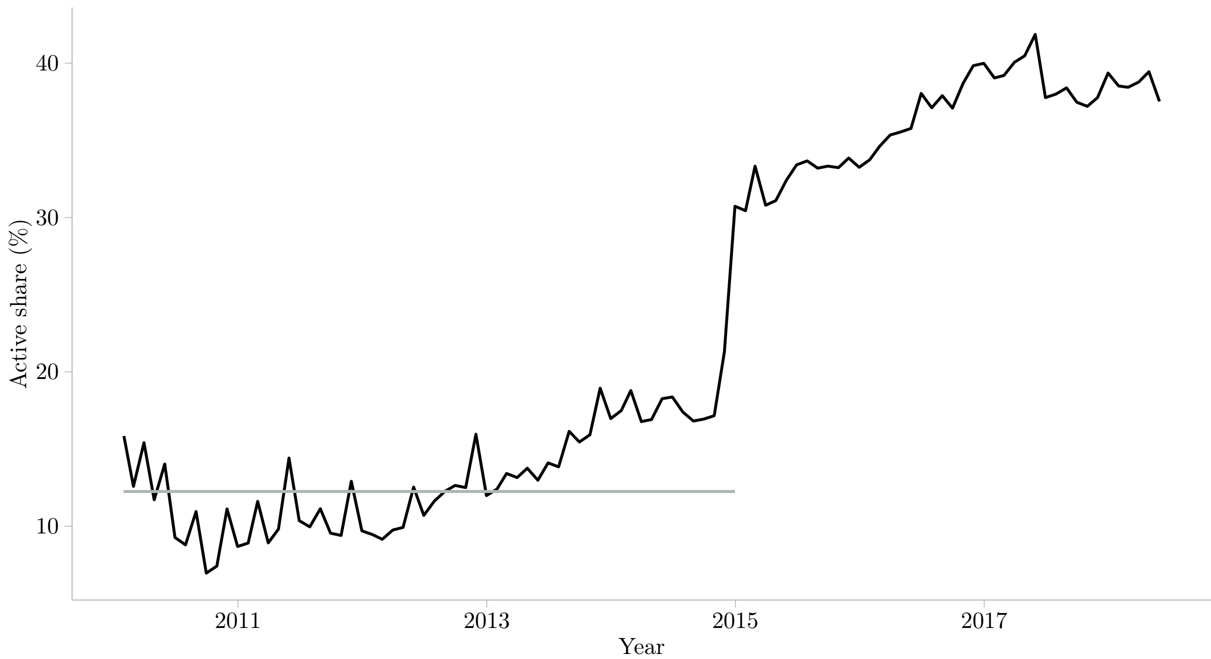
[5] There are some cases where the asset managers have accepted a fine (Portfolio Adviser article (2019)) or where there are ongoing lawsuits (Financial Times (Nov. 2021)).

[6] This corrective order is available with an English translation at Finanstilsynet.

[7] For details see Norwegian Consumer Council (2015), and Kjørven (2019).

Figure 2.1.1.
Active share for DNB Norge

This figure presents the development of active share for the mutual fund DNB Norge over the period 2010 to 2018. The horizontal line at 12.25% between 2010 and 2014 is the average active share for the fund leading up to the corrective order issued by the Norwegian Financial Supervisory Authority.



the second instance, Borgarting Court of Appeal ruled in favor of the consumers, and ordered DNB to pay back 0.8% of the management fee. DNB then appealed to the Norwegian Supreme Court, whose ruling was delivered on February 27, 2020, upholding the Court of Appeal's judgment. The verdict can be found here. Further details about the lawsuit can be found in Appendix 2.A. The DNB case is unique in that it has brought the closet index problem to court. However, as we show in this paper, the corrective order on DNB and other similar Scandinavian examples have had spill-over effects on potential index huggers, in that many funds changed behavior to avoid being classified as a closet index fund. The consequences of this new behavior are largely unknown. Closet index funds have become more active, but has this benefited the investors?

At one level, the impact of the intervention is by its very nature interesting. We attribute the changes in fund behavior to the intervention, however we cannot be certain whether the effects can solely be assigned to this or whether they may be due, for instance, to a general change in the conditions for active management. It could be that the timing of the attention from the FSAs coincided with unfavorable market conditions for active management, which means that we cannot definitively conclude what causes a potential low alpha in active management.[8] These endogeneity concerns must be addressed.

To address these endogeneity concerns and identify the impact of the intervention, we use the exogenous variation created by the policy scrutiny. Estimates of the scrutiny effect come from a difference-in-differences (diff-in-diff) research design, where we compare outcomes from Scandinavian closet index funds exposed to scrutiny (treated funds) with European closet index funds not under scrutiny (control funds). In robustness tests, found in Appendix 2.G, truly active Scandinavian funds serve as our control group. Despite the selection problem bias associated with these funds, it is reassuring that the results are largely consistent across control group specification.

We split the impact assessment into two "stages" based on the timing of the event. First, we investigate the impact of the policy scrutiny on active share and fee. In the second step, we examine the consequences for investors' alpha.

Regarding the first stage, we find the unsurprising result that fund managers respond to scrutiny by opting to increase activity over reducing the fee. The diff-in-diff estimate shows an increase in active share of over 9%. Fee estimates are indifferent from zero. For the second stage, we find that the closet index funds that become more active perform worse than a comparable control group of closet index funds. We find that the funds under scrutiny had, on average, a lower annual alpha of 1.3% than comparable funds unaffected by scrutiny.

The empirical analysis shows that scrutinizing closet index funds produces unwillingly active funds, which is the worst alternative. The best solution is not to force funds to become more active but to promote fee reduction. If the regulatory authorities in the DNB case complied with the alternative to update investor information and reduce fees, the investors would have fared better. Assuming the fund had performed as well as the average control fund and had reduced the fee by half of what DNB had to pay back, the annual alpha would have been 1.7% ($= 1.3\% + 0.4\%$) higher.[9]

The rest of the paper is structured as follows. In Section 2.2, we discuss related literature and emphasize our relative contribution. To assess the intervention's effect, we use a framework

[8] There are different ways of measuring value creation in active management. Most often, we use the term alpha. By that, we mean the difference between the actual portfolio and the benchmark return. This measure is also known as the active or excess return. Alpha can also be used to describe active returns risk-adjusted for factor risk. Finally, alpha can be calculated before or after fees, i.e., gross or net alpha, respectively.

[9] The estimate of loss comes from adding the diff-in-diff estimate for alpha and half the 0.8% DNB had to pay back to its investors.

proposed by the UK Financial Conduct Authority (Financial Conduct Authority, 2016). In Section 2.3: Problem diagnosis, we develop an understanding of the problems associated with closet indexing. Section 2.4: Intervention design describes the intervention and presents the data with summary statistics. In Section 2.5: Impact assessment, we examine how fund managers respond to scrutiny and how it affects investors. In Section 2.6, we go into more detail and examine the robustness of our previous findings. We have tried to keep the number of analyses in the main text to a reasonable level, and we refer to the appendices for additional material. Finally, we suggest some policy implications and conclude in Section 2.7.

2.2 Related literature and contribution

There is an extensive literature on closet indexing. Although Berk and Green (2004) mention the term in their paper, it was not until the introduction of the active share measure by Cremers and Petajisto (2009) that the problem became the subject of detailed study and media attention. Many papers have documented that closet indexing exists all over the world (Cremers et al., 2016).[10] However, there is a difference between identifying the problem and solving it. Market failure often calls for regulation, however it is not always the best solution and may, in fact, make matters worse. There is little literature on the regulation of closet index funds; one exception is Cremers and Quinn (2016), and even less literature examining such regulation. Hopefully, our assessment will be of interest to other markets such as the U.S. mutual fund market.

The intervention also provides insight into a more general question; whether regulatory agencies should intervene to correct market failures. The increased complexity of financial arrangements poses a challenge to households managing their financial affairs and to regulators attempting to assist them. There has long been a tension in economics between laissez-faire economists who appreciate and defend the performance of free markets and interventionists who identify market failures and argue that feasible policies can be found to correct them (Campbell, 2016). When households are unable to manage their financial decisions, they make mistakes that reduce their welfare with broader consequences for the economy. However, regulation can fail and instead exacerbate the problem it set out to solve. As a rule, it is hard to perform a cost-benefit analysis in financial regulation (Sunstein, 2015). Our study provides insight into

[10] Most of the literature focuses on the connection between active share and performance, see, for example, Cremers and Petajisto (2009), Cremers et al. (2021), and ESMA (2020).

how the effects of intervention can be assessed.[11]

Furthermore, we contribute to the literature on the scalability of active management. The extent to which an active fund can outperform its passive benchmark depends not only on its raw skill in identifying investment opportunities but also on the various constraints on the fund. A constraint subject to recent discussion is diseconomies of scale. If scale impacts performance, skill and scale interact: for example, a larger and more skilled fund can underperform a less skilled small fund. Therefore, to identify the skill, we must also understand scale effects.

The theoretical model developed by Berk and Green (2004) relies on the key assumption of diseconomies of scale in active management. Managers eventually run out of ideas and cannot generate additional alpha. Yet, the empirical evidence on the relationship between fund size and performance is mixed. Table 13 in Adams et al. (2021) provides an overview of the literature. For U.S. funds, at the fund-level, Chen et al. (2004), Edelen et al. (2007), and Yan (2008) document a significantly negative relationship between size and performance. However, these findings are challenged by studies that identify an endogeneity issue in the test of the return-to-scale property. Reuter and Zitzewitz (2021) examine the size-performance relationship in a natural experiment setting, applying a regression discontinuity approach. Pástor et al. (2015), Zhu (2018) and Pastor et al. (2021) address the omitted-variable bias by including fund fixed effects to account for heterogeneity in managerial skills. Phillips et al. (2018) use instrumental variables that are correlated with size but unrelated to recent performance. Finally, McLemore (2019) uses fund mergers as shocks to fund size. All these studies report a negative but mostly insignificant relationship. For non-U.S. funds, Ferreira et al. (2013) find increasing returns to scale.

This paper re-examines the size-performance relationship at the fund level with a novel identification strategy addressing the endogeneity issue for domestic European funds. If we accept that asset managers are forced to increase their activity level and have limited opportunities to use in- or outflow to do so, we introduce variation in the quantity of assets under active management. In such a quasi-natural experiment setting, we can identify the size-performance relationship.

Our findings suggest that we have decreasing returns to scale with the mean performance estimates in the treated groups being negative relative to the different control groups. The

[11] Financial Conduct Authority (2016) provides an overview of the British approach to an economic analysis of financial regulation. They outline a methodology for regulatory economic analysis that sets out a three-stage process, including problem diagnosis, intervention design, and impact assessment.

result also holds when we use a more direct measure of marginal bets. We find that the primary source of the underperformance comes from the new bets taken by managers under scrutiny. Our detailed data allow us to learn how managers rearrange their portfolios when pushed to increase their active share. We find that new bets from forced managers underperform new bets from untreated managers.

Finally, a problem associated with traditional activity measures, such as tracking error and active share, is that they do not detect whether funds perform true active management by taking new bets, increase existing ones, or engage in signal-jamming to appear truly active (Brown and Davies (2017) or Cremers et al. (2022)). Signal jamming, i.e., taking random bets, improves closet indexers' chance to pool with genuinely active funds by adding tracking error to their returns, thereby creating a false sense of active management. Moreover, this strategy implies that closet indexing may be more widespread than indicated by traditional measures, such as tracking error and active share.

To distinguish true active management from signal jamming, we examine a portfolio-level concentration measure (used, for example, by Brands et al. (2005) and Kacperczyk et al. (2014)). At the total active portfolio level, we identify more concentrated bets. When dividing the portfolios into sub-portfolios based on bet size, we find that managers take more concentrated bets in new stocks. This finding is in contrast to Pollet and Wilson (2008), who find that funds increase ownership in already owned stocks in response to increased inflows, especially if they operate in relatively illiquid markets. Finally, we find that the increased activity does not come from larger bets in managers' initial "best ideas" (Pomorski (2009) and Antón et al. (2021)).

2.3 Problem diagnosis

In this section, we develop an understanding of the problems associated with closet indexing and outline the drivers of poor outcomes resulting from the underlying market imperfections.

2.3.1 Closet indexing

Closet indexing is problematic because the investors do not get the actively managed funds they pay for and are promised by the fund company. When investors buy active mutual funds, they evaluate investor information, including fee structure.[12] This information then provides the foundation for the investors' expectation of active management. If this is incorrect, investors will base investment decisions on a false expectation of a more active fund management service than they actually receive (Cremers and Quinn, 2016). Closet indexing does not offer the same ex ante risk profile that investors should expect from genuine active management.

Evaluating the services rendered is arduous for mutual fund retail investors because manager effort is only partially observable, and, even if monitoring is possible, interpreting the information may be difficult. Therefore, to mitigate moral hazard, we need suitable measures of effort.[13] A natural activity measure is the actual outcome of active management, the alpha. However, this is not a reliable measure of the degree of active management. A manager can be very active, but if the bets cancel each other out, alpha is close to zero. Thus, realized returns cannot be used to identify closet index funds. Measures must be based on manager effort.[14]

The most used means of identifying closet indexing is active share, which evaluates the degree of active management for funds relative to the benchmark (Cremers and Petajisto, 2009).[15] An alternative is tracking error. We will go on to show that active share and tracking error are highly correlated for European domestic funds. If tracking error is low, active share is low, and vice versa. Therefore, we focus on active share as our activity measure.

To outperform its benchmark, the active portfolio must differ from the benchmark portfolio. As such, activity is a precursor to superior returns. Since closet index funds charge fees akin to truly active funds, while holding portfolios similar to the index, the net performance of closet index funds is on average lower than the net performance of the much cheaper index funds (ESMA, 2020).

2.3.2 Motives for closet indexing

To crack down on closet indexing, it is important to understand the motives behind it. According to Berk and Green (2004), the quality of marginal investment opportunities declines as active

[12] The Key Investor Information Document (KIID) includes the most important investor information. Commission Regulation 583/2010 provides a harmonized regime on the form and content of the document, ensuring that information in the UCITS markets is consistent and comparable. KIIDs include sections on objectives and investment policy, risk and reward profile, fees, past performance, and practical information.

[13] There may be different views on the manager's obligation. In the DNB Norge case, the fund claimed in court that the obligation was to use resources to search for bets, not to implement them. We assume that we can measure the obligation.

[14] In the law there is a distinction between the duty to achieve a specific result and the duty of best effort, for more about this, see, for example, UNIDRIOT Principles.

[15] Active share is defined as one half of the sum of the absolute value of the difference between portfolio and benchmark weights.

AUM increases. When not subject to an anti-closet-indexing constraint, managers will choose a (subjective) threshold alpha and closet-index assets when opportunities fall below that point. Gross investor returns are maximized when the threshold alpha is zero.[16]

However, managers may use different threshold levels. Being risk- or effort-averse leads to a positive threshold alpha. A negative one is chosen if they are overconfident, risk seeking (e.g., due to convex incentives), or have a desire to signal skill via activity. Cremers and Quinn (2016) suggested four alternative motives for closet indexing. We describe two of them, one resulting in a positive threshold alpha and one in a negative one.[17]

If managers want to preserve their current asset base and are afraid that underperformance may lead to a large outflow from the fund, they can set a positive threshold alpha. Although a higher active share is optimal for investors, the managers choose a closet indexing strategy. All else equal, larger funds generate more revenue than smaller funds. However, there is an asymmetry in the relationship between flows and performance. While small funds seek to create superior returns to grow their asset base, large funds may prefer to preserve their current assets and avoid substantial losses to the benchmark (Sirri and Tufano, 1998). One way to prevent underperformance is to put a larger share of the fund's assets in the benchmark. While this strategy reduces the fund's likelihood of beating the benchmark, average performance may be enough to maintain a large asset pool, and consequently its profitability. Therefore, closet indexing may be a valuable strategy for a risk-averse manager seeking to maximize assets under management. If the manager uses a positive threshold alpha, the investor will not benefit from the manager's skill, much like when a good soccer player is benched for most of a game. In this case, the investor would be better off if the manager increased the degree of activity.

If managers have run out of new ideas, they may set a threshold equal to or lower than zero. Assuming that managers possess private information about their skills, which are lower than the investors believe, they will run out of good ideas more quickly than set out in the investor information. Size may be a reason why managers run out of ideas. A large fund has a more limited set of investment opportunities consisting of only sufficiently large investments to make a difference. In this case, higher activity does not benefit the investor.

[16] With this framework, we relax the assumption that capital is competitively allocated (Berk and Green, 2004). In our setting, there are several reasons why this condition does not hold. Firstly, the event window is short and retail investors are typically slow at changing their positions, i.e., they "suffer" from inertia (see, for example, Bilias et al. (2010) or Agnew et al. (2003)). Secondly, tax motives do not favor reallocation between funds. Before and during the event, markets went up, meaning that most investors had large and unrealized investment gains. Had they sold, this would have triggered capital gains tax.

[17] We do not expand on the following two motives: 1) closet indexing is chosen due to the high cost of performing true active management, or 2) closet indexing is chosen due to time-varying active management possibilities.

2.3.3 The rationale for intervening

Investors buy active funds for the opportunity to beat the index alternative. However, closet indexing is a significant drag on mutual fund investors' returns, and closet index funds underperform the market, leaving investors worse off than for other investment choices. Thus, there is a potential to increase welfare by accurately regulating these funds.

Cremers and Quinn (2016) suggest two approaches to mitigate the closet indexing problem. Firstly, they suggest a disclosure regime that would incorporate more information, such as active share. Secondly, they examine whether closet indexing is potentially liable under existing laws. In the next section, we show that Scandinavian regulatory authorities used elements of both these approaches during the scrutiny period in our event.

2.4 Intervention design

This section describes the intervention design, and the laboratory (active domestic European funds) that we use to answer the questions.

2.4.1 Policy scrutiny in Europe

Although most of the literature focuses on U.S. mutual funds, some papers have investigated European funds (see, for example, Banegas et al. (2013), Ferreira et al. (2013), Cremers et al. (2016) and Leippold and Rueegg (2020)). Using a difference-in-differences design, we compare Scandinavian funds (treatment group) exposed to policy scrutiny with unaffected funds (control group) in other European countries where we are unable to identify a meaningful level of scrutiny. As well as being the first to have a regulatory focus on closet indexing in Europe (or the world), the Scandinavian countries also have the toughest policy interventions.[18] Below we describe each country's scrutiny in detail.

Denmark

Although we set an exact date in Table 2.4.1, these scrutiny processes take a considerable amount of time. Therefore, we define an event window of two years. At the beginning of 2014, the Financial Supervisory Authority of Denmark released an analysis of closet indexing in their report for 2013. Using limits of 60% for active share and 4% for tracking error, they found

[18] Two articles from Financial Times: Financial Times article (2016a) or Financial Times article (2016b).

Table 2.4.1.
Country selection

This table presents the countries in Europe divided into treated, control, or omitted from the sample, respectively. As a starting point, we use the countries included in Ferreira et al. (2013) and Cremers et al. (2016).

Country	Date and documentation
	Panel I: Treated
Denmark	*September 2014* Danish FSA publishes a report showing that 56 out of 188 funds studied were potential closet indexers. More details in text.
Norway	*March 2015* Norwegian FSA issues a corrective order to DNB asset management regarding active management. More details in text.
Sweden	*October 2015* Finansinspektionen (FI) publishes a report debating stricter rules on consumer protection in financial markets. More details in text.
	Panel II: Control
Austria	No scrutiny identified.
Belgium	No scrutiny identified.
Finland	No scrutiny identified.
Poland	No scrutiny identified.
Portugal	No scrutiny identified.
Switzerland	No scrutiny identified.
	Panel III: Omitted
Italy	*March 2016* The Italian regulator takes action against some of the largest investment companies in its domestic market for mis-selling actively managed funds that closely hugged an index.
Netherlands	*May 2016* AFM publishes a report on index hugging identifying 7 out of 85 funds investigated as closet trackers.
Germany	*September 2016* BaFin completes its investigation into closet indexing, identifying deficiencies in transparency.
France	*March 2017* AMF reminds asset management firms of the importance of clarity in the investment objective.
United Kingdom	*June 2017* FCA publishes their final report on the Asset Management Market study finding £109 bn invested in closet funds.
Luxembourg	*August 2017* CSSF issues a reminder on improving clarity in the “objectives and investment policy” section of the KIID.
Spain	*October 2018* CNMV analyzes the existence of these products without reaching any conclusion on how to act. More details here.
Ireland	*July 2019* Central Bank of Ireland publishes largest data driven study of industry about closet indexing to date.

that 56 out of 188 equity mutual funds had not practiced the active management strategy they marketed in their prospectuses (Financial Supervisory Authority of Denmark, 2013). When the FSA lowered the limits to 50% and 3% for active share and tracking error, respectively, the number of potential closet indexers was reduced to 22 funds. Based on the report, the FSA contacted the boards of these funds and requested explanations. Apparently, they were satisfied with the answers as no further action was taken. However, the funds were subsequently required to report active share and tracking error.

Norway

During the same period, the Norwegian Financial Supervisory Authority wrote an extensive report on the level of active management for a subsample of Norwegian mutual funds (Norwegian Ministry of Finance, 2015). Based on their findings, they chose to publicly criticize two funds in November 2014. We have already mentioned the case of DNB Norge. As illustrated in Figure 2.1.1 in the Introduction, the fund had an exceptionally low active share. The Norwegian FSA decided to impose a corrective order on DNB, to either bring the management of the fund in line with the characteristics of true active management, as reflected by the management fee and in the fund's prospectus, or adjust the fee to a level in line with the strategy applied. The second fund that the FSA criticized was Nordea Avkastning. They were given the same options, either change the level of activity or the fee, although the activity level was higher than for DNB Norge.[19]

Sweden

The Swedish Financial Supervisory Authority analyzed Swedish actively managed mutual funds in 2014. They examined whether the key investors' documents of funds marketed and sold in Sweden provided accurate and clear investment objective and policy information. The investigation is presented in the Swedish FSA's annual report on Consumer Protection 2015 (Financial Supervisory Authority of Sweden, 2015). The intervention started a debate on the legal issues related to closet indexing. In 2014, the Swedish Shareholders' Association (Sveriges Aktiesparares Riksforbund) initiated a class action against two mutual funds from one of the largest Swedish banks, deciding, however, not to go through with the lawsuit in July 2015.[20] For the Scandinavian countries, Kjørven (2019) analyzes how the European legal framework has been applied and discusses the need for legal measures to ensure that investors get what they pay for, as protection against closet indexing.

[19] The Financial Supervisory Authority of Norway, available in Norwegian at Finanstilsynet.
[20] For details about the funds, see case number 2014-11304 (Swedish Ministry of Finance, 2016).

Rest of Europe

Closely following Scandinavia, the European Securities and Market Authorities (ESMA, 2016) wrote an extensive report based on more than 2,600 European funds to map the prevalence of closet indexers. Their findings suggest that between 5% and 15% of the sample funds were potential closet indexers. As we see in Table 2.4.1, several other countries performed their own investigations. To our knowledge, none of these have resulted in any legal claims against funds.

2.4.2 Event design

To conduct an event study, we need to follow the mutual funds before, during, and after the intervention. We let the window of policy scrutiny start in January 2014 and last until December 2015. The pre-event period is the two years before January 2014, and the post-event period is the four years after December 2015, from January 2016 until December 2019.

We have two groups, treated and control funds. For the treated funds, we cannot always directly attribute the consequences of scrutiny to the behavior of the asset managers. In that sense, the DNB Norge case is an exception. However, based on the policymakers' and financial authorities' interventions and the subsequent media interest, we assume that the funds at risk of being labeled a closet index fund were rethinking their active management strategy during the two-year window from January 2014 to December 2015.

To identify a control group as free as possible from scrutiny, we have sorted all the countries in Europe, except Scandinavia, into two groups, those with and without identified policy scrutiny. In Panel II: Control in Table 2.4.1, we list the countries that constitute our control group, and in Panel III: Omitted, we list the countries that have scrutinized the fund industry. For the latter group, we also add links with further details.

We have also performed a robustness test where we replace our primary control group with truly active funds from Scandinavia. Due to a selection problem, i.e., comparing "sick" closet index funds with "healthy" truly active funds, we do not use this as our primary control group. However, we find almost identical results, indicating robust results. For details, see Appendix 2.G.

2.4.3 Data and summary statistics

This section describes the data, how we constructed our treatment and control groups, and presents summary statistics.

Sample selection

The dataset is built using two primary databases: Morningstar Direct and Lipper Fund Database. The data span from January 2010 through December 2019. Our focus is on domestic long-only equity funds.[21] The fund data include monthly assets under management and gross and net returns.[22] The fee is the price of active management and is calculated using the difference between gross and net returns. To calculate active share, we use monthly holdings for each fund.[23] Details on the sample selection and raw data are presented in Appendix 2.B.1.

A benchmark is defined for each fund to calculate active share and performance. In general, categorization is complicated for active funds, however the error is minimized by using domestic funds. These funds give us the most homogeneous group of funds both across and within countries. Benchmark constituents, weights, and returns come from Datastream. We use the Lipper technical benchmarks whenever these are available. This benchmark assignment method minimizes the concern that funds strategically choose benchmarks that may not accurately reflect their actual investment style. Details regarding benchmark data are provided in Appendix 2.B.2.

To avoid survivorship bias, we also include funds that died during the sample period. However, we exclude funds with less than 12 months of data in either the pre- or post-event period to draw meaningful inference. After these exclusions, we are left with a sample consisting of 353 funds, from which we define closet index funds. The full sample is presented in Table 2.B.2 in Appendix 2.B.3.

To evaluate factor returns, we collect size and style portfolio returns from MSCI. All variables, divided into outcome and controls, can be found in Table 2.B.3 in Appendix 2.B.4. Continuous variables are winsorized at the 1^{st} and 99^{th} percentiles to mitigate the potential impact of outliers.

Defining closet indexing

There are several ways to define a closet index fund, the two most common of which are active share and tracking error. In Appendix 2.B.5, we show that the factor structure is weak for domestic funds, which implies a high correlation between active share and tracking error. Thus, we choose to focus on active share in our analyses. To define a fund as a closet index fund, we

[21] We define domestic funds as funds with an investment area equal to the home country of the investment company and a domestic primary prospectus benchmark.

[22] Returns are in the local currency, while assets under management are in USD to establish a common currency for comparison across countries.

[23] We have used Morningstar data for funds missing data in Lipper.

set two cutoff points, 40% and 50%, with funds being classified at the start of the event window. We hypothesize that scrutiny exposure increases with lower activity levels. Therefore, when comparing the two samples, the estimates are largest for the lowest cutoff point.

2.4.4 Summary statistics

In Table 2.4.2, we report summary statistics for the treated sample (Scandinavian funds) and the control sample (European, non-Scandinavian funds). Active share and other fund characteristics are measured before the event window. We denote the benchmark-adjusted performance as alpha. We agree with Berk and van Binsbergen (2015) that a tradeable index-based adjustment is likely to adjust for fund style and risk more accurately than the loadings on risk factors. However, in robustness tests, factor-adjusted alphas are also analyzed. The variables are explained in Table 2.B.3 in Appendix 2.B.4.

Table 2.4.2.
Summary statistics: closet index funds

This table presents summary statistics for the sample of closet index funds in Scandinavia (Norway, Sweden, and Denmark) and the rest of Europe. Values are means over a two-year window before the event start. Gross alpha, expense ratio, and net alpha are annualized. Displayed values for competition are country-means for the Scandinavian countries and the rest of Europe, respectively (see Table 2.B.3 for details). Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. Inference on differences between treatment and control funds is based on Newey and West (1987) standard errors. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

Fund sample	Active share ≤ 40%			Active share ≤ 50%		
Group	Treated	Control	Difference	Treated	Control	Difference
Number of funds	46	33		75	47	
Active share (%)	32.4	32.2	0.2	38.0	36.5	1.5
Gross alpha (%)	1.04	0.59	0.46	1.24	1.03	0.20
Expense ratio (%)	1.14	1.18	−0.04	1.25	1.19	0.06
Net alpha (%)	−0.13	−0.60	0.47	−0.05	−0.18	0.12
AUM (million USD)	474	221	253**	377	225	152
Fund age (years)	16.2	16.0	0.2	15.0	15.4	−0.4
Competition	0.84	0.81	0.03	0.84	0.81	0.03

In columns 1 to 3, we report the numbers for the cutoff at 40%. Out of 79 funds, 46 are treated and 33 are controls. The average active share is about 32%. Regarding alpha, we find similar results as other studies in the literature, with a small positive alpha before fees and zero after. The treated and control sample is similar along all dimensions except for assets under management. We find that the treated funds are larger than the control funds. Control variables and fixed effects are used to correct for these differences.

In columns 4 to 6, we report the numbers for the 50% limit. In this case, there are more closet index funds. From a pool of 122 funds, 75 are treated and 47 are controls. The average active share for the pool of funds is around 37%. Again, the treated and control sample are similar along all dimensions, now also for size. It is worth noting that compared to a typical U.S. fund, the European funds are small in terms of their management teams and organization scope but large relative to their investment space.

2.5 Impact assessment

We study the effect of the interventions in two "stages". In the first stage, we examine the impact of scrutiny on the level of activity and fees. We have already hypothesized that the funds will increase activity over reducing fees. For the effect on alpha in the second stage, we explore the consequences of the potential new behavior. Depending on the motives for closet indexing, described in Section 2.3.2, we can expect either a positive or a negative impact on alpha.

Our goal in describing the event is twofold. Firstly, we want to document what really happened to the outcome variables, i.e., active share, fee, and realized alpha. Secondly, we want to understand whether differences in realized outcomes result from noise or can be interpreted as "significant" differences. Therefore, we start by describing the realized effects and then "add on" additional statistical tests.

2.5.1 First stage: Impact on active share

In this section, we empirically test what happens to active share for funds under scrutiny. Figure 2.5.1 illustrates the development of the realized mean active share for treated and control funds over time. We clearly find that funds exposed to scrutiny increase active share more than unaffected funds. We also see that the two groups not only display parallel trends before the event but also the same level of active share.[24]

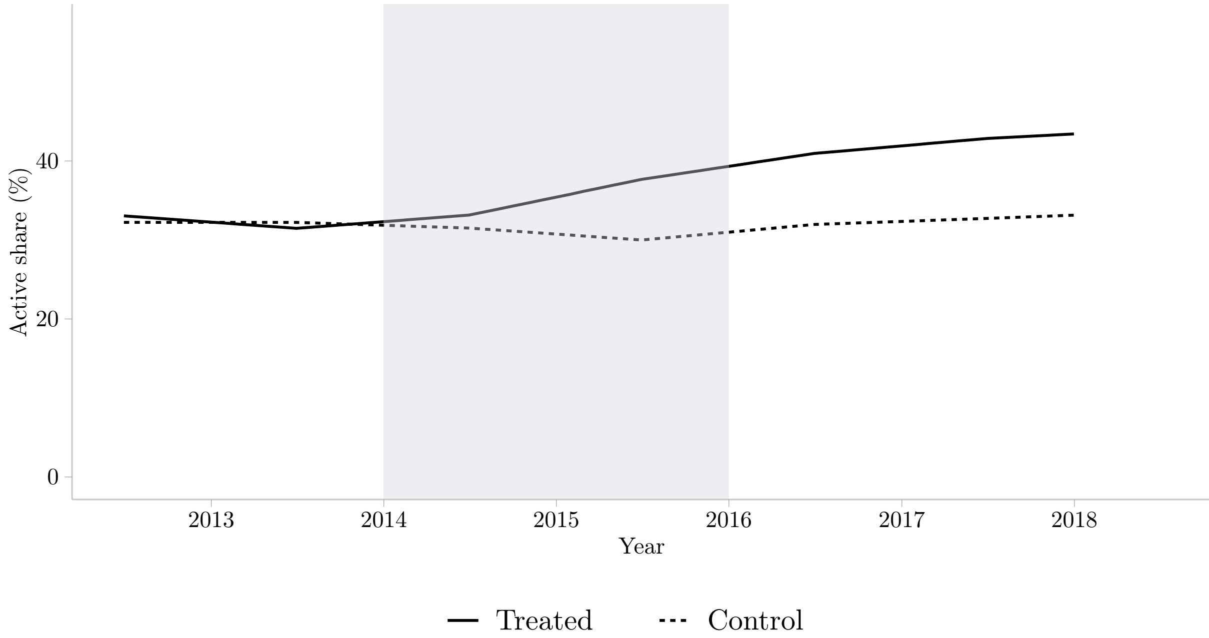
In Table 2.5.1, we report the realized differences for both the sample with active share below 40% and 50%. Results are reported using a two-year pre-and post-event window. For the below 40% sample, the difference between treated and control is 9.4% after the event. As shown in Figure 2.5.1, the funds' activity is similar before the event. Thus, the diff-in-diff estimate of 9.2%

[24] A prerequisite for a diff-in-diff analysis is a pre-trend evaluation. Both in the visual illustration and a formal test in Appendix 2.C.1, we document that the trends are parallel.

Figure 2.5.1.
Development of active share

This figure presents the development in active share for closet index funds[1] from 2012 through 2017. The time series are the annual averages of monthly cross-sectional, group-wise means, and the shaded area highlights the event window.

[1] Active share $\leq$ 40% in 2013.



is roughly the same as the post-event estimate. The increase is almost one third (9.24%/32.4%) relative to the pre-event level. For the sample of funds with active share below 50%, the post-event difference is 7.3%, and the diff-in-diff estimate is 5.7%. Thus, we find that scrutiny leads to a higher activity level, with the effect being larger for the most intensely scrutinized funds.

Regarding statistical inference, we also calculate the p-value for the different estimates in Table 2.5.1. In this first statistical analysis, we use the average difference between a treated and non-treated fund for the two years before and after the event. Using "collapsed" data, i.e., averages over time, we avoid any potential autocorrelation between the monthly observations (Bertrand et al., 2004). We find that all estimates have a p-value below 1%.

By splitting the sample into different groups based on fund characteristics, we examine whether the impact comes from certain funds. We divide the sample into a high or low category based on whether the fund is above or below the median value. The "high" treated funds are then tested against the "high" control funds, and similarly for "low" funds. The estimates from the triple diff-in-diff regressions are shown in Table 2.C.2 in Appendix 2.C. Overall, we find that

Table 2.5.1.
Effect of intervention on active share

This table reports the effects of policy interventions on **active share** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. Estimation is based on two-year pre- and post-event averages. The difference column presents post-event differences between treated and control funds. The difference-in-differences follows a classic setup, see Angrist and Pischke (2008) for details. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

	Active share	
Fund sample	Difference	Difference-in-differences
Active share $\leq$ 40%	9.41*** (2.50)	9.24*** (2.29)
Active share $\leq$ 50%	7.28*** (2.27)	5.70*** (2.05)

positive active share estimates are present for all fund categories: size, age, fee level, and past performance. However, the age characteristic is interesting; for both the 40% and 50% sample, the old funds respond significantly more than the young funds. This result may indicate that the older closet index funds are more concerned with scrutiny.

We also formally test the effect of scrutiny using fixed effects (fund and time) panel regressions to minimize endogeneity concerns, among other things. Fixed effects soak up any variation in active share due to cross-sectional differences in fund characteristics, meaning that identification comes from variation within a fund over time, and not from the cross-section of funds. These estimates are presented in Table 2.C.3 in Appendix 2.C. The coefficient sizes and significance levels are similar to the previous specifications, i.e., the estimates are large and significant.[25]

2.5.2 First stage: Impact on fee

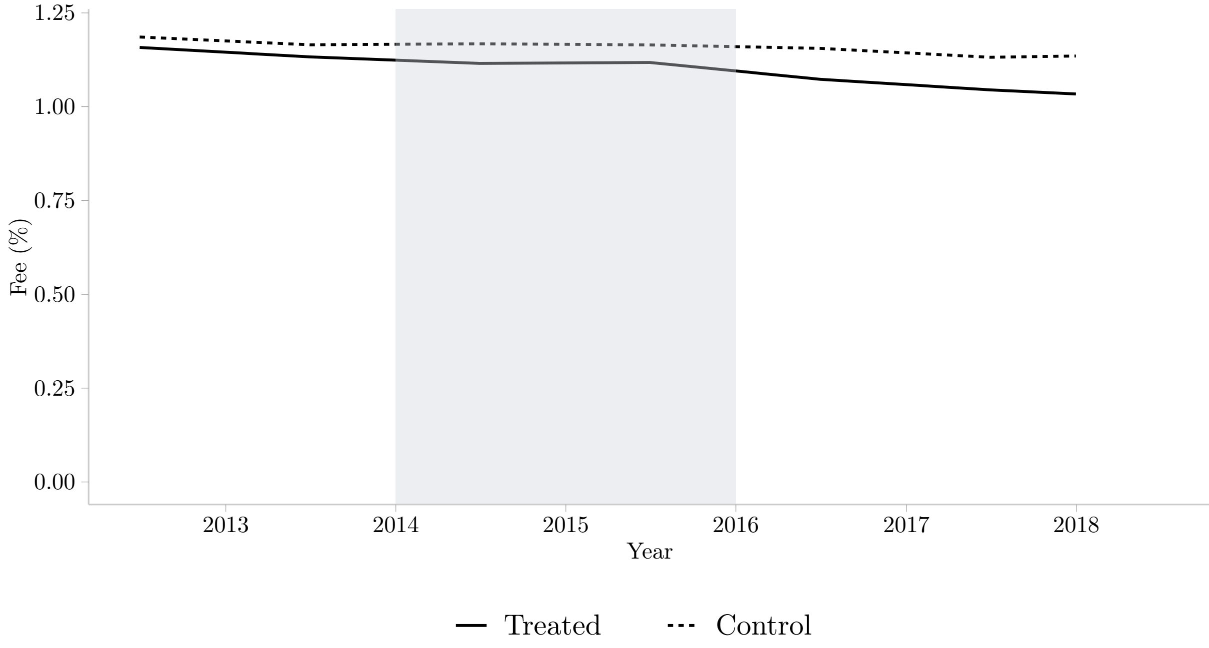
In this section, we empirically test the effect of scrutiny on fees. We have hypothesized that fund managers opt not to adjust fees. Again, we present two types of results, one documenting what happened and one where we statistically test for effects.

Figure 2.5.2 illustrates the development of the mean fee for treated and non-treated funds. For both groups, there is a decreasing trend, which is in line with the overall trend for active mutual funds (Morningstar, 2019). Intra-group discrepancies are minor and stable, meaning that funds under scrutiny did not reduce their fees more than unaffected funds. In Table 2.5.2, we report realized differences, given as annualized figures. The estimates show that scrutiny did

[25]The control variables are described in Appendix 2.B.4.

Figure 2.5.2.
Development of fees

This figure presents the development in fund fees for closet index funds[1] from 2012 through 2017. The time series are the annual averages of monthly cross-sectional, group-wise means, and the shaded area highlights the event window.
[1] Active share $\leq$ 40% in 2013.



not lead to a relatively lower fee level. We also check whether the p-value is below the regular level for statistical significance and find that none of the estimates are significantly different from zero.[26]

Table 2.D.2 in Appendix 2.D shows the sample split into different subsamples based on fund characteristics. We find no meaningful group-wise differences. If anything, scrutinized closet index funds with an active share below 50% and relatively low fees or age increase fees more than the control group. Table 2.D.3 in Appendix 2.D tests the effect of scrutiny using fixed effects (fund and time) panel regressions. Not surprisingly, the estimates are low and insignificantly different from zero.

To sum up for fees, every estimate is small and insignificant, thus confirming our hypothesis that managers under scrutiny choose to increase activity over reducing fees. This result supports the notion that the managers were "forced" to increase activity. The following section investigates the consequences of increased activity on performance.

[26]For a pre-trend analysis, see Appendix 2.D.1.

Table 2.5.2.
Effect of intervention on fees

This table reports the effects of policy interventions on **fees** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. Estimation is based on two-year pre- and post-event averages. The difference column presents post-event differences between treated and control funds. The difference-in-differences follows a classic setup, see Angrist and Pischke (2008) for details. Reported coefficients are annualized from fund-month level observations. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Fee	
Fund sample	Difference	Difference-in-differences
Active share ≤ 40%	−0.10 (0.12)	−0.07 (0.05)
Active share ≤ 50%	0.06 (0.12)	0.00 (0.04)

2.5.3 Second stage: Impact on alpha

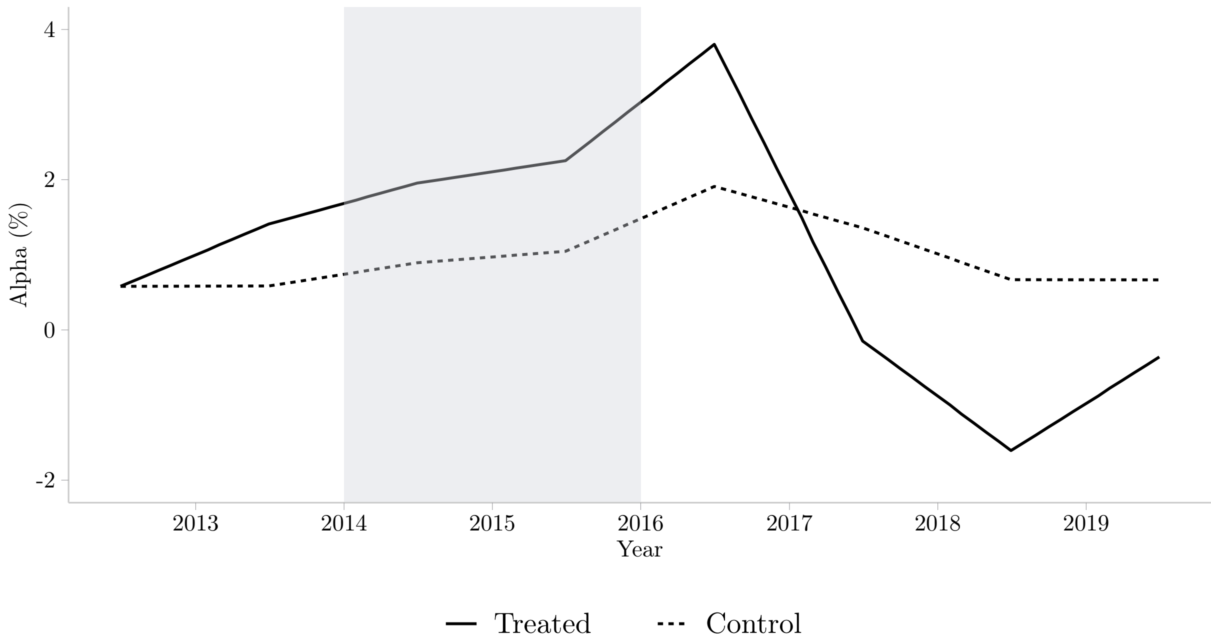
In this section, we study the total impact of policy scrutiny on fund performance. As mentioned above, arguments exist for both a positive and negative impact on performance. Our dependent variable is alpha, and, as demonstrated above, the changes in active share have already taken place going into the post-event period. To evaluate the impact, we use a four-year post-event period. Again, we first calculate the actual effect, i.e., how much return the group of scrutinized funds has generated relative to the group of unaffected funds, before formally testing whether this impact is statistically significantly different from zero.

Figure 2.5.3 illustrates the development of realized mean gross alpha for treated and non-treated closet index funds. In the pre-event period, treated funds appear to perform somewhat better than non-treated funds. However, this trend is insignificant, as reported in Table 2.E.1 in Appendix 2.E. After the intervention, we see a different pattern; treated funds underperform the control funds.

In Table 2.5.3, we report the realized differences. The annual difference between funds below 40% active share under scrutiny and not under scrutiny is 0.83% after the event. Since treated funds performed better before the event, the diff-in-diff estimate is 1.29%. One way of interpreting this result is that an investor that was long the average control fund and short the average treated fund before the intervention, and long the treated fund and short the control fund after the intervention would have lost 1.29% annually. Given the lower increase in active share for the 50% cutoff, we expect the results to be less negative. Surprisingly, we find similar

Figure 2.5.3.
Development of alpha

This figure presents the development in fund alpha for closet index funds[1] from 2012 through 2019. The time series are the annual averages of monthly cross-sectional, group-wise means, and the shaded area highlights the event window.
[1] Active share $\leq$ 40% in 2013.



point estimates for both samples.

Regarding statistical inference, there are many issues related to performance evaluation and there is no academic consensus on a particular method of testing performance (see, for example, Elton and Gruber (2020) for a recent review of the relevant literature). Therefore, we emphasize the importance of using alternative approaches.

Table 2.5.3.
Effect of intervention on alpha

This table reports the effects of policy interventions on **alpha** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. Estimation is based on two-year pre- and four-year post-event averages. The difference column presents post-event differences between treated and control funds. The difference-in-differences follows a classic setup, see Angrist and Pischke (2008) for details. Reported coefficients are annualized from fund-month level observations. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Alpha	
Fund sample	Difference	Difference-in-differences
Active share $\leq$ 40%	-0.83^{**} (0.34)	-1.29^{**} (0.58)
Active share $\leq$ 50%	-1.10^{***} (0.34)	-1.30^{***} (0.50)

Firstly, we calculate the p-value for the different estimates in Table 2.5.3, and find that they all have a p-value below 5%.

Secondly, we perform a placebo test. One potential concern is that the estimated impact of scrutiny is either a random effect or captures some spurious correlation(s) with omitted variables. If this were the case, we should obtain the same results independent of the assignment of treatment and control observations. We test this by means of a placebo test where we randomly assign funds to treatment and control groups, maintaining the same ratio of treated to non-treated funds as in the original sample (see Table 2.4.2).

Using these randomly assigned groups, we estimate the diff-in-diff model presented in Table 2.5.3. We repeat this exercise for 1,000 estimations and report the results in a histogram in Figure 2.E.1 in Appendix 2.E. We find a significantly negative effect (5% confidence level) for only 1 of the 1,000 trials (0.1%) when using 40% as the limit on active share. For the sample using 50% as a limit, the corresponding number of trials with a significantly negative estimate is 13 (1.3%). Thus, only 0.1% (40% limit) and 1.3% (50% limit) of the estimated coefficients are equal to or smaller than the coefficient estimated using the original sample, represented by dashed vertical lines in Figure 2.E.1. These results reassure us that our tests capture the treatment effect of regulatory scrutiny on fund alpha and not some random effect or omitted variable.

Thirdly, we split the sample into different groups based on fund characteristics. We examine whether the impact stems from certain types of funds. The triple diff-in-diff regression estimates are shown in Table 2.E.2 in Appendix 2.E. Overall, estimates are negative for all fund categories: size, age, fee level, and past performance. However, they do not all have a p-value below 5%. As we split the sample however, the number of observations decreases and the variance of the estimator increases, making it harder to achieve statistical significance.

Fourthly, in Table 2.5.4 we perform fixed effects (fund and time) panel regressions. The fund fixed effects soak up any variation in gross alpha due to cross-sectional differences in fund characteristics, and time fixed effects remove common variation related to time. We estimate the difference in alpha between funds exposed and not exposed to scrutiny. In the first two columns, we do not add controls, while in 3 and 4, we control for fund age, size, and fees. We find that the estimates are about the same as using standard diff-in-diff regressions.

Table 2.5.4.
Alpha-scrutiny relationship

This table reports estimated slope coefficients (Post × Scrutiny) from panel regressions testing the effects of policy interventions on **alpha** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All regressions include fund and month fixed effects, and columns 3 and 4 additionally control for fund age, size, and fees. Reported coefficients are annualized from fund-month level observations. Standard errors (in parentheses) clustered by fund are reported in columns 1 and 3 and independently by fund and month in columns 2 and 4. There are 5,492 and 8,410 fund-month observations in the 40% and 50% panels, respectively. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Post × Scrutiny			
Fund sample	(1)	(2)	(3)	(4)
Active share ≤ 40%	−1.13**	−1.13	−0.93*	−0.93
	(0.49)	(1.16)	(0.51)	(1.23)
Active share ≤ 50%	−1.12**	−1.12	−1.07**	−1.07
	(0.43)	(1.31)	(0.45)	(1.36)
Controls			×	×
Fund cluster	×	×	×	×
Month cluster		×		×
Adj. R^2 [40% | 50%]	[0.10 | 0.13]		[0.10 | 0.13]	

Regarding p-values, there are differences between the specifications in Table 2.5.4. In all specifications, standard errors are clustered along the fund dimension, however in columns 2 and 4 we independently cluster along the time dimension as well. This two-way clustering sharply increases the estimated standard errors. This is probably due to an intra-time dependence structure in mutual fund returns that is present even after time fixed effects are removed.[27]

As a final robustness test, we use two alternative measures of value creation, i.e., net alpha and factor-adjusted alpha. So far, we have assumed that the effect on net and gross alpha is the same. This assumption is based on documented small fee effects. Columns 1 and 2 in Table 2.E.3 in Appendix 2.E confirm our assumption that the results also hold when net alpha is the dependent variable.

Limiting our scope to comparing European domestic funds means we do not expect any funds to have a particular factor style, since the high correlation between active share and tracking error indicates a weak factor structure. Funds with a factor style often have a broader investment universe than a single European country. Consequently, our focus has been on adjusting for risk prescribed by the benchmark. However, as a robustness test, we adjust alpha for domestic factor risk and perform the same regression analyses as in the other tables. Columns 3 and 4 in Table 2.E.3 show that the estimates have the same sign but slightly lower magnitude after adjusting for factor exposure. Thus, the negative alphas are caused by a combination of poor stock picking and factor exposure.

To sum up, our analyses show that scrutinized closet index funds underperform non-scrutinized closet index funds. These findings relate to our suggested motives for closet indexing in that they support the notion that the manager runs out of ideas for new successful bets. Thus, regulation then forces managers to take bets in a sub-optimal manner, leading them to destroy investor value. Taking this into consideration, the best alternative for investors, conditional on supervisory intervention, is reduced fees.

2.6 Further analyses

Thus far, we have shown that scrutinized closet index funds increased activity and achieved lower alphas than non-scrutinized funds. In Section 2.6.1, we decompose the total active portfolios into sub-portfolios. Furthermore, we examine the source of increased active share and how

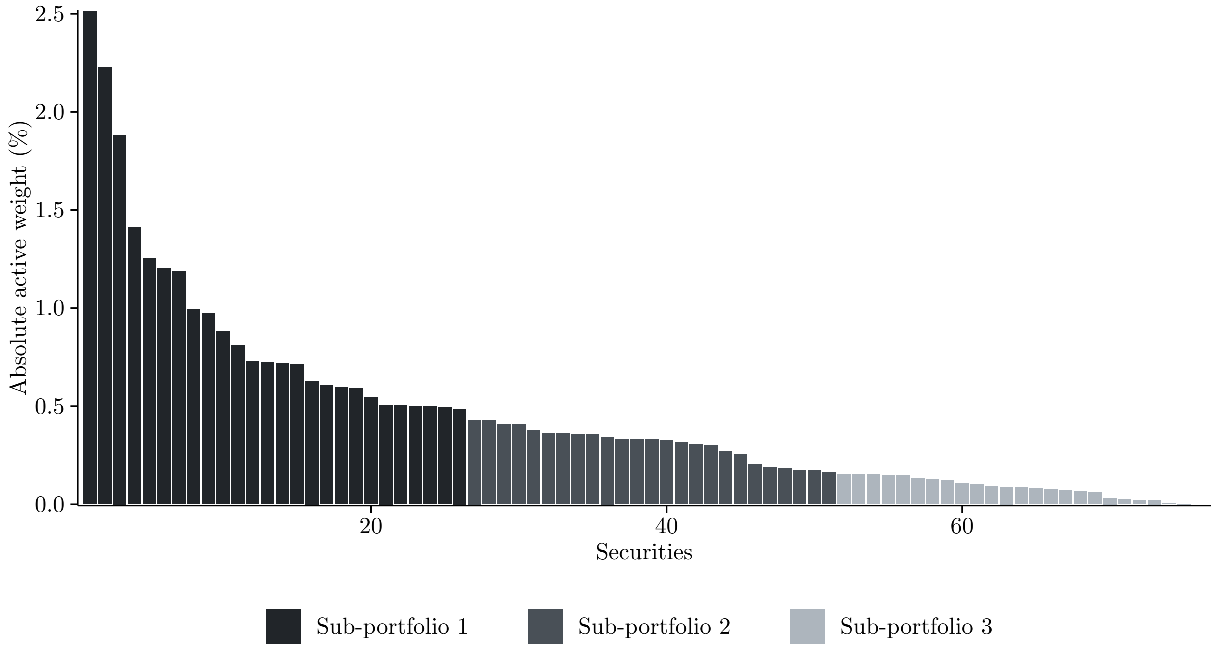
[27] Accurate standard errors are a fundamental component of statistical inference, but this issue seems to be in development, and it is not entirely clear what the best solutions are regarding fund returns. For example, see the recent well-published papers Pástor et al. (2015), Pástor and Vorsatz (2020), and Cremers et al. (2021).

this impacts performance. We argue that this detailed analysis can also be viewed as a test of diseconomies of scale in active management. Finally, in Section 2.6.2, we introduce an additional activity measure to determine whether the change in active share is a result of true active management or just signal jamming.

2.6.1 Decomposing the active portfolio

The nature of our data allows us to split the total active portfolio into subgroups, which is done based on active position size. This gives us three sub-portfolios per fund containing the largest, middle, and smallest positions based on absolute active weights. Inspired by Pomorski (2009) and Antón et al. (2021), we label the portfolio containing the largest bets the "best ideas" portfolio. To illustrate this categorization, Figure 2.6.1 shows how we allocate all the bets to a corresponding sub-portfolio based on all active weights for DNB Norge at the end of December 2013 (end of pre-event period).

Figure 2.6.1.
Active weights for DNB Norge in December 2013
This figure presents the active weights for DNB Norge in December 2013. Stocks with the largest absolute active weights are in portfolio 1 (best ideas) and those with the smallest absolute active weights are in portfolio 3.



We also identify a sub-portfolio including all the new bets obtained during the policy scrutiny period, which we label "new bets". These bets can come from existing stocks without an initial bet or from newly listed stocks.

In Table 2.6.1, we present our findings for the four sub-portfolios. Regarding changes in active share, we find that the difference between treated and control is positive for both the new and the smallest stock portfolio.[28] This applies to both the group of funds with an active share below 40% and 50%.[29]

Table 2.6.1.
Effect of intervention on sub-portfolios

This table reports the effects of policy interventions on sub-portfolio **active share** and **alpha** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All results are from a difference-in-differences model equivalent to the one presented in column 2 in Tables 2.5.1, 2.5.2, and 2.5.3. For active share, estimates are based on two-year pre- and post-event averages, while alpha estimates are based on two-year pre-event and four-year post-event averages. Reported coefficients for alpha are annualized from fund-month level observations. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Active share				Alpha			
Fund sample	New stocks	SP 1	SP 2	SP 3	New stocks	SP 1	SP 2	SP 3
Active share ≤ 40%	3.28***	0.60	0.30	0.72**	−0.44***	0.17	−0.18	0.16
	(0.44)	(1.35)	(0.48)	(0.35)	(0.13)	(0.34)	(0.16)	(0.12)
Active share ≤ 50%	3.10***	−0.70	−0.08	0.83***	−0.55**	0.25	−0.34**	−0.12
	(0.40)	(1.05)	(0.43)	(0.30)	(0.23)	(0.38)	(0.15)	(0.33)

Furthermore, we find that new bets taken by scrutinized funds underperform those taken by non scrutinized funds. This is contingent on the main source of scrutinized funds' portfolio underperformance set out in Section 2.5.3 stemming from these new bets.

In total, relative to active managers not scrutinized, treated funds generate roughly the same alpha on their best ideas and bets that they are already familiar with but lose out on new stocks. We can only speculate, but one potential explanation is that closet index managers are uncomfortable with being forced to increase activity. This result underpins a lack of new quality bets being the motive for closet indexing.

[28] Note that the sum of the characteristics we measure for the sub-portfolios is not identical to the total portfolio. This is because our definition of the sub-portfolios is limited to capturing the changes in stocks listed in December 2013 and does not capture all newly listed companies after the event window.

[29] All the funds follow the UCITS directives. The best known restriction is the so-called "5/10/40 rule". In summary, it sets out that a maximum of 10% of a UCITS fund's net assets may be invested in stocks from a single issuer and that investments of more than 5% with a single issuer may not make up more than 40% of the whole portfolio. One could argue due to this restriction; treated funds cannot increase their already largest bets. However, this is not likely since the restriction should also be binding for control funds.

Diseconomies of scale in active management?

One constraint discussed in the active management literature is diseconomies of scale. If scale impacts performance, skill and scale interact. A more skilled large fund can in such instance underperform a less skilled small fund. Consequently, we must understand scale effects to learn about skill.

Under certain assumptions, the intervention we examine enables us to test diseconomies of scale. If we assume that asset managers are forced to increase their level of activity and cannot use in- or outflows to change the degree of active management, we get a quasi-natural experiment that allows us to identify the scale-performance relationship. Given the diseconomies of scale predictions, described in Section 2.2, and the findings so far, we have indications of this type of dynamics.

2.6.2 Signal jamming?

As described in Section 2.2, one potential concern is that the change in activity does not come from real bets but from managers adding noise to their returns through signal jamming (Doshi et al. (2015) and Brown and Davies (2017)). A problem with the activity measures used so far is that they cannot detect whether funds engage in signal jamming to appear truly active. Therefore, we introduce a measure of stock-level portfolio concentration. This measure is used at the stock-level by Brands et al. (2005) and Kacperczyk et al. (2014), and at the industry-level by Kacperczyk et al. (2005). The portfolio concentration measure is calculated by summarizing the squared active weights, meaning that the measure places more weight on larger active positions.[30]

By comparing the development in portfolio concentration for treated and control closet index funds, we learn how the managers implement the activity change. If they spread the bets (signal jamming), we would expect no difference in portfolio concentration. Conversely, if more concentrated bets are taken, we expect the diff-in-diff estimate to be positive.

We use the same methodology to describe the development of the active portfolio concentration, as for testing the change in active share. Regression results are presented in Table 2.6.2. If we assume the active portfolios for treatment and control funds are equal before the event, we can see from column 1 that the treated post-event portfolios consist of more concentrated

[30] Portfolio concentration is defined as $\sum_{i=1}^{N}(w_i^a)^2$, i.e., the Herfindahl-Hirschman index of the portfolio's active weights.

active bets. The difference is only statistically significant for the sample with a cutoff at 40%. The diff-in-diff analysis yields the same results with smaller estimates. Similar to Kacperczyk et al. (2005), it is harder to use portfolio concentration to interpret economic significance than when using active share. In Appendix 2.F, we perform fixed effect regressions and obtain the same results.

Table 2.6.2.
Effect of intervention on portfolio concentration

This table reports the effects of policy interventions on **portfolio concentration** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. Estimation is based on two-year pre- and post-event averages. The difference column presents post-event differences between treated and control funds. The difference-in-differences follows a classic setup, see Angrist and Pischke (2008) for details. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Portfolio concentration	
Fund sample	Difference	Difference-in-differences
Active share $\leq$ 40%	1.16*** (0.26)	0.62*** (0.22)
Active share $\leq$ 50%	0.46 (0.51)	0.33 (0.26)

In Table 2.6.3, we perform similar analyses to investigate different parts of the active portfolios. We find that treated managers take more concentrated bets in stocks not previously part of the active portfolio for the cutoff at 40% and 50%. Thus, we find that scrutiny makes funds take more concentrated bets. A significant part of these bets is concentrated in the "new bets" part of the active portfolio.

Table 2.6.3.
Effect of intervention on sub-portfolio concentration

This table reports the effects of policy interventions on **sub-portfolio concentration** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All results are from a difference-in-differences model equivalent to the one presented in column 2 in Tables 2.5.1, 2.5.2, and 2.5.3. Estimation is based on two-year pre- and post-event averages. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p < 1\,\%$, $^{**}p < 5\,\%$, $^{*}p < 10\,\%$.

	Portfolio concentration			
Fund sample	New stocks	SP 1	SP 2	SP 3
Active share $\leq$ 40%	0.10***	−0.04	0.07**	0.04*
	(0.03)	(0.26)	(0.03)	(0.02)
Active share $\leq$ 50%	0.10***	−0.41*	0.04	0.04*
	(0.02)	(0.21)	(0.03)	(0.02)

2.7 Policy implication and conclusion

The current trend is that people are increasingly becoming their own money managers. The landscape in which they make decisions is becoming more and more complicated, and many often lack the information and knowledge they need to judge the quality of the products they purchase. This information asymmetry creates incentives for opportunistic behavior on the part of service providers, leading to market failures. The existence of closet indexing can be viewed as a realistic example of this failure. If investors were able to evaluate the quality of their fund managers' services, they would not choose to pay high fees for closet indexing. Clearly, financial authorities need to align the interests of financial intermediates and investors.

In this paper, we have attempted to identify what can be learned from the interventions in the Scandinavian countries. We highlight the following four results. Firstly, a potential closet index fund opts to increase activity and not reduce fees when placed under suspicion by the supervisory authorities. Secondly, when closet index funds are forced to become more active, they perform worse than unaffected comparable funds. Thirdly, the funds under scrutiny take positions in new stocks when they increase activity. Finally, the funds under scrutiny lose out on these new marginal bets relative to unaffected funds, which lends support to the hypothesis on diseconomies of scale in active management.

Regulators face two principal alternatives. They can either force funds to increase the active management level or reduce fees. So far, little research has investigated the effect of this intervention. We show that when given the choice, the funds increase their level of activity. We propose two different rationales for this behavior; managers are either skilled and in possession of new ideas but are afraid of losing assets under management or they lack ideas and want to harvest as much revenue as possible. Activity-increasing managers' failure to create value for their investors supports the narrative of revenue harvesting managers without a plethora of new, good ideas.

What can regulators thus learn from our findings? One of the most important motives for closet indexing is managers' lack of additional investment ideas, and fear of losing revenue by revealing that they are not as skilled as other fund managers. These funds should therefore not be forced to increase activity, but rather update their investor information and reduce the fees they charge. Regulatory authorities should be cautious about forcing potential closet index funds to become more active. However, this is not that simple in practice because a regulator cannot stop managers from increasing activity and thus justifying their high fees. If managers, who are not adequately skilled, choose to increase their activity, poor results inevitably lead to lower assets under management. In the most severe cases, these funds will be forced out of business. However, this process is slow and uninformed investors will sustain major losses on the way.

In sum, a regulator should not force funds to become more active. Instead, their role should be to identify closet index funds. If funds want to maintain a low degree of activity, this should be reflected in the fee levels and information set. If the trend of investigating closet index funds continues, the hope is that this will discourage new funds from choosing such a strategy. In this context, the DNB case from Norway illustrates that legal settlements can also be obtained. Intermediaries and information providers such as Morningstar should pay special attention to the performance of closet index funds that want to become a truly active fund. This point also illustrates the benefit of having independent fund providers (Stoughton et al. (2011)) and the problem with "own brand" funds (Jenkinson et al. (2020)). Naturally, information on active share and tracking error should also be available to investors.

Mutual funds manage trillions of dollars. The funds' investment decisions determine where a significant proportion of capital is allocated in the economy (Pástor et al. (2020)). A recent trend sees more capital being allocated away from active funds and into index funds (Investment Company Institute (2019)).[31] Eventually, the flow of funds from active to passive has to stop. The cost of active funds will decrease, making active funds more attractive relative to passive funds. However, we always need to pay attention to the content of the active funds. The debate on closet indexing gained momentum after the paper by Cremers and Petajisto (2009) and is still ongoing. The ruling against DNB in the Norwegian Supreme Court in February 2020 is a milestone in this debate. To our knowledge, this is the first time sanctions were imposed by the courts. Only the future can tell if and how the closet index problem will be resolved.

[31] There are many reasons for this development. One is the poor historical performance of actively managed funds, another is increased competition between funds due to new technology. A final one may be what we examine in this paper, the difference between what some funds promise and what they deliver.

References

Adams, J. C., Hayunga, D. K., and Mansi, S. (2021). Scale and performance in active management are not negatively related. *Critical Finance Review, forthcoming*.

Agarwal, V., Gay, G. D., and Ling, L. (2014). Window dressing in mutual funds. *The Review of Financial Studies*, 27(11):3133–3170.

Agnew, J., Balduzzi, P., and Sunden, A. (2003). Portfolio choice and trading in a large 401 (k) plan. *American Economic Review*, 93(1):193–215.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics*. Princeton university press.

Antón, M., Cohen, R. B., and Polk, C. (2021). Best ideas. *Available at `https://papers.ssrn.com/sol3/papers.cfm?abid=1364827`*.

Banegas, A., Gillen, B., Timmermann, A., and Wermers, R. (2013). The cross section of conditional mutual fund performance in european stock markets. *Journal of Financial economics*, 108(3):699–726.

Berk, J. B. and Green, R. C. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295.

Berk, J. B. and van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(1):1–20.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Bilias, Y., Georgarakos, D., and Haliassos, M. (2010). Portfolio inertia and stock market fluctuations. *Journal of Money, Credit and Banking*, 42(4):715–742.

Brands, S., Brown, S. J., and Gallagher, D. R. (2005). Portfolio concentration and investment manager performance. *International Review of Finance*, 5(3-4):149–174.

Brown, D. C. and Davies, S. W. (2017). Moral hazard in active asset management. *Journal of Financial Economics*, 125(2):311–325.

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

Chen, J., Hong, H., Huang, M., and Kubik, J. D. (2004). Does fund size erode mutual fund performance? the role of liquidity and organization. *American Economic Review*, 94(5):1276–1302.

Cremers, K. and Quinn, C. (2016). Do mutual fund investors get what they pay for: Securities law and closet index funds. *Virginia Law and Business Review*, 11:31.

Cremers, K. M., Fulkerson, J. A., and Riley, T. B. (2021). Benchmark discrepancies and mutual fund performance evaluation. *Journal of Financial and Quantitative Analysis*, pages 1–29.

Cremers, K. M., Fulkerson, J. A., and Riley, CFA, T. B. (2022). Active share and the predictability of the performance of separate accounts. *Financial Analysts Journal*, pages 1–19.

Cremers, K. M. and Petajisto, A. (2009). How active is your fund manager? a new measure that predicts performance. *Review of Financial Studies*, 22(9):3329–3365.

Cremers, M., Ferreira, M. A., Matos, P., and Starks, L. (2016). Indexing and active fund management: International evidence. *Journal of Financial Economics*, 120(3):539–560.

Doshi, H., Elkamhi, R., and Simutin, M. (2015). Managerial activeness and mutual fund performance. *The Review of Asset Pricing Studies*, 5(2):156–184.

Edelen, R. M., Evans, R. B., and Kadlec, G. B. (2007). Scale effects in mutual fund performance: The role of trading costs. *Available at* *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=951367*.

Elton, E. J. and Gruber, M. J. (2020). A review of the performance measurement of long-term mutual funds. *Financial Analysts Journal*, 76(3):22–37.

ESMA (2016). Supervisory work on potential closet index tracking. Available at https://www.esma.europa.eu/document/esma-updates-supervisory-work-closet-index-tracking.

ESMA (2020). Closet indexing indicators and investor outcomes. Available at https://www.esma.europa.eu/sites/default/files/library/esmawp-2020-2_closet_indexing.pdf.

Ferreira, M. A., Keswani, A., Miguel, A. F., and Ramos, S. B. (2013). The determinants of mutual fund performance: A cross-country study. *Review of Finance*, 17(2):483–525.

Ferreira, M. A., Matos, P., and Pires, P. (2018). Asset management within commercial banking groups: International evidence. *Journal of Finance*, 73(5):2181–2227.

Financial Conduct Authority (2016). Economics for Effective Regulation. Available at https://www.fca.org.uk/publication/occasional-papers/occassional-paper-13.pdf, Last accessed on 2020-09-10.

Financial Conduct Authority (2019). Closet trackers. Available at https://www.fca.org.uk/firms/authorised-and-recognised-funds/closet-trackers,.

Financial Supervisory Authority of Denmark (2013). Market developments 2013. Available at https://www.dfsa.dk/~/media/Tal-og-fakta/2014/Markedsudvikling/Market-developments-2013-Investment-associations.pdf?la=en, Last accessed on 2020-09-10.

Financial Supervisory Authority of Sweden (2015). Consumer protection on the financial market. Available at https://www.fi.se/contentassets/9791b37d1d074ed4a3fc193786d449f6/konsumentrapp_2015engny.pdf, Last accessed on 2020-09-10.

Huang, J., Sialm, C., and Zhang, H. (2011). Risk shifting and mutual fund performance. *The Review of Financial Studies*, 24(8):2575–2616.

Investment Company Institute (2019). A Review of Trends and Activities in the Investment Company Industry. Available at https://ici.org/pdf/2019_factbook.pdf, Last accessed on 2020-09-10.

Jenkinson, T., Martinez, J., Cookson, G., and Jones, H. (2020). Best buys and own brands: Investment platforms' recommendations of mutual funds. *Review of Financial Studies*.

Kacperczyk, M., Nieuwerburgh, S. V., and Veldkamp, L. (2014). Time-varying fund manager skill. *The Journal of Finance*, 69(4):1455–1484.

Kacperczyk, M., Sialm, C., and Zheng, L. (2005). On the industry concentration of actively managed equity mutual funds. *The Journal of Finance*, 60(4):1983–2011.

Kjørven, M. E. (2019). Closet index funds and retail investor protection - a scandinavian perspective. *Tilburg Law Review*, 24(1):125–138.

Leippold, M. and Rueegg, R. (2020). How rational and competitive is the market for mutual funds? *Review of Finance*, 24(3):579–613.

McLemore, P. (2019). Do mutual funds have decreasing returns to scale? evidence from fund mergers. *Journal of Financial and Quantitative Analysis*, 54(4):1683–1711.

Morningstar (2019). U.S. Fund Fee Study. Available at `https://www.morningstar.com/articles/925303/2018-morningstar-fee-study-finds-that-fund-prices-continue-to-decline`.

New York Office of the Attorney General (2018). Mutual Fund Fees and Active Share. Available at `https://ag.ny.gov/sites/default/files/ny_ag_report_on_mutual_fund_fees_and_active_share.pdf`, Last accessed on 2020-09-10.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Norwegian Consumer Council (2015). Consumer Council sues Norway's largest bank. Available at `https://www.forbrukerradet.no/side/consumer-council-sues-norways-largest-bank-on-behalf-of-180000-consumers/`, Last accessed on 2020-09-10.

Norwegian Ministry of Finance (2015). Financial Markets Report 2015. Available at `https://www.regjeringen.no/contentassets/fcbb121a9cb342f0b52d44300d3cc3b0/no/pdfs/stm201520160029000dddpdfs.pdf`, Last accessed on 2020-09-10.

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

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2020). Fund tradeoffs. *Journal of Financial Economics*, 138(3):614–634.

Pastor, L., Stambaugh, R. F., Taylor, L. A., and Zhu, M. (2021). Diseconomies of scale in active management: Robust evidence.

Pástor, L. and Vorsatz, M. B. (2020). Mutual fund performance and flows during the covid-19 crisis. *The Review of Asset Pricing Studies*, 10(4):791–833.

Petajisto, A. (2013). Active share and mutual fund performance. *Financial Analysts Journal*, 69(4):73–93.

Phillips, B., Pukthuanthong, K., and Rau, P. R. (2018). Size does not matter: Diseconomies of scale in the mutual fund industry revisited. *Journal of Banking & Finance*, 88:357–365.

Pollet, J. M. and Wilson, M. (2008). How does size affect mutual fund behavior? *Journal of Finance*, 63(6):2941–2969.

Pomorski, L. (2009). Acting on the most valuable information:'best idea' trades of mutual fund managers. *Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1108186`*.

Reuter, J. and Zitzewitz, E. (2021). How much does size erode mutual fund performance? a regression discontinuity approach. *Review of Finance*, 25(5):1395–1432.

Sensoy, B. A. (2009). Performance evaluation and self-designated benchmark indexes in the mutual fund industry. *Journal of Financial Economics*, 92(1):25–39.

Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *Journal of Finance*, 53(5):1589–1622.

Stoughton, N., Wu, Y., and Zechner, J. (2011). Intermediated investment management. *Journal of Finance*, 116(3):947–980.

Sunstein, C. R. (2015). Financial regulation and cost-benefit analysis. *The Yale Law Journal Forum*, 124:263–279.

Swedish Ministry of Finance (2016). SOU 2016:45, En hållbar, transparent och konkurrenskraftig fondmarknad. Available at `https://www.regeringen.se/49f197/contentassets/c1e668bf3373497d88e7b0ef014d9f2a/en-hallbar-transparent-och-konkurrenskraftig-fondmarknad-sou-201645`, Last accessed on 2020-09-10.

Yan, X. S. (2008). Liquidity, investment style, and the relation between fund size and fund performance. *Journal of Financial and Quantitative Analysis*, 43(3):741–767.

Zhu, M. (2018). Informative fund size, managerial skill, and investor rationality. *Journal of Financial Economics*, 130:114–134.

2.A Appendix describing the DNB Norge case

On June 21, 2016, the Norwegian Consumer Council instituted legal proceedings before Oslo District Court against DNB Asset Management AS (henceforth DNB), a wholly-owned subsidiary of DNB ASA providing asset management services. The Norwegian Consumer Council instituted a group action to pursue compensation on behalf of 180,000 investors in DNB Norge, a fund managed by DNB. The lawsuit alleged that the investor information and the high fees charged gave the investors reason to expect active management. In contrast, the funds were in fact managed very close to the index.

Oslo District Court passed its judgment on January 12, 2018, whereby the claim was rejected, and DNB was held not liable. On February 12, 2018, the Norwegian Consumer Council appealed the judgment to Borgarting Court of Appeal. The Court of Appeal's ruling was announced on May 8, 2019 and ruled in favor of the Norwegian Consumer Council in the group action.[32]

In short, the Court of Appeal describes the effort of active management as (i) performing analyses to identify potential good bets; and (ii) translating this into active positions such that the fund deviates from the index to a not insignificant extent. By comparing the investor information and the high fee on the one hand and the fact that the funds were managed very close to the index, the Court concluded that DNB had violated its obligations to investors. DNB was sentenced to pay approximately NOK 350 mill (approx. USD 35 mill).[33]

DNB appealed the case to the Norwegian Supreme Court. The appeal case started on January 21, 2020. The ruling was delivered on February 27, upholding the Court of Appeal's ruling.[34][35][36]

[32] Available at Lovdata TOSLO-2016-105341-2.
[33] Lovdata LB-2018-43087.
[34] Better Finance press release (2020).
[35] Norwegian Consumer Council press release.
[36] Lovdata HR-2020-475-A.

2.B Appendix describing the Data

2.B.1 Sample selection and raw fund data

In this section, we present details of the sample selection and construction. Next, we present information on the fund data. This section is intended to help replicate the paper's results.

Sample selection

The initial sample is constructed from lists generated in Morningstar Direct. As explained in the main part, and shown in Table 2.4.1, the treated countries are Denmark, Norway, and Sweden. These are the countries where the financial authorities have applied the most intense scrutiny. The initial sample of European countries are those where we have been unable to identify any scrutiny by the FSAs: Austria, Belgium, Finland, Poland, Portugal, and Switzerland. For each of these countries, we construct lists based on the fields *Global Broad Category Group*, *Investment Area*, *Firm Country*, and *Base Currency*. We set the field *Global Broad Category Group* equal to equity to extract equity-only funds. Next, we filter by *Firm Country* equal to *Investment Area* for funds with *Firm Country* for the countries included in the study, to obtain domestic funds. Last, we set the *Base Currency* equal to the domestic currency in each country. Moreover, a large part of the funds is structured with multiple share classes. We use the field *Oldest Share Class*, which takes the values of either Yes or No, to filter out the main share class of each fund. The initial sample consists of 1,148 funds, as presented in Table 2.B.1.

Next, we impose three additional filters based on the fund type. As this study's scope is to interpret the portfolios of actively managed funds, we require the funds to be active, i.e., manage a portfolio where the objective is to outperform a passive benchmark index and have a managed portfolio. For this we use the fields: *Index Fund*, *Enhanced Index Fund*, and *Fund of Funds*. These three fields take the values Yes or No, and we set all these parameters to No. For robustness purposes, we cross-check the fields from Morningstar with the Lipper database and find that our initial sample selection is not free of errors. Despite having removed index and enhanced index funds from the sample before matching with Lipper, there are still three Swiss funds flagged by Lipper as index funds. As the two data providers categorize the funds differently, we manually check the funds' investment objectives to determine which category is most appropriate. All of them state directly in the investment objective that they are either an index fund or replicate their benchmark index using either the physical or synthetic method.

Table 2.B.1.
Sample selection of domestic equity mutual funds

This table presents the outcome of our sample selection procedure. The number of funds at the initial step are those where the management company is located in the same geographic area as they invest in. At the fund type step, we exclude all the funds that are registered as either an index fund, enhanced index fund, or a fund of funds. To draw meaningful inference, we require funds to be alive one year before and after the event. Thus we exclude all funds that have an inception date after January 31, 2013 or an obsolete date before December 31, 2016 in the alive during event step. In order to form treatment and control groups, we need data on active share before and after the event. Finally, we require funds to have data on key variables such as returns, size etc. during the event, and thus exclude funds that lack observations over the two-year event period.

Step	Total	Treated	Control
Initial	1,148	624	524
Fund type	960	522	438
Alive during event	378	177	201
Data coverage	353	156	197
Total sample	**353**	**156**	**197**
Active share $\leq$ 50%	122	47	75
Active share $\leq$ 40%	79	33	46

After excluding funds based on fund type, we are left with a sample consisting of 960 funds in total.

The final requirement is that the funds have sufficient data before and after the event window, of January 2014 until December 2015. To draw meaningful inference, we require that each fund has data starting, at the latest, one year prior to and ending, at the earliest, one year after the event. We use the fields *Inception Date* and *Obsolete Date* to filter out funds. This means that funds with an inception date after January 2013 or an obsolete date before December 2016 are filtered out of the initial sample. This leaves us with a sample of 378 domestic actively managed equity mutual funds in our initial sample from which to draw treatment and control funds based on active share, with 177 potential treated funds and 201 potential control funds.

Fund data

After defining the initial sample, we collect fund returns, fund size, and portfolio holdings. The main source of the time series fund data is Morningstar Direct, while we use both the Morningstar and Lipper databases for the fund portfolio data. For each constituent in the lists explained in the previous section, we download the variables *Monthly Return*, *Monthly Gross Return*, and *Monthly Fund Size* aggregated over share classes.

For returns, all income and capital gains are reinvested monthly. The returns data is in the local currency, while assets under management are in USD to establish a common currency for comparison across countries. The Monthly Return includes management, administrative, and other costs that are deducted from the NAV, and gross returns are returns before fees. Thus, we use these two variables to compute the expense ratios in accordance with Morningstar Direct definition.

For the portfolio data, we use both the Morningstar and Lipper databases. However, we find that some of the other European countries' funds lack portfolio data in the Lipper database. For these funds, we download the portfolios from Morningstar to complete the data. We match the Morningstar (fund characteristics, performance, and portfolios) and Lipper data (fund portfolios) by ISIN or fund name if ISIN is missing. We end up with a link list between the two databases with ISIN, fund names, Lipper IDs (Lipper's internal fund identifier), and Sec ID (Morningstar's internal fund identifier).

2.B.2 Benchmark data

To measure active share and compare the fund returns to the returns of a benchmark, we must determine a benchmark index to evaluate the fund portfolios and performance against. We use Datastream to download the constituents and benchmark weights, as well as the benchmark returns. We use the primary prospectus benchmark from Lipper if available. For funds where the primary prospectus benchmark constituents are unavailable in Datastream, we choose to use the most common domestic benchmark within each country for that particular fund type. Moreover, for some of the indices, we cannot obtain the actual index weights from Datastream and use value-weighted weights based on market capitalization for the constituents. We match the benchmark portfolios with the fund portfolios based on stock ISIN. After downloading data for the initial fund sample, some funds lack either fund or portfolio data. After excluding funds with missing data, we end up with a sample consisting of 353 funds in total, where 156 are potential treated funds, and 197 are potential control funds.

The last row in Table 2.B.1 reports the final sample. The treated funds are funds from Scandinavia, and the control funds are funds from other European countries. We also show how many funds are closet index funds based on a limit of an active share of either 50% or 40%.

2.B.3 Sample

Table 2.B.2 reports summary statistics for all the domestic European active funds in the sample. The table presents the base sample of funds, where we form treatment groups in the main tests based on active share levels.

Table 2.B.2.
Summary statistics: equity mutual funds

This table presents summary statistics for the full sample of actively managed domestic equity mutual funds in Europe. Values are means over a two-year window before the event start. Gross alpha, expense ratio, and net alpha are annualized. Competition is defined as 1 − Herfindahl-Hirschman index (HHI), and displayed values are country-means for the Scandinavian countries and the rest of Europe, respectively. Scandinavian funds are labeled treated and the rest of Europe control. Inference on differences between treatment and control funds is based on Newey and West (1987) standard errors. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

Sample	Full sample	Treated	Control	Difference
Number of funds	353	156	197	
Active share (%)	55.0	50.9	59.0	-8.1^{***}
Gross alpha (%)	0.86	0.62	1.06	−0.44
Expense ratio (%)	1.34	1.37	1.31	0.06
Net alpha (%)	−0.50	−0.78	−0.27	−0.52
AUM (million USD)	240	336	165	171^{**}
Fund age (years)	12.0	13.0	10.7	2.3
Competition (1 − HHI)	0.82	0.84	0.81	0.03

2.B.4 Variables

In this section, we present the variables divided into outcome variables and control variables. The outcome variables are tested in the regressions, and the control variables are included in the vector of controls. We use control variables widely used in the literature.

Table 2.B.3.
Variable definitions

This table documents our variables and their definitions.

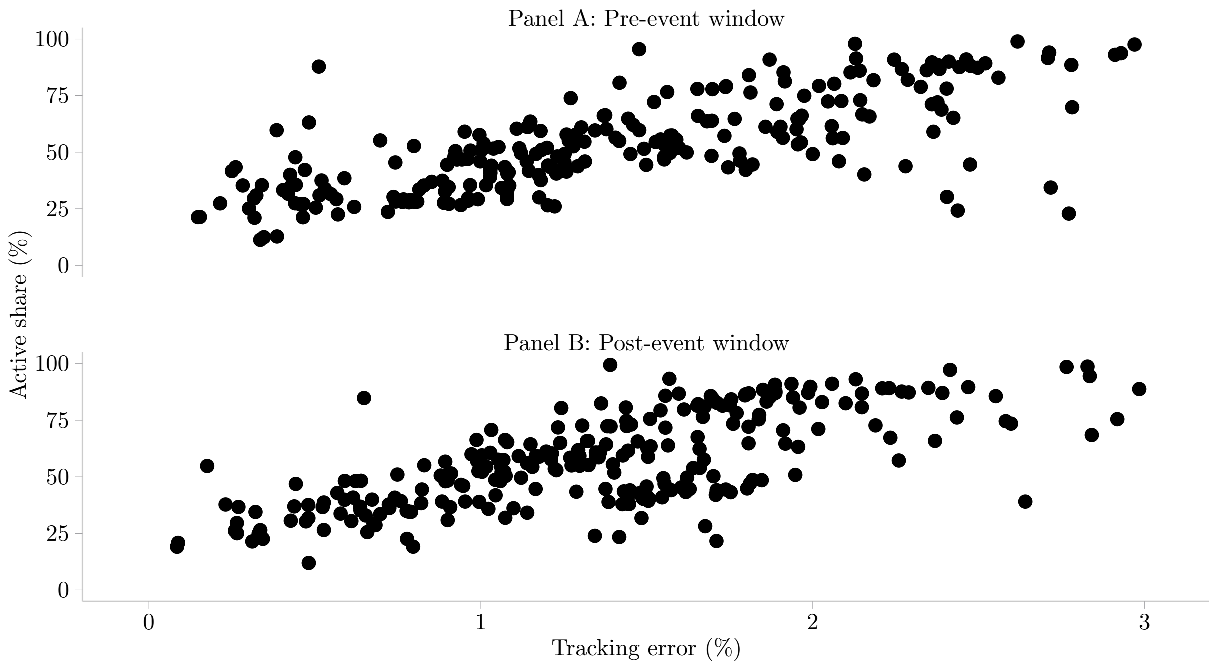
Name	Definition
Active share	Percentage of a fund's portfolio holdings that differ from its benchmark index holdings.
Expense ratio	Monthly expense ratio.
Gross alpha	Difference between the fund gross return and its benchmark return.
Net alpha	Difference between the fund net return and its benchmark return.
Factor-adjusted alpha	Three-factor alpha (percentage per month) with country-specific factors.
AUM	Total assets under management in USDm for all share classes.
Fund age	Number of years since the fund's launch date.
Industry competition	1 − Herfindah-Hirschman index (HHI). HHI is the fund family-level industry concentration in the country and defined as $\text{HHI} = \sum_{i=1}^{N} (w_i)^2$, where w_i is the AUM-based weight of fund family i in the country.

2.B.5 Relationship between active share and tracking error

An alternative measure of active management level is tracking error, i.e., the standard deviation of the funds' active returns. Active share and tracking error are often used in combination to determine whether funds are potential closet indexers, where the active share is forward-looking while tracking error requires historical data for calculation. Figure 2.B.1 plots the mean active share against tracking error of monthly return observations in the pre- and post-event window, in Panel I and II, respectively. The correlation coefficients between the two variables are 0.73 in the pre-event window and 0.74 in the post-event window. This shows that these two measures are highly correlated for the domestic funds in our sample. This confirms the findings from ESMA (2020).

Figure 2.B.1.
Tracking error and active share

This figure presents the relationship between tracking error and active share in the fund sample. Panel A plots it for the pre-event window, and panel B for the post-event window.



2.C Appendix with additional analysis for impact on active share

2.C.1 Testing for pre-trends

Table 2.C.1.
Pre-trend active share

This table reports estimated slope coefficients (Post × t) from panel regressions testing the presence of pre-trends in **active share** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All regressions include fund and month fixed effects, and column 2 additionally controls for fund age, size, and fees. Standard errors clustered by fund and month are in parentheses. There are 1,864 and 2,875 fund-month observations in the 40% and 50% panels, respectively. Asterisks denote statistical significance: $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

	Post × t	
Fund sample	(1)	(2)
Active share ≤ 40%	−0.13*	−0.14*
	(0.07)	(0.08)
Active share ≤ 50%	−0.02	−0.03
	(0.06)	(0.06)
Controls		×

2.C.2 Impact on active share based on fund characteristics

Table 2.C.2.
Heterogeneity in the effect of intervention on active share

This table reports heterogeneous effects of policy interventions on **active share** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. Funds are classified as "high" and "low" based on the median value in their respective country. The high (low) column reports difference-in-differences estimates using a high (low) sub-sample of treated and control funds. The difference column contains estimates from a tripe difference-in-differences model. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

Fund sample	Active share ≤ 40%		
Fund characteristic	High	Low	Difference
Size	8.64**	10.04***	−1.40
	(3.34)	(3.47)	(4.86)
Age	12.41***	5.65	6.76
	(2.45)	(4.01)	(4.79)
Fee	10.77***	8.11**	2.66
	(3.03)	(3.16)	(4.36)
Performance	9.92***	9.08***	0.85
	(3.42)	(3.13)	(4.61)
Fund sample	Active share ≤ 50%		
Fund characteristic	High	Low	Difference
Size	7.08***	3.54	3.54
	(2.65)	(2.98)	(3.94)
Age	9.26***	2.03	7.23*
	(2.49)	(3.10)	(3.99)
Fee	3.12	8.31***	−5.18
	(2.60)	(2.61)	(3.63)
Performance	6.28**	5.01*	1.27
	(2.52)	(2.65)	(3.64)

2.C.3 Impact on active share using fixed effects

Table 2.C.3.
Active share-scrutiny relationship

This table reports estimated slope coefficients (Post × Scrutiny) from panel regressions testing the effects of policy interventions on **active share** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All regressions include fund and month fixed effects, and columns 3 and 4 additionally control for fund age, size, and fees. Standard errors (in parentheses) clustered by fund are reported in columns 1 and 3 and independently by fund and month in columns 2 and 4. There are 3,711 and 5,582 fund-month observations in the 40% and 50% panels, respectively. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Post × Scrutiny			
Fund sample	(1)	(2)	(3)	(4)
Active share ≤ 40%	9.27***	9.27***	8.24***	8.24***
	(2.31)	(2.31)	(2.22)	(2.22)
Active share ≤ 50%	6.50***	6.50***	5.97***	5.97***
	(1.94)	(1.95)	(1.87)	(1.88)
Controls			×	×
Fund cluster	×	×	×	×
Month cluster		×		×
Adj. R^2 [40% | 50%]	[0.57 | 0.66]		[0.59 | 0.66]	

2.D Appendix with additional analysis for impact on fee

2.D.1 Testing for pre-trends

Table 2.D.1.
Pre-trend fees

This table reports estimated slope coefficients (Post × t) from panel regressions testing the presence of pre-trends in **fees** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All regressions include fund and month fixed effects, and column 2 additionally controls for fund age and size. Reported coefficients are annualized from fund-month level observations. Standard errors clustered by fund and month are in parentheses. There are 1,861 and 2,879 fund-month observations in the 40% and 50% panels, respectively. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Post × t	
Fund sample	(1)	(2)
Active share ≤ 40%	−0.00	−0.00
	(0.00)	(0.00)
Active share ≤ 50%	−0.00	−0.00*
	(0.00)	(0.00)
Controls		×

2.D.2 Impact on fee based on fund characteristics

Table 2.D.2.
Heterogeneity in the effect of intervention on fees

This table reports heterogeneous effects of policy interventions on **fees** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. Funds are classified as "high" and "low" based on the median value in their respective country. The high (low) column reports difference-in-differences estimates using a high (low) sub-sample of treated and control funds. The difference column contains estimates from a tripe difference-in-differences model. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

Fund sample	Active share ≤ 40%		
Fund characteristic	High	Low	Difference
Size	0.03	-0.19^{*}	0.21^{*}
	(0.05)	(0.11)	(0.12)
Age	-0.14	0.02	-0.16
	(0.09)	(0.03)	(0.11)
Fee	-0.19^{*}	0.06^{*}	-0.25^{**}
	(0.11)	(0.03)	(0.11)
Performance	-0.09	0.01	-0.10
	(0.08)	(0.05)	(0.10)
Fund sample	**Active share ≤ 50%**		
Fund characteristic	High	Low	Difference
Size	0.04	-0.03	0.07
	(0.04)	(0.07)	(0.08)
Age	-0.09	0.11^{**}	-0.20^{***}
	(0.06)	(0.05)	(0.07)
Fee	-0.04	0.08^{**}	-0.11
	(0.07)	(0.03)	(0.08)
Performance	0.00	0.04	-0.05
	(0.07)	(0.05)	(0.08)

2.D.3 Impact on fee using fixed-effects

Table 2.D.3.
Fee-scrutiny relationship

This table reports estimated slope coefficients (Post × Scrutiny) from panel regressions testing the effects of policy interventions on **fees** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All regressions include fund and month fixed effects, and columns 3 and 4 additionally control for fund age, and size. Reported coefficients are annualized from fund-month level observations. Standard errors (in parentheses) clustered by fund are reported in columns 1 and 3 and independently by fund and month in columns 2 and 4. There are 3,734 and 5,761 fund-month observations in the 40% and 50% panels, respectively. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Post × Scrutiny			
Fund sample	(1)	(2)	(3)	(4)
Active share ≤ 40%	−0.05 (0.05)	−0.05 (0.05)	−0.06 (0.05)	−0.06 (0.05)
Active share ≤ 50%	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)
Controls			×	×
Fund cluster	×	×	×	×
Month cluster		×		×
Adj. R^2 [40% | 50%]	[0.92 | 0.92]		[0.92 | 0.92]	

2.E Additional analyses regarding impact on alpha

2.E.1 Testing for pre-trends

Table 2.E.1.
Pre-trend alpha

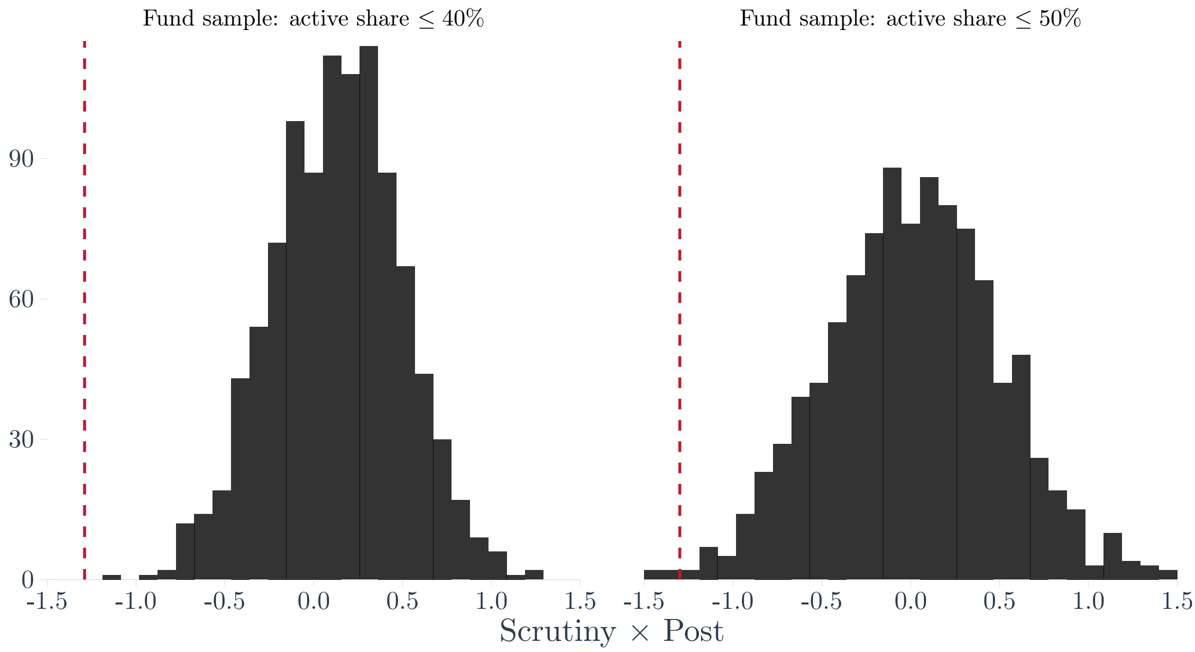
This table reports estimated slope coefficients (Post × t) from panel regressions testing the presence of pre-trends in **alpha** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All regressions include fund and month fixed effects, and column 2 additionally controls for fund age, size, and fees. Reported coefficients are annualized from fund-month level observations. Standard errors clustered by fund and month are in parentheses. There are 1,859 and 2,876 fund-month observations in the 40% and 50% panels, respectively. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Post × t	
Fund sample	(1)	(2)
Active share ≤ 40%	−0.04	−0.04
	(0.16)	(0.17)
Active share ≤ 50%	0.04	0.05
	(0.18)	(0.20)
Controls		×

2.E.2 Placebo tests

Figure 2.E.1.
Random assignment to treatment and control group

This figure presents histograms of the estimated coefficients of a falsification test for the difference-in-differences model for fund alpha, equivalent to the one presented in column 2 in Table 2.5.3. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. In each of the 1,000 separate estimations, the treatment and control groups are randomly assigned following a uniform distribution with the ratio of treated to control identical to that in the original sample (see Table 2.4.2). The model is then re-estimated using the randomly assigned treatment variable. The reported coefficients are for the interaction Scrutiny × Post. The dashed vertical lines mark the corresponding coefficient estimates from Table 2.5.3 column 2.



2.E.3 Impact on alpha based on fund characteristics

Table 2.E.2.
Heterogeneity in the effect of intervention on alpha

This table reports heterogeneous effects of policy interventions on **alpha** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. Funds are classified as "high" and "low" based on the median value in their respective country. The high (low) column reports difference-in-differences estimates using a high (low) sub-sample of treated and control funds. The difference column contains estimates from a tripe difference-in-differences model. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

Fund sample	Active share ≤ 40%		
Fund characteristic	High	Low	Difference
Size	−1.64**	−0.95	−0.70
	(0.81)	(0.60)	(1.02)
Age	−1.62**	−0.90	−0.71
	(0.75)	(0.69)	(1.09)
Fee	−0.93	−1.53*	0.60
	(0.64)	(0.88)	(1.11)
Performance	−1.03*	−1.00	−0.04
	(0.59)	(0.63)	(0.98)
Fund sample	Active share ≤ 50%		
Fund characteristic	High	Low	Difference
Size	−0.80	−1.84***	1.04
	(0.72)	(0.67)	(0.98)
Age	−1.13	−1.49**	0.36
	(0.69)	(0.64)	(0.95)
Fee	−1.31**	−1.22*	−0.08
	(0.57)	(0.73)	(0.96)
Performance	−1.27**	−1.25**	−0.01
	(0.57)	(0.61)	(0.78)

2.E.4 Value creation alternatives

Table 2.E.3.
Effect of intervention on alternative measures of value creation

This table reports the consequences of policy interventions on **net alpha** and **factor-adjusted alpha** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. Estimation is based on two-year pre- and four-year post-event averages. The difference column presents post-event differences between treated and control funds. The difference-in-differences (DiD) follows a classic setup, see Angrist and Pischke (2008) for details. Reported coefficients are annualized from fund-month level observations. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net alpha		Factor-adjusted alpha	
Fund sample	Difference	DiD	Difference	DiD
Active share ≤ 40%	-0.73^{**}	-1.19^{**}	-0.41	-1.07^{**}
	(0.31)	(0.54)	(0.39)	(0.48)
Active share ≤ 50%	-1.14^{***}	-1.27^{***}	-0.11	-0.64
	(0.31)	(0.47)	(0.38)	(0.42)

Table 2.E.4.
Value creation-scrutiny relationship

This table reports estimated slope coefficients (Post × Scrutiny) from panel regressions testing the effects of policy interventions on **net alpha** (panel A) and **factor-adjusted alpha** (panel B) for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All regressions include fund and month fixed effects, and columns 3 and 4 additionally controls for fund age, size, and fees. Reported coefficients are annualized from fund-month level observations. Standard errors (in parentheses) clustered by fund are reported in columns 1 and 3 and independently by fund and month in columns 2 and 4. Asterisks denote statistical significance: $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

	Panel A: Net alpha			
Fund sample	(1)	(2)	(3)	(4)
Active share ≤ 40%	−1.09**	−1.09	−0.94*	−0.94
	(0.48)	(1.13)	(0.51)	(1.22)
Active share ≤ 50%	−1.11**	−1.11	−1.09**	−1.09
	(0.43)	(1.29)	(0.45)	(1.35)
Controls			×	×
Fund cluster	×	×	×	×
Month cluster		×		×
Adj. R^2 [40% | 50%]	[0.10 | 0.13]		[0.10 | 0.13]	

	Panel B: Factor-adjusted alpha			
Fund sample	(1)	(2)	(3)	(4)
Active share ≤ 40%	−0.95*	−0.95	−0.69	−0.69
	(0.49)	(1.08)	(0.51)	(1.14)
Active share ≤ 50%	−0.56	−0.56	−0.46	−0.46
	(0.42)	(1.03)	(0.43)	(1.06)
Controls			×	×
Fund cluster	×	×	×	×
Month cluster		×		×
Adj. R^2 [40% | 50%]	[0.09 | 0.10]		[0.09 | 0.10]	

2.F Additional analyses regarding signal jamming

Table 2.F.1.
Portfolio concentration-scrutiny relationship

This table reports estimated slope coefficients (Post × Scrutiny) from panel regressions testing the effects of policy interventions on **portfolio concentration** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All regressions include fund and month fixed effects, and columns 3 and 4 additionally controls for fund age, size, and fees. Standard errors (in parentheses) clustered by fund are reported in columns 1 and 3 and independently by fund and month in columns 2 and 4. There are 3,709 and 5,577 fund-month observations in the 40% and 50% panels, respectively. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Post × Scrutiny			
Fund sample	(1)	(2)	(3)	(4)
Active share ≤ 40%	0.61*** (0.21)	0.61*** (0.22)	0.50** (0.22)	0.50** (0.23)
Active share ≤ 50%	0.34 (0.21)	0.34 (0.22)	0.29 (0.21)	0.29 (0.22)
Controls			×	×
Fund cluster	×	×	×	×
Month cluster		×		×
Adj. R^2 [40% | 50%]	[0.76 | 0.82]		[0.77 | 0.82]	

2.G True active Scandinavian funds as control funds

As an alternative control group, we use another group of funds not under scrutiny. This group consists of truly active funds in Scandinavia. However, these funds have differing characteristics. If we consider closet indexing to be a disease, contrasting closet index funds and truly active funds is tantamount to comparing sick funds with healthy ones. Such a comparison introduces a selection bias. Fund fixed effects may partly remedy this problem. As a robustness test of our main results, we carry out the same analyses as previously.

In Table 2.G.1, we present the same summary statistics as we did in Table 2.4.2, but the control group is now truly active funds. There are three main differences between the samples. Firstly, active share is lower for the closet index funds. Secondly, even if not by much (about 24 basis points annually), the group of closet index funds is cheaper than truly active funds. Finally, the closet index funds are older than truly active funds. This confirms that closet index funds may often be old funds with uninformed investors.

Table 2.G.1.
Summary statistics: Scandinavian funds

This table presents summary statistics for the sample of actively managed domestic equity mutual funds in Scandinavia (Norway, Sweden, and Denmark). Scandinavian non-closet indexers are assigned to the control group. Values are means over a two-year window before the event start. Gross alpha, expense ratio, and net alpha are annualized. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. Inference on differences between treatment and control funds is based on Newey and West (1987) standard errors. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

Fund sample	Active share ≤ 40%			Active share ≤ 50%		
Group	Treated	Control	Difference	Treated	Control	Difference
Number of funds	46	104		75	75	
Active share (%)	32.4	61.4	−29.0***	38.0	67.0	−29.0***
Gross alpha (%)	1.04	0.77	0.27	1.24	0.45	0.79
Expense ratio (%)	1.15	1.44	−0.29***	1.24	1.46	−0.22***
Net alpha (%)	−0.14	−0.73	0.59	−0.05	−1.05	1.00
AUM (million USD)	476	283	193*	379	306	73
Fund age (years)	16.2	12.0	4.2***	15.0	11.6	3.4***

In Table 2.G.2 and 2.G.3 we present how scrutiny influence closet index funds relative to truly active funds.

Table 2.G.2.
Robustness: effect of intervention on main outcomes

This table reports the effects of policy interventions on **active share**, **fees** and **alpha** for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. The control group comprises Scandinavian non-closet indexers. Values for fees and alpha are annualized. Estimation is based on two-year pre- and post-event averages (for alpha four-year post event). The difference column presents post-event differences between treated and control funds. The difference-in-differences (DiD) follows a classic setup, see Angrist and Pischke (2008) for details. For fee and alpha, reported coefficients are annualized from fund-month level observations. Newey and West (1987) standard errors in parentheses. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Active share		Fees		Alpha	
Fund sample	Difference	DiD	Difference	DiD	Difference	DiD
Active share ≤ 40%	−30.32***	3.16*	−0.41***	−0.09*	−1.62***	−2.21***
	(2.48)	(1.86)	(0.07)	(0.05)	(0.50)	(0.84)
Active share ≤ 50%	−26.96***	1.30	−0.27***	−0.05	−1.46***	−2.25***
	(2.09)	(1.72)	(0.06)	(0.04)	(0.50)	(0.80)

Table 2.G.3.
Robustness: main outcomes-scrutiny relationship

This table reports estimated slope coefficients (Post × Scrutiny) from panel regressions testing the effects of policy interventions on **active share** (panel A), **fees** (panel B), and **alpha** (panel C) for closet index funds. Fund sample denotes the active share cutoff limit used to classify funds as closet indexers. All regressions include fund and month fixed effects, and columns 3 and 4 additionally controls for fund age, size, and fees (not in panel B). For fee and alpha, reported coefficients are annualized from fund-month level observations. Standard errors (in parentheses) clustered by fund are reported in columns 1 and 3 and independently by fund and month in columns 2 and 4. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Panel A: Active share			
Fund sample	(1)	(2)	(3)	(4)
Active share ≤ 40%	3.86* (2.02)	3.86* (1.99)	3.39 (2.09)	3.39 (2.05)
Active share ≤ 50%	1.27 (1.85)	1.27 (1.83)	0.78 (1.93)	0.78 (1.89)
Controls			×	×
Fund cluster	×	×	×	×
Month cluster		×		×
Adj. R^2 [40% | 50%]	[0.78 | 0.77]		[0.77 | 0.77]	

	Panel B: Fees			
Fund sample	(1)	(2)	(3)	(4)
Active share ≤ 40%	−0.07 (0.06)	−0.07 (0.05)	−0.04 (0.05)	−0.04 (0.05)
Active share ≤ 50%	−0.03 (0.04)	−0.03 (0.04)	−0.02 (0.04)	−0.02 (0.04)
Controls			×	×
Fund cluster	×	×	×	×
Month cluster		×		×
Adj. R^2 [40% | 50%]	[0.83 | 0.83]		[0.85 | 0.86]	

	Panel C: Alpha			
Fund sample	(1)	(2)	(3)	(4)
Active share ≤ 40%	−1.96** (0.82)	−1.96 (1.31)	−1.82** (0.83)	−1.82 (1.31)
Active share ≤ 50%	−2.08** (0.81)	−2.08* (1.24)	−2.09** (0.82)	−2.09* (1.25)
Controls			×	×
Fund cluster	×	×	×	×
Month cluster		×		×
Adj. R^2 [40% | 50%]	[0.09 | 0.11]		[0.09 | 0.11]	

Do Fees Matter?

Investor's Sensitivity to Active Management Fees

Trond Døskeland, André Wattø Sjuve, and Andreas Ørpetveit

Abstract

The active mutual fund equilibrium model developed by Berk and Green (2004) predicts that fees should not matter for investors' mutual fund choices. However, there is often a difference between what a rational model indicates and what investors actually do. In this study, we look at how fees influence demand for active mutual funds by analyzing time variation in active funds' fees. Since investors should not pay "alpha fees" for "beta performance", we measure fees both as an excess fee above the passive alternative and as the unit price of active management where active share represents the activity level. Using international data, we identify a negative time-series relation between fee measures and fund flows. The results also hold after controlling for Morningstar ratings.

Keywords: Mutual funds; Active portfolio management; Fund Fees; Fund Flows

JEL Classification D14; G11; G24

3.1 Introduction

If I could have a law named after me, it'd be that no investment product is so good that high enough fees cannot make it a bad investment.

—Cliff Asness, Ilmanen (2022)

Investors face a trade-off between the potential for excess returns, created by a perceived manager-specific skill, and the extra cost for active management. A rich literature examines whether funds actually create excess returns, but equally crucial for final wealth - and a less studied area - is the extra costs associated with active management. This gap motivates us to take a fresh look at whether fees guide investors trading in actively managed funds.

We analyze how fees influence the demand for active mutual funds by analyzing time variation in active funds' fees. Even where fee levels have been quite sticky; we document significant variations within funds for the extra cost of active management. We refer to this extra cost as the excess fee and define it as the difference between the total expense ratio of the fund and the cost of the index fund substitute. Since, investors should not pay alpha fees for beta performance, we also calculate the unit cost of active management, i.e., the excess fee divided by the activity level of the fund. We represent the activity level by active share (Cremers and Petajisto, 2009). The demand for the funds is defined as net fund flows. The main question we seek to answer is thus: Does investors' capital flow into (out of) funds that decrease (increase) the cost of active management?

It is not obvious what to expect from the relationship between fees and demand for active management, as reflected in the first part of the title of this paper: "Do Fees Matter?". According to the neoclassical view of mutual funds, when in equilibrium, gross-of-fee excess returns (or alphas) will on average be equal to fees, and average net alphas (after fees) will be zero (Berk and Green, 2004; Berk and Van Binsbergen, 2015). Thus, with competitive fund markets and rational investors, fees should not matter to mutual fund investors; investors expect zero net alpha independent of the fee level.

Despite the important theoretical equilibrium argument, the answer to whether fees *actually* matter when it comes to how investors allocate capital is partly unanswered. One strand of the literature claims that investors behave rationally and use a full asset pricing model when deciding which funds to buy and sell (see Barber et al. (2016) and Berk and van Binsbergen (2016)). While the idea that investors allocate capital using risk-adjusted returns is attractive, many studies find that investors display behaviors that may be considered suboptimal or unsophisticated; see, for example, Ben-David et al. (2022).[1] With a narrative containing imperfections, it is more intuitive that fees matter for how investors allocate their capital. The more (less) you pay for a product, the less (more) you buy.

Based on these conflicting flow-fee relations, we formalize the fee sensitivity predictions by developing a simple model of fund flow in the presence of time-varying fees. In the partial version of the model, the flow response is negative, i.e., negative development of the investor's allocation with increased excess fees and positive for increased activity level. The model also predicts a negative time-series relation between the unit cost of active management and subsequent fund flow. When we expand the model to an equilibrium setting in line with, for example, the model of Berk and Green (2004), we do not find any relationship between fees and flows.

Empirically, we identify the flow-fee relation by running a time-series regression between fees and corresponding flow changes with fund fixed effects. Although we report similar cross-sectional results in the Appendix, the main focus is on the time-series relation. We believe this estimate to closest to our question of how sensitive investors are to changes in fees.

We add to the previous flow-fee literature along two dimensions. Firstly, we examine the time-series relation between fees and flows for a given fund. This is in contrast to prior studies that have looked at whether there is a flow-fee relation across funds. Sirri and Tufano (1998) and Barber et al. (2005) find a negative relationship between total fund fee and mutual fund flow. The empirical strategy of this study is intended to enable an analysis of the causal relationship between fees and flows within funds.

Secondly, we examine more detailed fee measures than has previously been done. This firstly comprises the excess fee of active management above the fee level of passive funds in the given country, where we propose that the price for the active management service is the difference between the passive alternative and the current active fund. The second fee measure defines how much activity you get for the excess cost of active management. This definition is similar to what Cremers and Curtis (2016) call *active fee*. The numerator is the same as in the first

[1] Examples of "signals" investors respond to are external rankings (Morningstar: Del Guercio and Tkac (2008), Ben-David et al. (2022), and Evans and Sun (2021), Wall Street Journal: Kaniel and Parham (2017), sustainability: Hartzmark and Sussman (2019)) and past returns (Chevalier and Ellison (1997), Choi and Robertson (2020)).

measure, i.e., the excess fee. The denominator is the activity level, defined as active share by Cremers and Petajisto (2009).[2] To our knowledge, there is no previous literature on something as basic as price sensitivity to how much active management you get for what you pay, i.e., "the bang for the buck". In the same way as when purchasing other products, the buyer should therefore compare the product's unit price.[3]

Compared to other products, the properties of the mutual fund industry are unique in the sense that the funds do not have to decide on the number of shares (quantity) to supply. However, the funds can adjust their fees by changing the total fee level or adjust their unit price by changing the activity level.

Mutual fund flows help us to understand the factors influencing consumers' economic thinking since the decisions aggregate a large portion of society. As the mapping from observed behavior to beliefs is not one-to-one, investors' true beliefs are unknown. However, we show that in practice, investors utilize fees to pick active funds. Firstly, we find that a fund's excess fee negatively predicts the fund's subsequent flow. Secondly, we find that a fund's activity level in the form of active share positively predicts the fund's subsequent flow. Combining these two measures to the unit price, we find that a fund's active fee negatively predicts its subsequent flow. For example, a one standard deviation increase in a fund's active fee translates to a decrease in annual net flow of 22.8 bps ($0.228\% = -3.51 \cdot 0.065$).

The estimated effects given in this paper are more or less identical to the effects found by Barber et al. (2005), which they claim have moderate economic significance.[4] As a benchmark, we compare the estimated effects to the corresponding effects of a one-standard-deviation increase in Morningstar ratings. It is well-known that Morningstar ratings are an important explanatory variable for net flow. We find that the estimated effects of excess and active fee account for around 10% of the estimated effects of Morningstar ratings. This is an interesting finding since we claim that fees are more informative when it comes to future profit for investors. Fees and active share affect a fund's ex-ante potential to outperform its benchmark index. Investors can identify and respond to these variables, while they cannot directly control the excess return and the subsequent Morningstar rating.

[2] The first edition of the paper is from August 2006, see `https://users.nber.org/~confer/2006/apf06/cremers.pdf`. Our sample period is 2006 to 2019.

[3] With unit pricing, we take the price of a product, for example, 1.59 for a 24-ounce jar of spaghetti sauce and divide the total cost by a standard unit of measurement (such as ounces). We then have a simple price comparison point, i.e., 1.59 divided by 24 ounces = 0.07 per ounce. We can then use that unit price to compare whether we get the best deal. For example, what is the unit price of a 45-ounce jar of sauce at 3.69? When we do the math, 3.69 divided by 45 ounces = 0.08 per ounce, we find that the smaller jar is the best buy.

[4] They find that a 100-basis point decrease in total expenses is associated with 0.39% growth in new money.

The effect size is approximately the same when we control for Morningstar ratings. We also find an interaction between the flow-fee relation and funds with high and low ratings, where investors in the high-rated (low-rated) funds are more fee sensitive. Moreover, when examining differences in the flow-fee relation across funds based on fund characteristics, we find evidence that investors are less (more) fee sensitive to high (low) performing funds. We find little evidence for a fee interaction with fund size. We also examine differences based on fund style, using the Morningstar 3x3 style box. We find that the sensitivity is more prominent for value funds than growth funds.

There is a debate among academics, practitioners, the justice system, and regulators regarding the competitiveness of mutual fund markets and related questions concerning mutual fund fees and their impact on investor performance. Given investors' ability to switch among funds and the number of active mutual funds, almost all funds face multiple close substitutes. Thus, the opportunity for investors to choose among rival funds is well established. However, empirical evidence suggests significant price dispersion even when products are homogeneous. Carlin (2009) shows that complexity increases market power because it prevents some consumers from becoming knowledgeable about prices in the market. Although the level of competition is debated, our results show that investors are price sensitive to a certain degree. While excess fees and active share are increasingly becoming more salient for investors, the active fee is not displayed in investor information. On this basis, we suggest that increased awareness may lead to better investment decisions and greater competition in the industry.

The rest of the paper is organized as follows: In Section 3.2, we propose a simple model and develop the hypothesis. In Section 3.3, we present the research design, including the data and sample. Section 3.4 shows the main results, while further analysis results are described in Section 3.5. Finally, we conclude the findings in Section 3.6.

3.2 Hypothesis development

We will firstly outline a simple ex-ante model of selecting active funds before presenting our hypotheses.

3.2.1 Active fund selection model

We want to see how investors respond to changes in fees and begin by modelling the profit for investors and fund managers. The gross value-added created by a manager is the product of benchmark-adjusted return, α_g, and the assets under management, AUM: $V^g = AUM \cdot \alpha^g$. The revenue for the manager is the product of quantity and a fee, EF: $AUM \cdot EF$. The surplus for the investor is the value added after the fee: $V^n = AUM(\alpha^g - EF)$. Below we discuss different models for the division of total value added between investors and asset managers.

In the simplest version of the model, we add two more properties. First, we assume that the manager chooses how much of their assets under management to actively manage. As suggested by Berk and Green (2004), the manager performs a "beta" versus "alpha" allocation, where the alpha part is actively managed while the beta part is indexed. We represent the activity level, or active part, with AS, and $1-AS$ thus becomes the index part. When testing the model, we will use active share (Cremers and Petajisto, 2009) as a representation of activity level. We could alternatively have used another measure, such as tracking error. We assume that the indexed capital earns no alpha, so the value added comes from the active part.

Second, investors pay fees to the manager for active management services. These fees are paid on the total assets under management and not on actively managed assets. As argued in Equation 15 on page 1278 of Berk and Green (2004), we can express total revenue as either a fee of total AUM ($AUM \cdot EF$) or as an "adjusted" fee of actively managed funds ($AS \cdot AUM \cdot AF$). We can interpret this adjustment as adjusting the fees for the beta part of the portfolio. Since the size of the revenue is the same, we can "back out" the adjusted fee level given by: $AF = \frac{EF}{AS}$. This measure is similar to active fee of Cremers and Curtis (2016).

Putting this all together, we can express the net value added for investors as follows:

$$V^n = AUM(AS \cdot \alpha^g - AS \cdot AF) \tag{3.1}$$

Dividing by AUM, we find net alpha to be:

$$\text{(Partial model)} \qquad \alpha^n = AS \cdot \alpha^g - EF = AS(\alpha^g - AF) \tag{3.2}$$

Based on the simple expression above, the net alpha is reduced with increasing fee, EF. For increasing AS, net alpha increases for positive α^g. Finally, given a positive AS, net alpha is reduced with increasing AF. Along the time dimension, we assume that a higher (lower) net alpha will lead to inflow (outflow) in the next period.

We can easily expand the partial model above to illustrate the Berk and Green model results. We start by assuming decreasing returns-to-scale in active management. These "costs" can have different shapes. For simplicity, we assume a negative linearly relation in actively managed capital: $\alpha^g = a - b \cdot AS \cdot AUM$, where a and b are above 0.

The second assumption is that investors chase any positive net present value investment opportunity. Investors will compete aggressively for skillful managers' services and pay so high fees or invest so much capital as to bring investors' expected net alpha to zero. This assumption implies that all assets earn an expected return commensurate with the risk of the asset. Consequently, all funds have, in equilibrium, net alphas of zero.

Incorporating these new assumptions into Equation 3.2, we get the following expression:

$$\text{(Equilibrium model)} \qquad \alpha^n = AS \cdot (a - b \cdot AS \cdot AUM) - EF = 0 \tag{3.3}$$

While investors set AUM, managers can freely adjust AS and EF in order to maximize their revenue.[5] Since managers can benchmark excess capital, they are indifferent to a large AUM with a small fee or a small AUM with a high fee.[6] In any case, since expected net alpha is zero in equilibrium, there should not be any relevant information in these two parameters that would enable the prediction of investors' demand.

The Berk-Green model is restrictive because it assumes full competition and that both managers and investors know the manager's skill level. In reality, neither a nor b are likely to be known to investors or managers. If managers are also competing for capital from investors and there is uncertainty about the manager's true skill level, the negotiating power should be more balanced. A richer model that better reflects this has been developed by Garleanu and Pedersen (2019), where value added is divided more evenly between investors and asset managers.

3.2.2 Hypotheses

Main tests

We explore the relationship between excess fee, EF, active share, AS, and active fee, AF and fund flow. We implicitly test whether investors follow the first partial model (see Equation 3.2) or a more sophisticated equilibrium model (see Equation 3.3).

[5] See, for example, section 5.1 in Pástor et al. (2020) for a description of the interplay between activity and fee level.

[6] In our tests, we examine time-variation in fees. According to Berk and Green (2004) (page 1278), it is not natural for managers to adjust fees, and it is assumed that managers adjust the activity level. We seek to show that the different fee measures vary over time and that both AS and EF (and subsequently AF) can predict demand.

We have already referred to literature that claims that fees are negatively related to flows in the cross-section; see for example Sirri and Tufano (1998) and Barber et al. (2005). Based on the partial model, we find that taking the derivative of net alpha with respect to the fee level gives a negative flow-fee relation. Regarding active share, Cremers et al. (2016) find that funds with a higher active share attract more flows. The partial model shows that the sign of the derivative of net alpha for active share is positive and depends on skill level. The third relation we examine relates the unit price of active management or active fee to net alpha. We postulate a negative time-series relation between active fee and subsequent fund.

Additional tests

To deepen our understanding of the main results above, we performed several robustness tests. As mentioned, Morningstar ratings have an important influence on fund flows and could therefore interfere with the flow-fee relation. An important difference between the variables from our model and Morningstar ratings is that our variables are decided ex-ante by the funds themselves, while ratings are given by an external provider based on realized historical performance. The tests firstly look at whether there is a relation between fee and subsequent Morningstar rating. Since mutual fund ratings are purely quantitative backward-looking measures of a fund's past performance (Huang et al., 2020), this is an implicit test of the fee-performance relation. Based on findings in, for example, Cooper et al. (2021), we suspect no such relation.

However, investors may use ratings as a proxy for skill. A higher skill will lead to a more attractive fund in our simple model. A natural question as such is whether investors remain price-sensitive after controlling for star ratings.

The effect may also vary for different segments of funds. We examine how past performance interferes with the relation between fee and fund flow. Past performance is easily accessible to investors and is a predictor of subsequent flows (see, for example, Ferreira et al. (2012)). Since Morningstar ratings build on past performance, we suggest that past performance should have the same impact as the ratings. Fee sensitivity can also interact with fund size. With decreasing-returns-to scale, as suggested by Berk and Green (2004), skill is, all else equal, lower for larger funds than for smaller funds. We therefore examine the fee sensitivity for different fund sizes. Finally, we examine whether the fund style interferes with the relationship between fee and fund flow. The flow-fee relation is explored for different Morningstar fund styles.

3.3 Research design

This section describes the research design used to test the flow-fee relation. Here we explain the main variables and main empirical strategy, provide details about our data collection and sample construction, and present descriptive statistics.

3.3.1 Main variables

Firstly, we will provide a brief overview of the main variables: net flow, excess fee, active share, and active fee. Table 3.A.1 in Appendix 3.A.1 presents a complete list of all variables.

Net flow

Drawing on the work of Chevalier and Ellison (1997) and Sirri and Tufano (1998), we define new money growth rate as the net growth rate in total net assets (TNA) that is not due to dividends and capital gains on the assets under management but to new external money. Net flow (NF) for fund i at time t is calculated as:

$$\mathrm{NF}_{i,t} = \frac{\mathrm{TNA}_{i,t}}{\mathrm{TNA}_{i,t-1}} - (1 + r_{i,t}), \tag{3.4}$$

where $\mathrm{TNA}_{i,t}$ is the total net asset value in USD for fund i at the end of time t and $r_{i,t}$ is fund i's total return from $t-1$ to t. Equation 3.4 assumes that flows occur at the end of each period, as we have no information on the timing of new investments. To compute quarterly flows with the monthly TNA observations, we sum the monthly dollar flows within the quarter, and divide the flows with TNA at the end of the previous quarter. This is done to avoid the assumption that all flows occur at the end of each quarter.

Excess fee

An active fund's total portfolio can be divided into an index fund and a "residual" fund with over- (long) or underweights (short). The excess fee, EF, is the fee of the fund in excess of the cost of the passive index fund alternative. It consists of the difference between the total expense ratio of the fund, TER, and the cost of the index fund substitute. In order for the index alternative to be a substitute, it must be set as the cost of an index fund investing in the same universe, and tracking the same underlying benchmark as the fund is measured against:

$$\mathrm{EF}_{i,t} = \mathrm{TER}_{i,t} - \text{Index fee}_t. \tag{3.5}$$

Active share

The relative weighting of the active and passive parts of the fund is determined by the fund's activity level, for which there are several different measures. We chose to use active share (AS), as introduced by Cremers and Petajisto (2009).[7] A beneficial feature of active share is that at one, all assets are actively managed, while at zero, all assets are invested in a manner similar to the benchmark.

Active fee

The second fee measure defines how much activity you get for the excess cost of active management. This definition is similar to what Cremers and Curtis (2016) call *active fee*. The numerator is, as mentioned above, active share.

3.3.2 Data

We use Lipper as our primary database as it provides a comprehensive sample of mutual funds offered across many countries. Our primary interest is the relation between fee and flow in managed portfolios and we therefore choose to focus on actively managed open-end equity mutual funds and exclude index funds, fund-of-funds, and closed-end funds. We download cross-sectional fund information (fund name, domicile, benchmark) and monthly times-series with returns, total net assets (TNA), fees and expenses (TER), as well as detailed fund holdings. Since the database provides active, merged, and liquidated funds, it is survivorship bias-free. Lipper treats separate share classes within the same fund as different observations, despite them having the same holdings and the same returns before expenses. In line with Cremers et al.

[7] Two alternative measures are tracking error and activeness (Pástor et al., 2020). With our forward-looking model, we prefer to use active share rather than ex-post tracking error. The information needed to compute active share is also easier to obtain than ex-ante tracking error. Active share is also more readily available than the activeness measure, as proposed by Pástor et al. (2020). However, there is a clear overlap between the measures. For a formal analysis, see Marmoiton (2021). Active share is computed by:

$$\mathrm{AS}_{i,t} = \frac{1}{2}\sum_{j=1}^{N_{i,t}} |w_{j,i,t} - w_{j,b,t}|, \tag{3.6}$$

where $w_{j,i}$ is the weight of security j in fund i's portfolio at time t and $w_{j,b,t}$ is security j's weight in the benchmark b at time t.

(2016), our unit of observation is the share class that Lipper identifies as the primary share class. Fund-level variables are consequently aggregated across the different share classes.

The level of observations in our tests are quarterly. For returns and fees, we calculate cumulative quarterly variables while for the other variables, we use the last value of the quarter. To estimate the relationship between flows and fee, we require a fund to have at least 12 observations on size and returns and at least two years of quarterly portfolio data.

To obtain relevant index fees for calculating excess fees, we use our data on 3,100 index mutual funds and ETFs to compute benchmark-quarter fees for the benchmarks these index funds track. We then match index fund fees with our active funds based on their respective benchmarks.[8]

The fund portfolios are used to calculate active share. This method was introduced by Cremers and Petajisto (2009), but was known a few years before the paper's publication. Consequently, our sample period is 2006−2019. The availability of fund portfolios varies across countries. When calculating active share, we follow the approach taken by Cremers et al. (2016) and use the Lipper-assigned benchmarks instead of the self-declared fund benchmarks to infer the investment opportunity sets of funds. In so doing, we avoid the problem of funds strategically picking their benchmarks. To arrive at the securities' benchmark weights, we use the weights of index funds and ETFs replicating the benchmark. To define a fund's nationality, we employ the approach taken by Schumacher (2018), who argues that the location of the management company should dictate residence rather the fund's legal domicile.

To complement the data from Lipper, we gather additional information from Morningstar Direct on fund ratings and Morningstar's style-box classifications. Morningstar's Overall Rating is a time-series variable that assigns a fund a rating between one and five stars on a monthly basis.[9] The style-box classification is a 3x3 grid variable indicating where a fund's portfolio lies in the dimensions of size (small, mid, large) and style (value, growth, blend).[10] We match the Morningstar data into the Lipper data by ISIN or fund name (text-matching) if ISIN is missing.

All variables used in the paper are defined in Table 3.A.1. The final sample consists of 10,814 funds (404,687 fund-quarter observations) with 2,7 USDtn in AUM as of December 2019. Table 3.A.2 lists the 30 sample countries along with the number of funds and fund-quarter pairs, assets under management, average net flow, excess fee, active share, and active fee for each domicile.

[8] If a benchmark-quarter fee is missing, we substitute in the average domicile-quarter index fee. If both are missing, we use the global-quarter index fee.

[9] See Ben-David et al. (2022) and Morningstar Ratings for details on Morningstar ratings and their methodology.

[10] See Morningstar Style-Box for details.

3.3.3 Descriptive statistics

Table 3.3.1 presents the sample summary statistics of our test and control variables across all funds and quarters. We find that the average fund has a quarterly outflow of 0.20%. On average, the funds in the sample charge 0.23% quarterly for active management over the passive alternative, have an active share of 77%, and a quarterly unit price of active management of 0.32%. We also note that the net flow variable has high variation, but that the mean is comparably low.

Berk and Green (2004) claim that fees, in general, are quite sticky. Therefore, to examine the time-variation for the funds over the sample period, we also present the within-fund standard deviation of the main variables. In contrast to the standard deviations measured across both funds and time, these standard errors isolate the variation across time for each fund.[11] The within-fund standard deviations of excess fee and active fee are around 25% of the standard deviations across the whole sample, illustrating that fees are relatively sticky for funds. This decline in standard deviation when examining within-fund variation is not surprising, as the cross-section of funds contains funds from many different countries and industries, with differences in general fee-level. However, we argue that the variables exhibit sufficient variation for meaningful inference in a fund fixed effects set-up.[12]

To illustrate the development in our main variables during the sample period, we plot the annual equal-weighted times series of excess fee, active share, and active fee in figures 3.A.1-3.A.3 in Appendix 3.A. Figure 3.A.1 shows that the total expense ratio has been fairly stable during the sample period, but that the average index fee decreased in 2019, leading to an increase in the excess fee in the same year. Active share increased during the sample period, from around 75% to 80%. As our sample period consists of the decade after the introduction of active share (Cremers and Petajisto (2009)), it is not surprising that fund managers have become more aware of this activity measure and increased active share. We also note that active share decreased somewhat towards the end of the sample period. The increase in excess fee together with a small decline in active share at the end of the sample period leads to an increase in active fee in 2019. In the years prior to 2019, active fee was stable, with some variation from year to year.

[11] We use these variations to interpret the economic significance of the estimates from the main variables relevant to our regression specification.

[12] In addition, we plot the distribution of deviations from within-fund means in Figure 3.A.4. These plots also show that the data has variation in the main variables.

Table 3.3.1.
Summary statistics

This table presents summary statistics for our main variables and additional control variables used in the panel regressions. For each variable, we report the number of fund-quarter observations, mean and medians, standard deviation, within-fund standard deviations, as well as the minimum and maximum values based on the full set of observations. The variables are described in Table 3.A.1 in Appendix 3.A.1. To ensure that extreme values do not drive our results, continuous variables are winsorized at the bottom and top 1% level of the distribution. The within-fund standard deviation is the average standard deviation for each fund.

Variable	N	Median	Mean	Std.dev	Min	Max	Within-fund Std.dev
Net flow (%)	404,687	-1.36	-0.20	12.52	-29.84	61.00	11.05
Excess fee (%)	401,940	0.22	0.23	0.15	0.00	0.76	0.04
Active share (%)	280,200	81.10	77.40	18.50	26.60	100.0	5.90
Active fee (%)	279,040	0.29	0.32	0.21	0.00	1.08	0.06
Gross alpha (%)	404,687	-0.22	-0.22	3.43	-11.02	10.59	3.12
TNA (USDm)	404,687	72.00	375	959	1.00	6,835	134
Fund age (years)	404,687	10.80	12.70	9.50	0.30	50.60	2.70
Family TNA (USDbn)	404,687	1.98	11.31	36.43	0.00	298.11	2.51
Industry size (USDbn)	404,687	136	632	1,123	0.00	5125	161
Competition (1-HHI)	404,687	97.20	95.90	4.90	30.50	99.50	1.00
Morningstar star rating	310,680	3.00	3.08	1.07	1.00	5.00	0.63

3.3.4 Empirical strategy

Our main test is related to the time-series relation between net flow and our main variables, as derived in Section 3.2. We specify the following regression:[13]

$$\mathrm{NF}_{i,t} = \varphi_{k,i} + \gamma_k \mathrm{X}_{k,i,t-1} + \theta \mathbf{Z}_{i,t-1} + \varepsilon_{k,i,t} \tag{3.7}$$

where $\mathrm{NF}_{i,t}$ is fund i's net flow at time t. Since we have three independent variables of interest, we denote that $\mathrm{X}_{k,i,t-1}$ can be the main variable k for fund i at time $t-1$, where $k \in \{\mathrm{EF}, \mathrm{AS}, \mathrm{AF}\}$.

As pointed out by Ben-David et al. (2022), researchers should be careful with interpreting inference when combining time-series and cross-sectional variation. Since we model active capital allocation as a function of time-varying fees, our primary interest lies in the time-series relationship between the two. Therefore, Equation 3.7 includes the i subscript on the constant term $\varphi_{k,i}$, indicating that we do not require equal intercepts across funds. The main results

[13]We label the slope coefficient capturing the flow-excess fee relation γ_{EF}, the slope coefficient for flow-active share γ_{AS}, and flow-active fee γ_{AF}

of this paper are thus based on the estimated relationship between net flow and the main test variables from a fund fixed effects panel regression.[14] In the same way as fund fixed effects isolate the time variation, using time fixed effects alone is equivalent to using only cross-sectional variation.[15]

Detecting the relation between flow and our main variables requires variation in the dependent variables. If a fund does not change its fee or vary its activity level when we observe it, it will not contribute to the results. Therefore, a potential concern is that our results depend on only a small sub-sample of funds. In Figure 3.A.4 in Appendix 3.A, we plot the distribution in our main test variables demeaned fund-by-fund and show that the variables exhibit sufficient variation.

$\mathbf{Z}_{i,t-1}$ is a vector of control variables expected to affect the net flow of the funds, as explained in Table 3.A.1. Our main specification includes only fund fixed effects, while we present results adding time fixed effects in the Appendices.

To assess the statistical significance of the flow-fee slope estimates, we cluster independently by fund and domicile$-$quarter. Our rationale for clustering by fund is that net flows are shown to be persistent (see Coval and Stafford (2007)). Furthermore, we cluster on domicile$-$quarter as well to allow for cross-sectional dependence in net flows within each country, for example as a result of country-specific macroeconomic events or differences in countries' flow-characteristics at specific dates.

3.4 Main results

In this section, we report the estimated slope coefficients on excess fee ($\hat{\gamma}_{EF}$), active share ($\hat{\gamma}_{AS}$), and active fee ($\hat{\gamma}_{AF}$), from our main regression specification in Equation (3.7), both with and without additional control variables. Since our main interest is the time-series relationship between flow and the test variables, all regressions include fund fixed effects. However, we also report results from cross-sectional tests in Appendix 3.B. As outlined in the hypothesis development in Section 3.2, we hypothesize a negative time-series relation between excess fee and flow and a positive relation between active share and flow. This in turn implicates a negative relation between active fee, i.e., the unit price of active management, and flow.

The results are shown in Table 3.4.1. Starting with the results from testing excess fee on

[14]Fund fixed effects can isolate the time-series variation and exclude the cross-sectional variation when using OLS to estimate γ, as shown by Pástor et al. (2017).

[15]Here, the $\hat{\gamma}$ coefficient becomes a weighted average of yearly cross-sectional OLS regressions (see Pástor et al. (2017) for details).

net flow in Columns 1 and 2, we find that the slope coefficients are negative and statistically significant at 1% without control variables and 5% with control variables. The reduction in the reported slope's magnitude when adding control variables shows that the additional control variables soak up some of the variation, but the slope is still statistically significant. The estimated slopes imply that a one standard deviation increase in a fund's excess fee translates to a reduction in subsequent quarterly net flow of 22.2 bps ($-0.222\% = -5.54 \cdot 0.04$) without control variables, and 7.3 bps ($-0.073\% = -1.83 \cdot 0.04$) with control variables.

In Columns 3 and 4, we test the flow-active share relation. The sign of $\hat{\gamma}_{AS}$ matches our ex-ante expectations, and both coefficient estimates are statistically significant at the 1% level. The slope estimates in Columns 3 and 4 suggest that a one (within-fund) standard deviation increase in active share increases expected net flow by 11.9 bps ($0.119\% = 0.02 \cdot 5.93$).

Our final tests relate to an expected negative time-series relation between active fee and subsequent flows. These regression results are presented in Columns 5 and 6. Both estimates of $\hat{\gamma}_{AF}$ are negative and statistically significant at the 1% level. As regards excess fee, the magnitude of the coefficient declines when we add control variables, but the reduction is lower compared to the tests on excess fee. The estimated slopes imply that a one standard deviation increase in a fund's active fee translates to a decrease in annual net flow of 22.8 bps ($-0.228\% = -3.51 \cdot 0.065$) without control variables or 13.4 bps ($-0.134\% = -2.06 \cdot 0.065$) with control variables.

To further analyze these relationships, we perform two sets of robustness tests. Firstly, we add time fixed effects to the main regression in order to rule out the possibility that our findings are the result of time-specific shocks to net flow. Related to the regression in Equation (3.7), this implies the addition of a vector of domicile-quarter dummy variables (μ_t). This addition controls for any unobserved variables that change over time but not across funds within each sample country. These results are presented in Table 3.B.1 in Appendix 3.B. Statistically, the results hold, with some changes to the magnitude of the coefficients.

Secondly, we run cross-sectional regressions with domicile-quarter fixed effects. These regressions allow for cross-fund variation across different dates in the domiciles. In Equation (3.7), the fund fixed effects are replaced by time fixed effects (μ_t), and test how investors allocate capital across funds within domicile-quarters at each time. The regressions are presented in Table 3.B.2. Although they are not a direct test of our model, it is reassuring to see that the relations for the most part also hold in the cross-section.[16]

[16]These results are also in line with other studies testing the flow-fee relation in the cross-section (see, e.g., Sirri and Tufano (1998) and Barber et al. (2005)).

Table 3.4.1.
Net flow-main variables relationship

This table reports estimated slope coefficients ($\hat{\gamma}_{EF}$, $\hat{\gamma}_{AS}$, and $\hat{\gamma}_{AF}$) from panel regressions testing the relationship between net flow and our main fee variables, on the form presented in Equation (3.7). Columns 1-2 present the results for excess fee, Columns 3-4 for active share, and Columns 5-6 for active fee. All regressions include fund fixed effects. Other control variables are described in Table 3.A.1 in Appendix 3.A.1. Standard errors (in parentheses) are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow$_t$					
X	Excess fee		Active share		Active fee	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{X}_{t-1}$	-5.54^{***} (1.09)	-1.83^{**} (0.88)	0.02^{***} (0.004)	0.02^{***} (0.004)	-3.51^{***} (0.59)	-2.06^{***} (0.50)
Other controls		×		×		×
Fund FE	×	×	×	×	×	×
Adjusted R^2	0.080	0.141	0.085	0.135	0.085	0.134
Observations	401,940	401,940	280,200	280,200	279,040	279,040

To sum up, we have established a clear statistical negative time-series relationship between fee and ensuing net flows in the following quarter. Thus, the results in this section are consistent with the hypotheses derived from our model of active management in Section 3.2. Our findings suggest that investors use fee and active share when buying and selling funds, and this implies that also the unit cost of the funds is an important variable in fund selection.

These results suggest that investors use these variables to evaluate the funds' ex-ante upside potential to outperform the benchmark index. These results are in agreement with the proposal that excess fees matter (Sirri and Tufano (1998), and Barber et al. (2005)). As regards active share, Cremers et al. (2016) also find that this is related to fund flows. However, their results rely on the time-series variation at the benchmark level while we use time-series variation at the fund level, meaning that our coefficient estimates are not directly comparable.

In the next section, we elaborate on the flow-fee relationship by, among other things, comparing the effect size with that of Morningstar ratings.

3.5 Further analysis

In this section, we perform a number of additional analyses to deepen the understanding of the results from the main analysis above. Firstly, we examine how the above estimates are related to Morningstar ratings before going on to examine differences in the flow-fee relation across fund characteristics.

3.5.1 Morningstar ratings and the flow-fee relation

The Morningstar rating, also known as "star ratings", is a purely quantitative, backward-looking measure of a fund's past performance, measured from one to five stars each month.[17] Our quarterly ratings are the last observed rating in each quarter.

First, we test how the Morningstar ratings are related to excess fee, active share, and active fee. We run the baseline regressions in Equation (3.7) with ratings as the dependent variable. The results are shown in Table 3.5.1. We find that the fee variables are statistically significant, and have the same sign as the flow-fee regressions. Active share has the opposite sign of the regressions in the last section without control variables, and changes sign and becomes statistically insignificant with other control variables. However, with the Morningstar rating ranging from 1-5 and with a sample mean of 3.07, these effects are comparably small. Using the within-fund standard errors of the test variables from Table 3.3.1 to interpret the effects, we argue that the magnitude of the effect on rating of a one standard deviation change in the variables is small. The common idea that the more you pay, the better product you get, is therefore not valid here.

Next, we include ratings as a control variable in the regression in Equation (3.7), both as a stand-alone variable and as an interaction variable to examine whether our findings from the main tests hold, also when controlling for Morningstar ratings. Since Morningstar ratings are performance based, this estimate can be seen as a flow-performance test. We argue that by controlling for the ratings, we improve the identification in two ways. Firstly, it extracts the effect of ratings, giving a cleaner estimate of the flow-fee relationship. Secondly, it allows the economic significance of the estimated flow-fee relationship to be interpreted relative to the estimated flow-rating relationship.[18]

[17] For more information about the Morningstar Star Ratings see: Morningstar Ratings.

[18] Multiple studies document a strong relationship between Morningstar ratings and flow (see, e.g., Del Guercio and Tkac (2008),Ben-David et al. (2022) and Evans and Sun (2021)). Moreover, ratings are highly available for investors and thereby function as a benchmark for how important our fee variables are for flows.

Table 3.5.1.
Morningstar rating-main variables relationship

This table shows the estimated slope coefficients ($\hat{\gamma}_{EF}$, $\hat{\gamma}_{AS}$, and $\hat{\gamma}_{AF}$) from panel regressions testing the relationship between Morningstar ratings and our main fee variables on the form presented in Equation (3.7). Columns 1-2 present the results for excess fee, Columns 3-4 for active share, and Columns 5-6 for active fee. All regressions include fund fixed effects. Other control variables are described in Table 3.A.1 in Appendix 3.A.1. Standard errors (in parentheses) are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Morningstar rating$_t$					
X	Excess fee		Active share		Active fee	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{X}_{t-1}$	-0.82^{***} (0.08)	-0.23^{***} (0.08)	-0.001^{**} (0.001)	0.001 (0.001)	-0.34^{***} (0.05)	-0.12^{***} (0.05)
Other controls		×		×		×
Fund FE	×	×	×	×	×	×
Adjusted R^2	0.54	0.57	0.56	0.58	0.56	0.58
Observations	318,533	314,150	233,377	233,373	233,007	233,003

With interaction variables, we test whether the flow-fee relation is different in funds with a high and low rating. High-rated funds are funds with 4 or 5 stars and low-rated funds are funds with 1 or 2 stars, leaving medium-rated funds (3 stars) at the base of the regression. Using this method, we test whether there are differences in how those investing in funds with different star ratings are influenced by fees when selecting funds.

The regression results are presented in Table 3.5.2. In Columns 1 and 2 we test excess fee, with lagged ratings as an additional control variable in Column 1 and interaction variables in Column 2. We find that the negative relation between excess fee and subsequent net flow still holds when controlling for ratings. Moreover, we find a similar coefficient for the mid-rated funds, and the coefficients of the interaction variables show no statistically significant difference in the flow-excess fee response in the fund groups. We also run an F-test of the difference in the coefficient of the interaction with high and low rated funds and excess fee, but the difference is not statistically significant. Comparing the coefficient of excess fee with the corresponding coefficient in Table 3.4.1, the magnitude of the coefficient is almost twice as large.

Columns 3 and 4 present the equivalent results for active share. Again, our main results hold in terms of statistical significance, but the magnitude is reduced to around half of the corresponding magnitude in the main results. In Column 4, we find that the effect is larger for funds with high ratings and smaller for funds with low ratings. The difference between the interaction terms is statistically significant at the 1% level.

Finally, we test active fee in Columns 5 and 6. Again, the statistical significance holds, and the magnitude of the coefficient increases somewhat compared to the regression in Table 3.4.1. This slight increase is a result of the increase in magnitude from the flow-excess fee relation and the reduction of magnitude in the flow-active share relation. The relation is similar for the funds with a 3-star rating, and the coefficients of the interaction variables show that the negative relation is stronger in magnitude for the high-rated funds and weaker for the low-rated funds. As regards active share, the difference between the interaction terms is statistically significant at the 1% level.

In line with the main results in the previous section, we add domicile-quarter fixed effects to examine whether this affects the results. The regressions are presented in Table 3.C.2. The addition of time fixed effects does not change the inference and leads to minor changes in the coefficient sizes.

The estimated effects of fee on flow are statistically significant. However, the economic significance of the results in Table 3.4.1 and 3.5.2 is modest. The size of the effects is more or less identical to the findings of Barber et al. (2005).[19] Therefore, we also interpret the results by benchmarking the effects to corresponding estimated effects of Morningstar ratings. We multiply the coefficients of $\mathbf{X}_{t-1}$ and $Rating_{t-1}$ from Table 3.5.2 by the within-fund standard deviations to compare the effects.

Table 3.5.3 presents the estimated effects of fee variables and Morningstar ratings on flow, as well as the ratio of the fee estimate in comparison with the Morningstar estimate. Excess fee and active fee constitute around 12% of the Morningstar rating, while active share explains less. Given the documented strength of Morningstar ratings' ability to predict flows, together with the fact that the net flow variables typically contain a lot of noise, we argue that fees still explain a fair share of the subsequent net flows.[20]

The results from these tests show a statistically significant relationship between Morningstar ratings and fee variables, but the economic significance of this relationship is moderate. Furthermore, the main results hold when we add Morningstar ratings to the vector of control variables.

[19] They find that a 100 basis-point decrease in expense ratio leads to an increase of 0.39% in net flow the following quarter.

[20] A comparison of the coefficient size from regressions testing the fee and Morningstar ratings independently yields similar results with control variables and a higher ratio without control variables. This ratio can be obtained by comparing the effects from Table 3.4.1 with the effects in Table 3.C.1.

Table 3.5.2.
Net flow-main variables relationship and Morningstar ratings

This table reports estimated slope coefficients ($\hat{\gamma}_{EF}$, $\hat{\gamma}_{AS}$, and $\hat{\gamma}_{AF}$) from panel regressions testing the relationship between net flow and our main fee variables when controlling for Morningstar ratings, on the form presented in Equation (3.7). Columns 1-2 present the results for excess fee, Columns 3-4 for active share, and Columns 5-6 for active fee. The regressions in Columns 1, 3, and 5 add lagged Morningstar rating as a control variable. The regressions in Columns 2, 4, and 6 add interaction terms for funds with high and low Morningstar ratings. The difference and associated p-value is a test of the difference in the interaction terms' coefficients. All regressions include fund fixed effects. Other control variables are described in Table 3.A.1 in Appendix 3.A.1. Standard errors (in parentheses) are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance:
$^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow$_t$					
X	Excess fee		Active share		Active fee	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{X}_{t-1}$	-3.64^{***}	-3.61^{***}	0.01^{***}	0.01^{**}	-2.44^{***}	-2.35^{***}
	(0.84)	(0.90)	(0.004)	(0.005)	(0.48)	(0.51)
Rating$_{t-1}$	1.91^{***}		2.05^{***}		2.05^{***}	
	(0.07)		(0.07)		(0.07)	
High$_{t-1}$		2.88^{***}		0.91^{***}		3.29^{***}
		(0.17)		(0.35)		(0.18)
$\mathbf{X}_{t-1} \cdot$ High$_{t-1}$		-0.82		0.02^{***}		-1.54^{***}
		(0.58)		(0.005)		(0.45)
Low$_{t-1}$		-1.61^{***}		-0.48		-1.99^{***}
		(0.17)		(0.31)		(0.18)
$\mathbf{X}_{t-1} \cdot$ Low$_{t-1}$		0.38		-0.01^{***}		1.06^{***}
		(0.52)		(0.004)		(0.38)
Difference		-1.20		0.04^{***}		-2.61^{***}
P-val.		0.11		<0.01		<0.01
Other controls	×	×	×	×	×	×
Fund FE	×	×	×	×	×	×
Adjusted R^2	0.11	0.11	0.12	0.11	0.12	0.11
Observations	308,970	308,970	230,360	230,360	229,989	229,989

When comparing the effects of a one standard deviation change in the fee variables to a one standard deviation change in ratings, we find that the fee variables explain around 12% of ratings. This is an interesting finding, as fees are more informative when it comes to future performance than stars based on past performance. Many studies show that performance is not particularly persistent, and past performance is not therefore a good signal of future performance. Moreover, the funds themselves can set the fee and active share, while they cannot set the performance.

Table 3.5.3.
Effect of main variables and Morningstar ratings on flow

This table reports the estimated effects of a one within-fund standard deviation change in excess fee, active share, active fee, and Morningstar rating for the slope estimates with other control variables in Table 3.5.2. The within-fund standard deviations are presented in Table 3.3.1. **X**/Rating is how much of the variation the fee variables explain presented as a share of how much variation the Morningstar ratings explain.

X	Excess fee	Active share	Active fee
X estimate (%)	0.15	0.06	0.15
Rating estimate (%)	1.20	1.29	1.29
X/Rating	0.125	0.046	0.116

This makes it more difficult for the funds to attract flows through the performance channel than through the fee and active share channel.

Moreover, we find evidence of differences in the effects across funds based on ratings, which suggests that investors in these groups of funds use the signals from our model differently. In general, investors in highly rated funds appear to respond more to active share and active fee than investors in funds with a low rating. One interpretation is that investors in funds with a high rating are more financially sophisticated or rational. In the same way, investors in funds with a low rating are less inclined to collect information and respond to changes in active fee.

These results show how well-grounded and important Morningstar ratings are in the mutual fund industry, and their importance as input for investors. They are also highly accessible. Based on our model, however, we would claim that fee signals are more rational than Morningstar ratings, as they are more informative when it comes to future performance.

3.5.2 Fund characteristics

In this section, we test differences in the flow-fee relation based on the fund characteristics of past performance, fund size, and fund style. In general, fund performance is not particularly persistent (see Carhart (1997)). Despite this lack of persistence, multiple studies document that fund flows chase funds with a high past performance (see e.g., Ippolito (1992), Chevalier and Ellison (1997), Sirri and Tufano (1998), and Del Guercio and Tkac (2002)). Ferreira et al. (2012) show in an international study that the flow-performance relationship is most present in developed countries.[21] In this study, we group the funds into terciles based on the benchmark adjusted returns over the previous year, and the funds are sorted within each country-benchmark segment to determine the rankings.

Fund size is related to the ex-ante potential of the fund to outperform the benchmark by decreasing returns to scale. Several studies show that larger funds often underperform relative to smaller funds, and if investors are aware of this, the relation between fund flow and fee might vary across fund sizes. As suggested by the theoretical model of Berk and Green (2004), funds have decreasing returns to scale. This is also documented in several empirical studies in the mutual fund literature (see e.g., Pástor et al. (2015) and Zhu (2018)). We can relate this feature to our setting, with a lower skill for large funds than small funds. This relation implies a less negative relation between fees and fund flow for larger funds. To test this prediction, we sort funds into terciles based on the fund TNA at the end of the previous quarter within each domicile.

Table 3.5.4 reports the regression results. The reported slopes on excess fee ($\hat{\gamma}_{EF}$), active share ($\hat{\gamma}_{AS}$), and active fee ($\hat{\gamma}_{AF}$) are in these regressions the slopes of the mid-groups based on performance and size.

Columns 1 and 2 show the results for excess fee. We find that the negative flow-excess fee relationship is weaker in magnitude for funds with a high past performance. This implies that when investors chase past performance, they care less about the cost of the fund. The difference in interaction terms is significant at the 10% level. As regards fund size, we find that the negative flow-fee relationship is weaker in magnitude for large funds. Columns 3 and 4 present the corresponding results for active share. Here we find that the positive flow-active share relation is stronger in magnitude for high-performing funds and weaker for low-performing funds, with the difference in the interaction terms significant at the 1% level. For size, there is a weak result showing that larger funds have a stronger positive flow-active share relation. The results for active fee, presented in Columns 5 and 6 show that the negative relation is stronger in magnitude for funds with a high past performance and weaker in magnitude for funds with a low past performance.

The coefficients for the stand-alone dummy variables based on performance, $High_{t-1}$ and Low_{t-1}, which can be interpreted as the average flow into these funds relative to the middle

[21] Since Morningstar ratings are based on past performance, these tests are highly related to the tests in the previous section, but the ranking of past performance across funds is not as accessible as for Morningstar ratings. In addition, Morningstar ratings are based on a longer historical period, while we examine performance in the past quarter.

group of funds, are highly significant and of opposite signs. These findings confirm the flow-performance relationship of previous studies, where money flows into the top performers and out of the bottom performers. As regards fund size, the results are less clear when it comes to average flow.

The results in this section show some evidence on differences in the flow-fee relationship across funds based on fund characteristics. However, for excess fee and active fee, the relation regarding high and low past performance is of opposite signs. This implies that investors chasing past performance also care about active share. For past performance and excess fee, the results here are the opposite of the corresponding results for Morningstar ratings in the previous section, which may be due to of the differences in methodology.[22] In relation to size, we find weak evidence of differences in the relations to flow for large funds. For fee and active fee, the estimated slope on the interaction term is positive, hinting at a stronger flow-fee relation for larger funds, but this is only statistically significant for excess fees.

It may be argued that if investors primarily chase past performance, they might care less about other attributes. However, our results show that this is not necessarily the case, and that fees remain an important determinant. We do not find strong evidence suggesting that investors in small and large funds use the signals from excess fee, active share, and active fee differently to select funds. The model of Berk and Green (2004) suggests that larger funds have lower skill or potential to outperform the benchmark, but it does not seem to be the case that this information is used in interaction with ex-ante fee variables in fund selection.

Next, we examine differences in the flow-fee relation across funds based on the style of their strategies. To classify funds based on style, we use the Morningstar 3x3 style-boxes.[23] In the tests, we focus on the styles independently, i.e., large vs. small-cap funds and value vs. growth funds. We are agnostic here and do not form any specific predictions.

We add interaction variables between our main variables and the style-dummies to the regression in Equation (3.7). The regression results are presented in Table 3.5.5, with the mid-groups forming the base of the regression.[24] Columns 1 and 2 present the results for excess fee. Here we find that the negative flow-fee relationship is weaker for growth funds. For active share, shown in Columns 3 and 4, the positive flow-active share is weaker for large-cap funds and

[22] Although the ratings are based on past performance, past performance in this section is limited to the past quarter, while Morningstar ratings are based on a longer period of performance.

[23] The nine styles in the style-box include small-cap value, small-cap blend, small-cap growth, mid-cap value, mid-cap blend, mid-cap growth, large-cap value, large-cap blend, and large-cap growth. See the Morningstar Style-Box for more information.

[24] The stand-alone dummies are soaked up by fund fixed effects since they do not vary over time.

Table 3.5.4.
Net flow-main variables relationship and fund characteristics

This table shows the estimated slope coefficients ($\hat{\gamma}_{EF}$, $\hat{\gamma}_{AS}$, and $\hat{\gamma}_{AF}$) from panel regressions testing the relationship between net flow and our main fee variables across funds based on fund characteristics, on the form presented in Equation (3.7). Columns 1-2 present the results for excess fee, Columns 3-4 for active share, and Columns 5-6 for active fee. The regressions in Columns 1, 3, and 5 add interaction terms for funds with a high and low past performance. The regressions in Columns 2, 4, and 6 add interaction terms for large and small funds in terms of TNA. The difference and associated p-value is a test of the difference in the interaction terms' coefficients. All regressions include fund fixed effects. Other control variables are described in Table 3.A.1 in Appendix 3.A.1. Standard errors (in parentheses) are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow$_t$					
X	Excess fee		Active share		Active fee	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{X}_{t-1}$	-3.57^{***}	-2.36^{**}	0.02^{***}	0.01^{**}	-2.74^{***}	-2.20^{***}
	(0.92)	(0.98)	(0.005)	(0.01)	(0.53)	(0.59)
High$_{t-1}$	1.53^{***}		0.07		1.99^{***}	
	(0.13)		(0.28)		(0.14)	
$\mathbf{X}_{t-1} \cdot$ High$_{t-1}$	0.96^{**}		0.02^{***}		-0.67^{*}	
	(0.48)		(0.004)		(0.36)	
Low$_{t-1}$	-1.05^{***}		-0.14		-1.38^{***}	
	(0.11)		(0.25)		(0.12)	
$\mathbf{X}_{t-1} \cdot$ Low$_{t-1}$	-0.09		-0.01^{***}		0.70^{**}	
	(0.38)		(0.003)		(0.29)	
Small$_{t-1}$		-0.44^{*}		-0.97^{*}		-0.35
		(0.25)		(0.51)		(0.26)
$\mathbf{X}_{t-1} \cdot$ Small$_{t-1}$		0.24		0.01		-0.11
		(0.77)		(0.01)		(0.57)
Large$_{t-1}$		0.41^{*}		0.09		0.72^{***}
		(0.24)		(0.44)		(0.26)
$\mathbf{X}_{t-1} \cdot$ Large$_{t-1}$		1.68^{**}		0.01^{*}		0.57
		(0.80)		(0.01)		(0.52)
Difference	1.04^{*}	-1.44	0.04^{***}	-0.00	-1.37^{***}	-0.68
P-val.	0.07	0.19	<0.01	0.68	<0.01	0.33
Other controls	×	×	×	×	×	×
Fund FE	×	×	×	×	×	×
Adjusted R^2	0.12	0.14	0.13	0.13	0.13	0.13
Observations	364,126	401,940	261,954	280,200	260,864	279,040

stronger for value funds. In Columns 5 and 6, we find that the negative flow-active fee relation is stronger for value funds and weaker for growth funds. We also note that in most regressions, the interaction variable coefficients have opposite signs, and the tests of the differences are statistically significant.

Table 3.5.5.
Net flow-main variables relationship and fund style

This table shows the estimated slope coefficients ($\hat{\gamma}_{EF}$, $\hat{\gamma}_{AS}$, and $\hat{\gamma}_{AF}$) from panel regressions testing the relationship between net flow and our main fee variables across funds based on fund style, on the form presented in Equation (3.7). Columns 1-2 present the results for excess fee, Columns 3-4 for active share, and Columns 5-6 for active fee. The regressions in Columns 1, 3, and 5 add interaction terms for small-cap and large-cap funds. The regressions in Columns 2, 4, and 6 add interaction terms for value and growth funds. The difference and associated p-value is a test of the difference in the interaction terms' coefficients. All regressions include fund fixed effects. Other control variables are described in Table 3.A.1 in Appendix 3.A.1. Standard errors (in parentheses) are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow$_t$					
X	Excess fee		Active share		Active fee	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{X}_{t-1}$	-1.62	-2.38^{**}	0.05^{***}	0.01^{**}	-2.20^{***}	-2.26^{***}
	(1.40)	(1.14)	(0.01)	(0.01)	(0.79)	(0.65)
$\mathbf{X}_{t-1}$ · Small-cap	-4.22		-0.03		-3.09	
	(2.57)		(0.02)		(1.95)	
$\mathbf{X}_{t-1}$ · Large-cap	0.53		-0.03^{***}		0.38	
	(1.50)		(0.01)		(0.86)	
$\mathbf{X}_{t-1}$ · Value		-1.33		0.02^{**}		-1.77^{*}
		(1.61)		(0.01)		(0.90)
$\mathbf{X}_{t-1}$ · Growth		3.33^{**}		-0.004		2.30^{**}
		(1.55)		(0.01)		(0.96)
Difference	-4.76^{*}	-4.66^{**}	0.00	0.02^{**}	-3.47^{*}	-4.07^{***}
P-val.	0.05	0.01	0.86	0.03	0.07	<0.01
Other controls	×	×	×	×	×	×
Fund FE	×	×	×	×	×	×
Adjusted R^2	0.14	0.14	0.13	0.13	0.13	0.14
Observations	382,616	382,616	266,096	266,096	265,717	265,717

In sum, the findings of these tests show that there are differences in the active fee-fund flow relation dependent on fund style. The opposite signs of the interaction coefficients show that there may be differences in the investor base of these types of funds, based on their ability to behave rationally, in line with the model derived in this paper, and follow signals from excess fee, active share, and active fee when buying and selling funds.

3.6 Conclusion

This paper develops a simple partial model showing that excess and active fees can be a signal for future value creation and hypothesizes that investors will follow the model when selecting their mutual funds. Our empirical results are consistent with our predictions. Excess fee is negatively related to subsequent quarterly net flows, and active share is positively related to subsequent quarterly net flows, leading to a negative flow-active fee relationship. Thus, the relation between active fee and flow comes from both the excess fee and the activity level.

We examine the relation further by introducing Morningstar ratings as a control variable and compare the effects to the corresponding effects of Morningstar ratings, as well as examining differences in the relation across funds based on fund characteristics and style. We highlight the following three results. Firstly, the fee variables are only moderately related to Morningstar ratings, and our main results hold statistically when controlling for Morningstar ratings. A one within-fund standard change in excess fee and active fee accounts for around 10% of a one standard deviation change in ratings. Secondly, investors in funds with high ratings are more rational, following the fee-signals to a larger extent. Thirdly, investors in value funds are more price sensitive than investors in growth funds.

We demonstrate that in a market where it is hard for both researchers and investors to identify what creates value for the investors, at least some investors use a "rational" signal such as active fee. The fee variables are decided ex-ante by the funds themselves and are not decided based on historical performance by an external agency. We therefore argue that fee variables are more informative of future performance than Morningstar ratings, and that active mutual fund investors should therefore have responded more than they were shown to do.

It is not easy to regulate financial markets, but active fee should perhaps form part of investor information. As suggested by Cremers and Curtis (2016), disclosure of active fee can, for example, help to prevent closet-indexing.[25] Moreover, since Morningstar ratings are based on a different information set than active fee, a combination of both can be valuable.This question should be the subject of future research.

[25] The alternative to disclosing this type of information ex-ante is a potentially costly ex-post intervention and correction by regulatory authorities. For an example of the latter, see Bjerksund et al. (2020).

References

Barber, B. M., Huang, X., and Odean, T. (2016). Which Factors Matter to Investors? Evidence from Mutual Fund Flows. *The Review of Financial Studies*, 29(10):2600–2642.

Barber, B. M., Odean, T., and Zheng, L. (2005). Out of sight, out of mind: The effects of expenses on mutual fund flows. *The Journal of Business*, 78(6):2095–2120.

Ben-David, I., Li, J., Rossi, A., and Song, Y. (2022). What do mutual fund investors really care about? *The Review of Financial Studies*, 35(4):1723–1774.

Berk, J. B. and Green, R. C. (2004). Mutual fund flows and performance in rational markets. *Journal of political economy*, 112(6):1269–1295.

Berk, J. B. and Van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(1):1–20.

Berk, J. B. and van Binsbergen, J. H. (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, 119(1):1–23.

Bjerksund, P., Doskeland, T., Sjuve, A. W., and Ørpetveit, A. (2020). Forced to be active: Evidence from a regulation intervention. *Available at SSRN 3635718*.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

Carlin, B. I. (2009). Strategic price complexity in retail financial markets. *Journal of financial Economics*, 91(3):278–287.

Chevalier, J. and Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of political economy*, 105(6):1167–1200.

Choi, J. J. and Robertson, A. Z. (2020). What matters to individual investors? evidence from the horse's mouth. *The Journal of Finance*, 75(4):1965–2020.

Cooper, M. J., Halling, M., and Yang, W. (2021). The persistence of fee dispersion among mutual funds. *Review of Finance*, 25(2):365–402.

Coval, J. and Stafford, E. (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 86(2):479–512.

Cremers, K. and Curtis, Q. (2016). Do mutual fund investors get what they pay for: Securities law and closet index funds. *Va. L. & Bus. Rev.*, 11:31.

Cremers, K. M. and Petajisto, A. (2009). How active is your fund manager? a new measure that predicts performance. *The review of financial studies*, 22(9):3329–3365.

Cremers, M., Ferreira, M. A., Matos, P., and Starks, L. (2016). Indexing and active fund management: International evidence. *Journal of Financial Economics*, 120(3):539–560.

Del Guercio, D. and Tkac, P. A. (2002). The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds. *Journal of Financial and Quantitative Analysis*, pages 523–557.

Del Guercio, D. and Tkac, P. A. (2008). Star power: The effect of monrningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis*, 43(4):907–936.

Evans, R. B. and Sun, Y. (2021). Models or Stars: The Role of Asset Pricing Models and Heuristics in Investor Risk Adjustment. *The Review of Financial Studies*, 34(1):67–107.

Factbook, M. F. (2021). Investment company institute. Available at `https://www.ici.org/system/files/2021-05/2021_factbook.pdf`.

Ferreira, M. A., Keswani, A., Miguel, A. F., and Ramos, S. B. (2012). The flow-performance relationship around the world. *Journal of Banking & Finance*, 36(6):1759–1780.

Garleanu, N. and Pedersen, L. H. (2019). Active and passive investing. *Available at SSRN 3253537*.

Hartzmark, S. M. and Sussman, A. B. (2019). Do investors value sustainability? a natural experiment examining ranking and fund flows. *The Journal of Finance*, 74(6):2789–2837.

Huang, C., Li, F., and Weng, X. (2020). Star ratings and the incentives of mutual funds. *The Journal of Finance*, 75(3):1715–1765.

Ilmanen, A. (2022). *Investing Amid Low Expected Returns: Making the Most When Markets Offer the Least*. John Wiley & Sons.

Ippolito, R. A. (1992). Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. *The Journal of Law and Economics*, 35(1):45–70.

Kaniel, R. and Parham, R. (2017). Wsj category kings – the impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics*, 123(2):337–356.

Marmoiton, A. (2021). An approximate formula linking active share and tracking error. *Available at SSRN 3771144*.

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

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2017). Do funds make more when they trade more? *The Journal of Finance*, 72(4):1483–1528.

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2020). Fund tradeoffs. *Journal of Financial Economics*, 138(3):614–634.

Schumacher, D. (2018). Home bias abroad: Domestic industries and foreign portfolio choice. *The Review of Financial Studies*, 31(5):1654–1706.

Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *The journal of finance*, 53(5):1589–1622.

Zhu, M. (2018). Informative fund size, managerial skill, and investor rationality. *Journal of Financial Economics*, 130(1):114–134.

3.A Research design

This part of the appendix contains additional tests relating to Section 3.3.

3.A.1 Variable definitions

Table 3.A.1 describes the variables used in the tests. Other control variables are the variables included in the main regressions, and the additional control variables are variables used in robustness tests relating to fund ratings and style.

Table 3.A.1.
Variable definitions

This table documents our variables, their definitions, unit of measurement, and frequency. We make a distinction between our main variables and other controls.

Variable	Definition
Main outcome variable	
Net flow	Percentage growth in TNA (in USD), net of internal growth (assuming reinvestment of dividends and distributions). See Section 3.3.1, Equation (3.4) for details.
Main test variables	
Excess fee	Total cost of active management, fund fee less comparable index fee. See Section 3.3.1 for details.
Active share	Percentage of a fund's portfolio holdings that differ from its benchmark index holdings.
Active fee	Unit cost of active management, excess fee over activity level. See Section 3.3.1 for details.
Other control variables	
Benchmark-adjusted return (gross alpha)	Difference between a fund's gross return and its benchmark return.
TNA	Total net assets in millions of U.S. dollars. Used in log form in panel regressions.
Fund age	Number of years since the fund's launch date. Used in log form in panel regressions.
Family TNA	Family total net assets in millions of U.S. dollars of other equity funds in the same management company excluding the own fund TNA. Used in log form in panel regressions.
Fund industry size	Sum of TNA for all funds within each domicile-year. Used in log form in panel regressions.
Fund industry competition	1 - HHI, where HHI is the sum of squared market shares of fund management companies for mutual funds in the fund's country.
Additional control variables	
Morningstar rating	A purely quantitative, backward-looking measure of a fund's past performance, measured from one to five stars. Star ratings are calculated at the end of every month. See Morningstar Ratings for more information.
Morningstar style-box	A style-based measure based on the tilts toward size (Small, Mid-Cap, or Large) and value tilt (Value, Blend, or Growth) of the fund's actual portfolio holdings. See Morningstar Style-Box for more information.

3.A.2 Sample details

Table 3.A.2 presents summary statistics by country. We note that there is a large variation in both active fee and net flows across our global sample of mutual funds. At one end of the range, we have countries such as Canada and Ireland that have experienced an average net inflow of capital over the sample period, while countries such as Japan and Taiwan have large average outflows. Across countries, we see a negative trend in an average country's net outflow of 2.1%

per year over the sample period.[26] The active fee is, as net flow, widely dispersed, ranging from an average of 0.5% in the USA to 2.2% in Italy, i.e., a spread of 170 bps between the countries. Overall, the average level of active fee is 1.4% per year. For excess fee, i.e., the total cost of active management minus indexing, the global country average is 1% per year. Also here, the USA is the country with the lowest fee, charging on average only 40 bps per year for active management. Active share at the country level is in the interval between 66% (Sweden) and 86% (China).

Table 3.A.2.
Summary statistics per country

This table presents summary statistics on the number of actively managed equity mutual funds, the total net assets as of December 2019 in USDbn, the average net flow and average excess fee, active share, and active fee per country.

Country	Funds	N	TNA	Net flow (%)	Excess fee (%)	Active share (%)	Active fee (%)
Australia	109	4,454	21	1.44	0.66	69	0.93
Austria	209	8,546	13	0.11	1.00	76	1.42
Belgium	163	6,447	22	0.24	0.66	68	1.12
Canada	1,198	44,324	248	1.30	1.47	82	1.82
Chile	76	3,146	1.68	2.81	2.53	77	3.55
Denmark	235	9,884	31	-0.32	0.75	70	1.19
Finland	184	7,155	21	1.35	1.02	78	1.41
France	750	30,135	100	-0.05	0.97	74	1.37
Germany	428	15,106	98	-0.78	0.84	75	1.20
Hong Kong	109	4,643	31	0.13	0.74	75	1.02
India	290	12,088	85	0.14	1.28	74	1.85
Ireland	374	13,214	59	-0.42	0.93	77	1.33
Italy	117	4,031	12	-1.41	1.27	69	2.05
Japan	797	34,632	67	-2.97	0.81	76	1.13
Liechtenstein	101	3,739	5.85	0.25	1.22	84	1.53
Luxembourg	65	2,880	26	-0.24	0.80	79	1.01
Malaysia	170	7,427	6.14	-0.82	0.57	83	0.73
Netherlands	86	3,370	16	-1.06	0.53	76	0.72
Norway	143	5,872	31	0.97	0.81	78	1.20
Portugal	38	1,821	1.71	-1.36	1.11	72	1.30
Singapore	108	4,719	8.64	-1.66	0.92	80	1.17
South Africa	222	8,837	23	1.26	0.88	73	1.30
South Korea	381	13,775	11	-2.33	0.84	79	1.12
Spain	348	11,136	24	0.55	1.00	71	1.53
Sweden	208	8,399	103	1.13	0.74	66	1.25
Switzerland	237	9,060	20	-0.97	0.77	67	1.26
Taiwan	286	11,260	12	-2.69	1.53	81	1.83
Thailand	243	9,883	20	0.54	0.88	72	1.43
UK	1,278	46,058	380	0.33	0.80	78	1.08
USA	1,861	71,634	1,236	0.03	0.71	82	0.89
All countries	**10,814**	**417,675**	**2,734**	**–0.15**	**0.97**	**75**	**1.36**

[26] This decline resembles that identified in other sources; see, for example, Figure 3.14 in Factbook (2021).

3.A.3 Development in main variables

Figure 3.A.1.
Development of fees

This figure presents the development in average total expense ratio, index fee, and excess fee for the sample of funds.

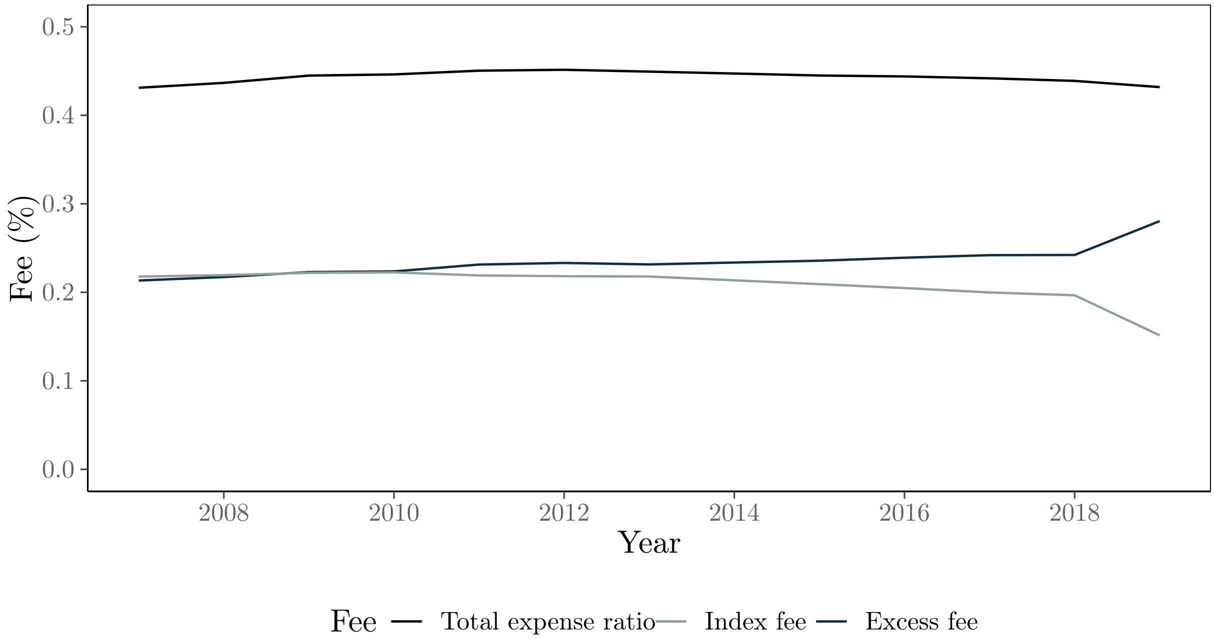


Figure 3.A.2.
Development in active share

This figure presents the development in average active share for the sample of funds.

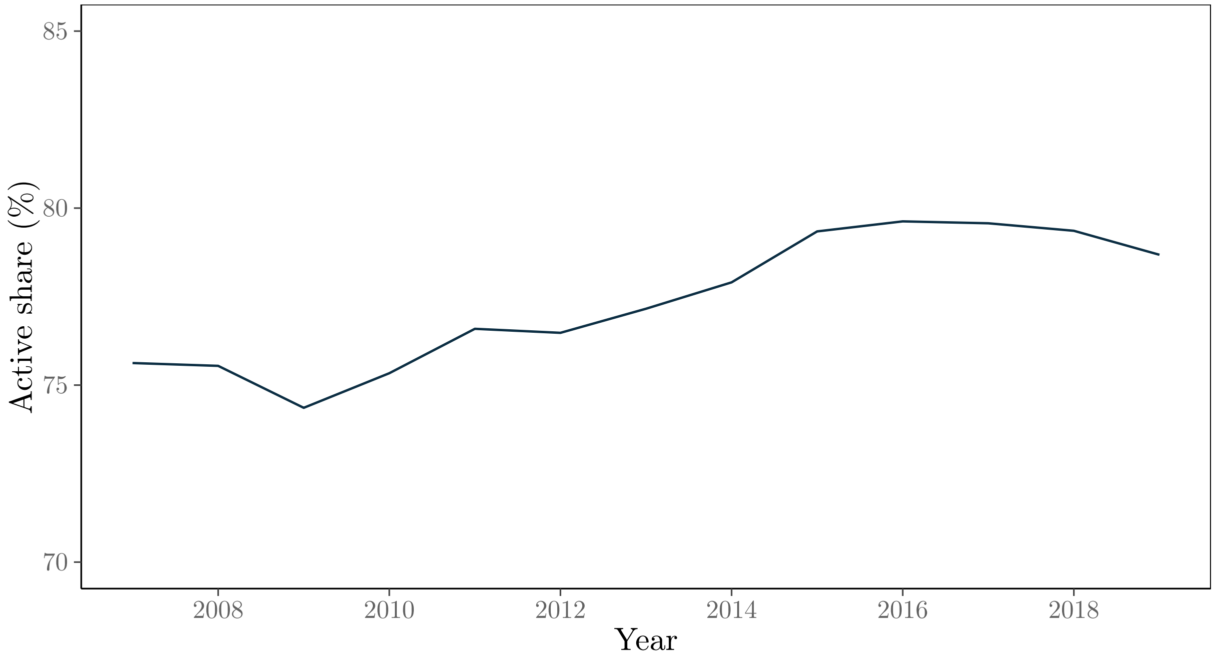
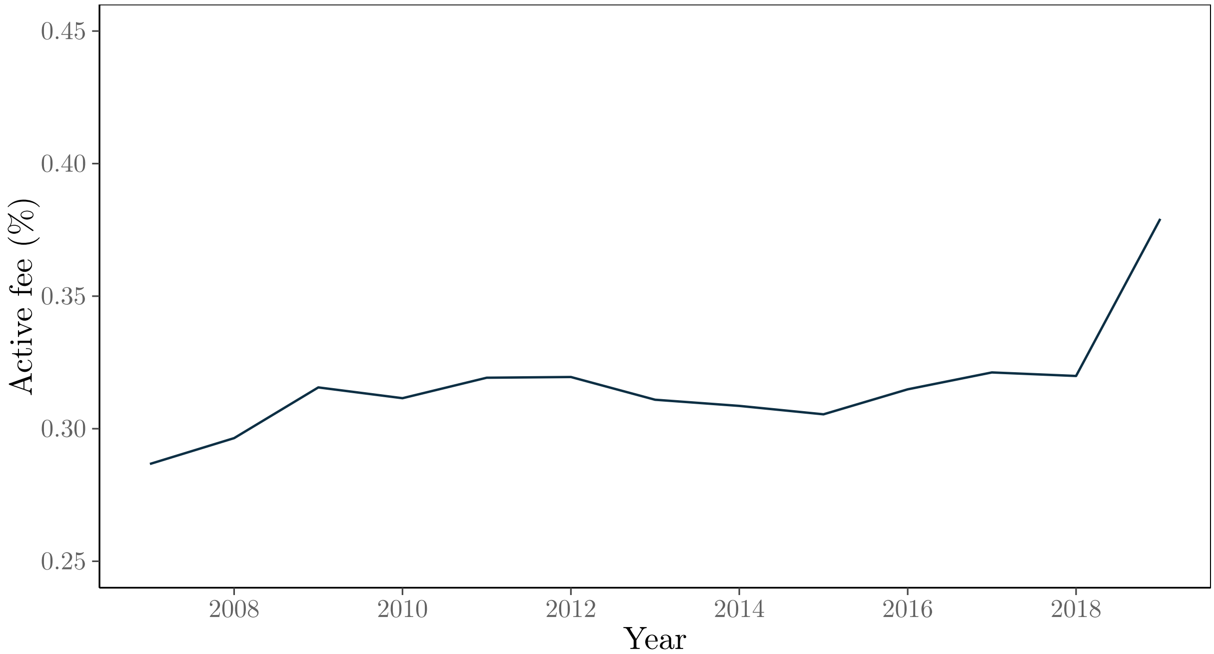


Figure 3.A.3.
Development in active fee

This figure presents the development in average active fee for the sample of funds.

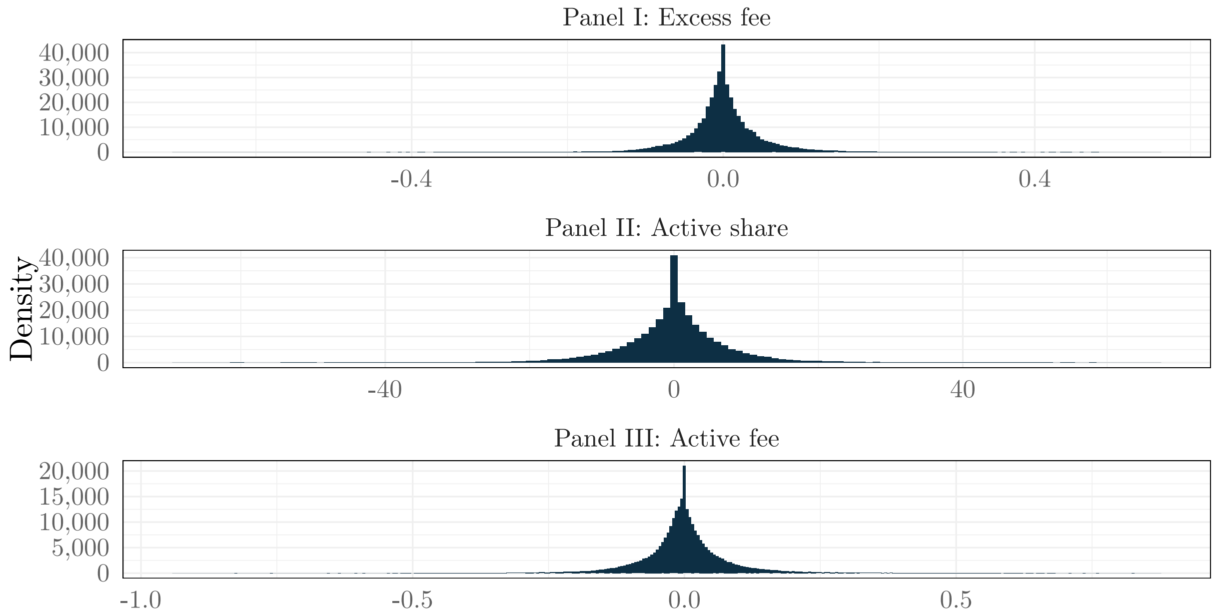


3.A.4 Distribution of main variables

In Figure 3.A.4 we present the distribution of the main fee variables used as dependent variables in the regressions. The variables are demeaned fund by fund to illustrate the distribution relevant to the main regressions with fund fixed effects.

Figure 3.A.4.
Distribution of main variables

This figure presents the distribution of the main fee variables of the paper. The variables are explained in Section 3.3.1. All distributions are deviations from the within-fund means. Panel I presents the distribution of excess, Panel II of active share, and Panel III of active fee.



3.B Main results: additional results

In this appendix, we provide additional analyses on the flow relation with our main variables as analyzed in Section 3.4. We firstly add time fixed effects to our main results for robustness purposes before testing whether the results also hold in the cross-section of funds.

3.B.1 Adding time fixed effects

Table 3.B.1.
Net flow-main variables relationship with time fixed effects

This table reports estimated slope coefficients ($\hat{\gamma}_{EF}$, $\hat{\gamma}_{AS}$, and $\hat{\gamma}_{AF}$) from panel regressions testing the relationship between net flow and our main fee variables, on the form presented in Equation (3.7). Columns 1-2 present the results for excess fee, Columns 3-4 for active share, and Columns 5-6 for active fee. All regressions include fund and time fixed effects. Other control variables are described in Table 3.A.1 in Appendix 3.A.1. Standard errors (in parentheses) are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow$_t$		
X	Excess fee (1)	Active share (2)	Active fee (3)
$\mathbf{X}_{t-1}$	-3.11^{***} (0.79)	0.02^{***} (0.004)	-2.05^{***} (0.42)
Other controls	×	×	×
Fund FE	×	×	×
Time FE	×	×	×
Adjusted R^2	0.18	0.17	0.17
Observations	401,940	280,200	279,040

3.B.2 Cross-sectional regressions

Table 3.B.2.
Net flow-main variables relationship

This table reports estimated slope coefficients ($\hat{\gamma}_{EF}$, $\hat{\gamma}_{AS}$, and $\hat{\gamma}_{AF}$) from cross-sectional regressions testing the relationship between net flow and our main fee variables. Columns 1-2 present the results for excess fee, Columns 3-4 for active share, and Columns 5-6 for active fee. All regressions include time fixed effects. Other control variables are described in Table 3.A.1 in Appendix 3.A.1. Standard errors (in parentheses) are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance:
$^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

	Net flow$_t$					
X	Excess fee		Active share		Active fee	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{X}_{t-1}$	-0.87^{***}	-2.44^{***}	0.02^{***}	0.004	-1.65^{***}	-1.51^{***}
	(0.33)	(0.33)	(0.002)	(0.002)	(0.22)	(0.21)
Other controls		×		×		×
Time FE	×	×	×	×	×	×
Adjusted R^2	0.05	0.09	0.05	0.08	0.05	0.08
Observations	401,940	401,940	280,200	280,200	279,040	279,040

3.C Robustness: additional results

This part of the appendix contains additional tests relating to Section 3.5.

3.C.1 Flow-rating relationship for comparison

Table 3.C.1.
Net flow-Morningstar rating relationship

This table presents the estimated slope coefficients from panel regressions testing the relationship between net flow and Morningstar star rating. All regressions include fund fixed effects. Other control variables are described in Table 3.A.1 in Appendix 3.A.1. Standard errors (in parentheses) are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance: $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

	Net flow$_t$	
	(1)	(2)
Star rating$_{t-1}$	1.71***	1.90***
	(0.06)	(0.06)
Other controls		×
Fund FE	×	×
Adjusted R^2	0.08	0.11
Observations	310,680	310,680

3.C.2 Adding time fixed effects

Table 3.C.2.
Net flow-main variables relationship and Morningstar ratings

This table reports the estimated slope coefficients from different panel regressions of net flows on excess fee. They differ in their treatment of fixed effects and inclusion of additional controls beyond excess fee. The control variables are described in Table 3.A.1 in Appendix 3.A.1. Net flow is fund i's net flow in year-quarter t. Active fee is fund i's excess fee in the quarter ending before net flow is measured. We report the level of fixed effects for each model. Standard errors are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow$_t$					
X	Excess fee		Active share		Active fee	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{X}_{t-1}$	-2.64^{***}	-2.68^{***}	0.01^{***}	0.01^{**}	-1.95^{***}	-1.91^{***}
	(0.70)	(0.75)	(0.004)	(0.004)	(0.39)	(0.43)
Rating$_{t-1}$	1.89^{***}		2.03^{***}		2.03^{***}	
	(0.06)		(0.07)		(0.07)	
High$_{t-1}$		2.80^{***}		1.09^{***}		3.25^{***}
		(0.16)		(0.33)		(0.18)
$\mathbf{X}_{t-1} \cdot$ High$_{t-1}$		-0.62		0.02^{***}		-1.45^{***}
		(0.55)		(0.004)		(0.43)
Low$_{t-1}$		-1.55^{***}		-0.57^{*}		-1.93^{***}
		(0.16)		(0.30)		(0.17)
$\mathbf{X}_{t-1} \cdot$ Low$_{t-1}$		0.23		-0.01^{***}		0.98^{***}
		(0.51)		(0.004)		(0.37)
Other controls	×	×	×	×	×	×
Fund FE	×	×	×	×	×	×
Time FE	×	×	×	×	×	×
Adjusted R^2	0.15	0.14	0.15	0.15	0.16	0.15
Observations	308,970	308,970	230,360	230,360	229,989	229,989

Table 3.C.3.
Net flow-main variables relationship and fund characteristics

This table shows the estimated slope coefficients from different panel regressions of net flows on excess fee. They differ in their treatment of fixed effects and inclusion of additional controls beyond excess fee. The control variables are described in Table 3.A.1 in Appendix 3.A.1. Net flow is fund i's net flow in year-quarter t. Active fee is fund i's excess fee in the quarter ending before net flow is measured. For each model we report the level of fixed effects. Standard errors are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance: $^{***}p<1\,\%$, $^{**}p<5\,\%$, $^{*}p<10\,\%$.

X	Net flow$_t$					
	Excess fee		Active share		Active fee	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{X}_{t-1}$	-3.36^{***}	-3.84^{***}	0.01^{***}	0.01^{**}	-2.27^{***}	-2.27^{***}
	(0.71)	(0.87)	(0.004)	(0.01)	(0.41)	(0.49)
High$_{t-1}$	1.36^{***}		0.02		1.79^{***}	
	(0.12)		(0.27)		(0.13)	
$\mathbf{X}_{t-1} \cdot$ High$_{t-1}$	0.91^{**}		0.02^{***}		-0.54	
	(0.45)		(0.004)		(0.33)	
Low$_{t-1}$	-1.06^{***}		0.03		-1.33^{***}	
	(0.11)		(0.27)		(0.12)	
$\mathbf{X}_{t-1} \cdot$ Low$_{t-1}$	0.48		-0.01^{***}		0.89^{***}	
	(0.39)		(0.003)		(0.30)	
Small$_{t-1}$		-0.49^{**}		-0.88^{*}		-0.21
		(0.22)		(0.50)		(0.23)
$\mathbf{X}_{t-1} \cdot$ Small$_{t-1}$		0.91		0.01		0.06
		(0.72)		(0.01)		(0.54)
Large$_{t-1}$		0.36^{*}		0.15		0.48^{**}
		(0.21)		(0.41)		(0.22)
$\mathbf{X}_{t-1} \cdot$ Large$_{t-1}$		1.14		0.01		0.62
		(0.75)		(0.01)		(0.48)
Other controls	×	×	×	×	×	×
Fund FE	×	×	×	×	×	×
Time FE	×	×	×	×	×	×
Adjusted R^2	0.16	0.18	0.16	0.17	0.16	0.17
Observations	364,126	401,940	261,954	280,200	260,864	279,040

Table 3.C.4.
Net flow-main variables relationship and fund style

This table shows the estimated slope coefficients from different panel regressions of net flows on excess fee. They differ in their treatment of fixed effects and inclusion of additional controls beyond excess fee. The control variables are described in Table 3.A.1 in Appendix 3.A.1. Net flow is fund i's net flow in year-quarter t. Active fee is fund i's excess fee in the quarter ending before net flow is measured. For each model we report the level of fixed effects. Standard errors are clustered along the fund and year-domicile-quarter dimension. Asterisks denote statistical significance: $^{***}p<1\%$, $^{**}p<5\%$, $^{*}p<10\%$.

	Net flow$_t$					
X	Excess fee		Active share		Active fee	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{X}_{t-1}$	-3.00^{**}	-3.09^{***}	0.03^{***}	0.01^{*}	-1.84^{***}	-1.88^{***}
	(1.32)	(1.06)	(0.01)	(0.01)	(0.69)	(0.60)
$\mathbf{X}_{t-1}$ · Small-cap	-2.74		-0.03		-2.02	
	(2.41)		(0.02)		(1.83)	
$\mathbf{X}_{t-1}$ · Large-cap	0.83		-0.02^{*}		0.04	
	(1.45)		(0.01)		(0.79)	
$\mathbf{X}_{t-1}$ · Value		-1.03		0.02^{**}		-1.34
		(1.57)		(0.01)		(0.85)
$\mathbf{X}_{t-1}$ · Growth		2.11		-0.01		1.23
		(1.47)		(0.01)		(0.88)
Other controls	×	×	×	×	×	×
Fund FE	×	×	×	×	×	×
Time FE	×	×	×	×	×	×
Adjusted R^2	0.18	0.18	0.17	0.17	0.17	0.17
Observations	382,616	382,616	266,096	266,096	265,717	265,717