Michael Cooper, Michael Halling and Wenhao Yang
1645 E. Campus Center Drive
Salt Lake City, UT 84112
December 16, 2016

SEC Investor Advisory Committee Members

Dear SEC Investor Advisory Committee Members:

We would like to submit our paper “The Mutual Fund Fee Puzzle” to the SEC Investor Advisory Committee, as we think that our empirical results are of relevance to the mutual fund cost disclosure reform.

Sincerely,

Michael Cooper, Michael Halling and Wenhao Yang

[redacted]

cc:

The Honorable Mary Jo White
Chairman
Securities and Exchange Commission

The Honorable Michael Piwowar
Commissioner
Securities and Exchange Commission

The Honorable Kara M. Stein
Commissioner
Securities and Exchange Commission

David Grim
Director
Division of Investment Management
Securities and Exchange Commission

The Mutual Fund Fee Puzzle

Michael Cooper, Michael Halling and Wenhao Yang*

October, 2016

Keywords: mutual funds, fund fees, fund expenses, price dispersion, price persistence.

JEL Classifications: G10, G11, G23.

---

* Cooper is with the David Eccles School of Business, University of Utah. Halling is with the Stockholm School of Economics and the Swedish House of Finance. Yang is with the University of South Carolina. We thank Yakov Amihud, Jack Bogle, Darwin Choi, James Choi (WFA discussant), Susan Christoffersen, Magnus Dahlquist, Monika Gehde-Trapp (discussant), Amit Goyal, Joseph Halford, Rachel Hayes, Keith Jacob, Stephan Jank (discussant), Greg Kadlec, E. Han Kim, Dong Lou (discussant), Veronika K. Pool (discussant), Jonathan Reuter, David Robinson, Norman Schuerhoff, Mark Seasholes, Henri Servaes, Paolo Sodini, Laura Starks, Paul Tetlock, Charles Trzcinka, and Sunil Wahal. We also thank participants of the CEPR ESSFM meetings (2011, asset pricing week), the Northern Finance Association Meetings (2011), the 2012 WFA meetings, the Austrian Working Group on Banking and Finance (2013), the John McConnell's Contributions to Finance Conference (2015), the 12th FBI Symposium in Karlsruhe (2014), and seminar participants at Arizona State, HEC Lausanne, the Helsinki Finance Seminar, the Institute of Advanced Studies in Vienna, KAIST, Nanyang Tech University, the Shanghai Advanced Institute of Finance, the University of Arizona, the University of Colorado, the University of Gothenburg, the University of Montana, the University of Luxembourg, the University of Virginia, the University of California, San Diego, and VU University Amsterdam for their comments. An earlier version of this paper was titled “Violations of the Law of One Fee in the Mutual Fund Industry.” Particular thanks go to Michael Lemmon for his contributions to this earlier version of the paper.

The Mutual Fund Fee Puzzle

Abstract

Previous work shows large differences in fees for S&P500 index funds. We expand this work to compare fees across all US equity funds using two methods, regression-based pricing models and holdings-based fund matching, to control for fund heterogeneity. We find economically large, robust, persistent and pervasive fee dispersion in the mutual fund industry. Importantly, fee dispersion exists among the largest funds (top TNA quintile) as well as among institutional funds. Most surprisingly, fee dispersion has noticeably increased over the last twenty years, even as the industry has experienced enormous growth in capital invested and the number of funds.

1. Introduction

Two important papers, Elton, Gruber, and Busse (2004) and Hortacsu and Syverson (2004) document substantial price dispersion for essentially identical S&P500 index funds.[1] These results are surprising because in competitive markets, prices for close to identical products should have similar prices. Elton, Gruber, and Busse (2004) conclude that a combination of the inability to arbitrage (i.e., one cannot short sell open-ended mutual funds) and uninformed investors is sufficient to have the law of one price fail in the S&P500 index fund market. Hortacsu and Syverson (2004), in contrast, link the fee dispersion to a combination of nonfinancial fund differentiation and search frictions. Since the publication of these two studies, the mutual fund markets have experienced dramatic growth. Along with this growth has come the continued debate, on the part of academics, practitioners, and regulators concerning the competitiveness of the mutual fund markets and related questions concerning mutual fund fees and their impact on performance.[2,3]

In this paper, we investigate whether the price dispersion effects documented in Elton, Gruber, and Busse (2004) and Hortacsu and Syverson (2004) are still present for S&P500 index funds. More importantly, however, we examine the degree to which these pricing effects are present in *all* US equity funds using two different approaches to control for fund heterogeneity. Specifically, we use an extensive time-series and cross-section of fund data ranging from 1966 to 2014[4]. The key advantage of this rich, empirical setup is that it allows us to document novel and interesting results regarding the determinants of

---

[1] See also Elton, Gruber, and Rentzler (1989) who find that public commodity funds exist that underperform the risk free rate and Christoffersen and Musto (2002) who find a wide dispersion in expenses across similar money market funds. Theories on optimal fund fees are rather scarce (see, for example, Nanda, Narayanan and Warther (2000), Das and Sundaram (2002), Berk and Green (2004), and Pastor and Stambaugh (2010)).

[2] Haslem, Baker and Smith (2006), Gil-Bazo and Ruiz-Verdu (2009), and Barras, Scaillet and Wermers (2010) argue that there is lack of price competition among funds. In contrast, Khorana and Servaes (2009), Wahal and Wang (2011) and Cremers, Ferreira, Matos, and Starks (2015) find evidence that the mutual fund industry behaves like a competitive industry.

[3] For example, the Wall Street Journal (May 3, 2013), in reporting on the fees charged by index funds, writes that “You might think that a plain-vanilla index mutual fund is synonymous with low fees. But some such funds are charging expenses that might make even a high-turnover momentum-fund manager blush.” In the legal arena, the New York Times (August 9, 2016) reports in an article titled “M.I.T., N.Y.U. and Yale Are Sued Over Retirement Plan Fees,” on a class action suit accusing prominent academic institutions of allowing their employees to be charged excessive fees on their retirement savings. The Wall Street Journal (May 18, 2015) reports on a recent case at the US Supreme court related to mutual fund fees charged within a company 401(k) plan. The plaintiffs in the case of Tibble et al. v. Edison International allege that the “…California utility company Edison International breached its duty to plan participants by using high-cost shares of some mutual funds when lower-cost alternatives were available.” The topic of mutual fund fees has also received attention from the Executive branch of the US government. In a New York Times (February 28, 2015) editorial, they write that “A new study by the White House Council of Economic Advisers has found that financial advisers seeking higher fees and commissions drain $17 billion a year from retirement accounts by steering savers into high-cost products and strategies rather than comparable lower-cost ones.”

[4] Elton, Gruber, and Busse (2004) and Hortacsu and Syverson (2004), for example, are limited to a sample period of 6 years between 1995/1996 and 2000/2001 and a sample size of less than 100 funds.

fund fees in the cross-section and the dynamics of fees and fee dispersion over time for a broad set of funds. Thus, in contrast to these earlier studies using a very narrowly defined set of funds, we are able to evaluate the pervasiveness of fee dispersion in a much broader context including the quantification of investor welfare costs associated with investing in high fee funds when similar lower fee funds are available.

To control for the heterogeneity among funds in our sample, we use two approaches. First, we make the fees charged by different funds comparable by following a widely used approach in the economics literature to standardize prices: we examine the residuals from yearly, cross-sectional regressions of total annual expenses (i.e., annual operating expenses, including management fees and 12b-1 fees) on lagged fund characteristics, such as risk and performance characteristics, extent of active management, service levels, fund size and age (we label these residuals "residual expenses" in the paper).[5] The residual approach lets us compare prices across "identical" funds, under the assumption that we have controlled for the relevant fund characteristics. Second, we use holdings data to identify similar funds. This approach is inspired by Wahal and Wang (2011) who identify similar funds for their analysis based on holdings. One important advantage of this approach is that it is completely model-free; i.e., it neither depends on the linear pricing framework nor on specific fund-expense models. For simplicity and unless we observe important qualitative differences, we focus on the results from the regression-based approach in the general discussion.

In our empirical analysis, we first extend the work of Elton, Gruber, and Busse (2004) and Hortacsu and Syverson (2004) and find large levels of fee dispersion across similar US market S&P500 index funds in the most recent 15 years' worth of data after their studies. Elton, Gruber, and Busse (2004), for example, report the value of 34 bps for the interquartile spread in the reported annual expense ratios for their sample.[6] Remarkably, we find that the average interquartile spread in reported expenses has increased to 50 bps for S&P500 index funds in the period after their studies. Further, the 10-90th percentile spread in reported expenses over this more recent period is 116 bps. When we examine residual expenses, we also find large spreads; the average interquartile spread in residual expenses is 32 bps across the S&P500 index sample while the 10-90th percentile spread is 61 bps. Thus, the level of fee dispersion reported for S&P 500 index funds is still large and economically important; it has not decreased after the publication of these influential studies but, in contrast, it has increased. It is also noteworthy that our regression-based framework to control for fund heterogeneity yields a sizeable reduction in fee dispersion even in the case of S&P 500

[5] See Bakos (2001), Brown and Goolsbee (2002), Brynjolfsson and Smith (2000), Lach (2002), Nakamura (1999), Pratt, *et al*. (1979), Scholten and Smith (2002), and Sorensen (2000).
[6] Note that we also find the value of 34 bps if we limit our sample to the 1996 to 2001 period of the Elton, Gruber, and Busse (2004) study.

index funds. This is surprising, as it suggests that S&P 500 index funds might not be as homogeneous as one might think once one controls for fund characteristics beyond past performance and risk exposure.

The important remaining questions, however, are whether the phenomenon of fee dispersion extends beyond index funds and how it has evolved over a more extensive period of time. Thus, we turn our attention to the comprehensive sample of all US equity funds since 1966 using both the residual and holdings based approaches to control for fund heterogeneity. Using the residual approach, we find that the average spread in residual expenses between the 25th and the 75th percentile (between the 10th and 90th percentile) across all funds over the sample is 62 bps (123 bps). These are economically very large numbers.

While we explicitly control for fund size in our regression framework, we also separately replicate our analysis for the funds in the largest TNA quintile to rule out that our results are driven by some obscure and small funds. Note that these large funds represent on average almost 85% of the market value of our fund universe. For these funds we find an interquartile (10-90th percentile) spread in residual expenses of 41 bps (92 bps). While these numbers are smaller than in the case of the full sample, they are still economically very large given the substantial size of the funds included in this sub-sample.

In a next step, we analyze the dynamics of fee dispersion. Interestingly, the dispersion in residual expenses has not decreased over time despite the huge growth in the industry. In fact, we find quite the opposite, namely a pronounced increase in fee dispersion over time. If we split the sample into halves, we find an interquartile (10-90th percentile) spread in residual fees of 41 bps (104 bps) in the first and 64 bps (125 bps) in the second half. Surprisingly, this pattern is even much stronger for the sub-sample of large funds. In this case, we find that spreads have more than doubled – the interquartile (10-90th percentile) spread in residual regressions has increased from 20 bps (46 bps) to 43 bps (96 bps).

Importantly, the results on fee dispersion are very robust, qualitatively but also quantitatively, to a host of alternative specifications of the linear pricing framework, including variations on the fund characteristics used to generate the residual expenses and the estimation of residual expenses at the share class level instead of the fund level. For example, we repeat our basic fund level analysis with three different approaches to aggregate share classes. For the full sample of funds, the interquartile (10-90th percentile) spread experiences a small drop to 54 bps (116 bps).

Finally, we use a completely different, model-free approach to control for fund heterogeneity. Using the holdings-based approach, we find results that are qualitatively very similar to the regression-based approach. For example, we find that the average spread in reported expenses for all funds matched with their closest-neighboring-fund in terms of holdings is 54 bps (103 bps) for the interquartile spread (10th to 90th percentile) spread.

One might assume that an annual average fee dispersion of 50 to 100 basis points is small and not economically important. To address this concern, we examine the implications of our findings for investors. Based on residual expenses, an investor purchasing the lowest expense funds (i.e., the bottom decile) would have earned compounded abnormal returns 71% higher than an investor purchasing the most expensive (i.e., the top decile) funds. We also measure the dollar amount associated with positive residual fee funds. Those fees can be interpreted as capturing excessive fees, as reported fees exceed expected fees (according to our pricing model) by this amount. For example, in the case of the sub-sample of large funds and the most recent year in the sample, 2014, we find that the cumulative dollar amount associated with positive residual fees is USD 7 billion or 22% of total revenues earned by these funds in that year. Obviously, these numbers have to be interpreted cautiously, as they depend on several assumptions. However, we still find them interesting, as they clearly illustrate, even if one disagrees with the specific values, that investors bear significant costs from investing in high expense mutual funds when similar lower fee funds are available.

In a final step, we exploit the richness of our data sample and investigate whether (a) the degree of competition or (b) distribution channels affect fund fees and matter for fee dispersion. Specifically, we follow Wahal and Wang (2011) and create a fund-level competition measure based on holdings information. We also construct empirical proxies for two specific frictions that could potentially inhibit competition and, as a consequence, boost dispersion. Specifically, we single out funds that are potentially attractive to investors because of salient advertising but are difficult, because of back-end loads, to exit (i.e., funds with high marketing expenses and high back-end loads – we deem these funds "easy-in hard-out" funds). Furthermore, we calculate a flow-performance measure to capture investor attention, as Christoffersen and Musto (2002) and Gil-Bazo and Ruiz-Verdu (2008) argue that funds with performance-insensitive investors tend to have higher fees.

In terms of expected fees, the empirical evidence for competition is mixed. For the full sample, in our pricing regressions, we find some evidence counter to a competition story – fees are positively, but only marginally significantly, related to competition. In contrast, for the largest-TNA-funds, we find a statistically significant negative relation between fees and competition – thus more competition is associated with lower expected fees. While the easy-in-hard-out fund characteristic is strongly and robustly associated with higher fees, we do not find any evidence that funds with lower flow-performance sensitivity charge higher fees. Importantly, however, controlling for any of these additional variables does not lead to a substantial reduction in fee dispersion.

Finally, we investigate the role of distribution channels using two approaches. First, we split our sample into institutional and retail funds. We find higher reported expenses for retails funds, but surprisingly, we also find large levels of residual fee dispersion for institutional funds. Second, we follow

Bergstresser, Chalmers and Tufano (2009) and distinguish between directly-sold and broker-sold retail funds using 12b-1 fees. We find higher reported fees for broker-sold funds relative to directly sold funds, but the amount of residual fee dispersion remains high in both groups. Thus, while distribution channels help explain levels of fees, they do not help in explaining fee dispersion across similar funds.

We see the following broader contributions of this paper. First, we show that the previously documented spread in fees of S&P 500 index funds is still prevalent nearly 15 years after it has been documented in influential and widely disseminated papers. Second, we provide comprehensive and robust evidence that the phenomenon of fee dispersion extends to the *entire* US equity fund industry including the largest funds. Third, the dynamics of fee dispersion are stunning, as fee dispersion has not decreased over time but, instead, has increased substantially despite the enormous growth in terms of assets under management and funds. Fourth, the levels of fee dispersion documented in this paper are economically large and correspond to substantial welfare losses for investors who pick the wrong funds. Clearly the stakes are high. Fifth, we shed some light on how fees are set in the industry. Importantly, however, we are still – after exploiting two different methods to control for fund heterogeneity and a multitude of proxies for fund characteristics – unable to make sense of the levels of fee dispersion that prevail in the fund industry leaving us with a puzzle about the drivers of mutual fund fees.

The remainder of the paper is organized as follows. In Section 2 we describe the data used in our analysis and describe the characteristics of high and low expense funds. In Section 3 we present results based on the regression-based fund pricing approach, while Section 4 summarizes results from a holding-based matching approach. In Section 5, we look in more detail into the importance of competition and distribution channels on fees. Section 6 concludes.

2. Data

2.1 Sample Construction

The sample selection follows Pastor and Stambaugh (2002). Accordingly, we select only domestic equity funds and exclude all funds not investing primarily in equities such as money market or bond funds. In addition, we exclude international funds, global funds, balanced funds, flexible funds, and funds of funds. The ICDI classification codes that were used by Pastor and Stambaugh (2002) are, however, no longer available. Thus, we follow Bessler et al. (2008) who use a combination of Lipper codes, Wiesenberger codes and Strategic Insight codes to identify domestic equity funds. Table A in the Data Appendix lists the specific codes that we use to identify the funds in our sample.

In short, the above screens result in our sample focusing on active and passive US domestic equity funds. Our sample includes approximately 38% of all funds covered in the CRSP Mutual Fund Database (our sample consists of a total of 20,926 funds while the CRSP Mutual Fund Database universe has approximately 55,109 funds). As measured by total net assets, our sample covers approximately 42.5% of the cumulative net assets represented in the database. The sample period spans 1966 to 2014 and the data frequency is yearly, as we focus on fund expenses.

2.2 Descriptive Statistics

Table 1 Panel A reports summary statistics of our fund sample. Details of the variable construction can be found in Table B in the Data Appendix. Throughout the paper we distinguish between a pre-1999 (up to and including 1998) and a post-1999 (including 1999) sample because several important variables such as fund family information (i.e., information on management companies) and flags for institutional funds only became available in the CRSP Mutual Fund Database in 1999.

The descriptive statistics show the dramatic increase in mutual funds over the past 30 years. In the pre-1999 sample the mean number of funds per year is 1051, while it increases to 7931 in the post-1999 sample. Note that the mean fund size (*TNA*) also increases from 361 Million USD pre-1999 to 530 Million USD post-1999. Thus, the mutual fund industry has experienced a considerable increase in assets under management.

Intuitively, given more funds and thus presumably increased competition, we would have expected to find that the rapid expansion of the mutual fund industry was also accompanied by a decrease in average expense ratios – but this is not the case.[7] Average annual expense ratios (*expense ratio*) increased from 121 basis points (bps) to 138 bps. It is also interesting to observe that yearly changes of expense ratios (*Δ(expense ratio)*) are on average close to zero. This is mostly driven by the fact that, on average, a similar fraction of funds increases and decreases their fees, namely 32% and 33% of all funds in a given year, respectively. Thus, if we remove the signs of the fee changes and calculate the time-series average of cross-sectional mean *absolute* fee changes, we find that it corresponds to 22 basis points; a relatively large number in economic terms.

The average performance of our sample funds, as measured by annual four-factor alphas (Carhart (1997)), is negative (-1.4% in pre-1999 and -1.2% in post-1999), consistent with Carhart (1997) and others who show that funds do not earn positive abnormal returns net of expenses. The average fund, over both time periods, has a market beta (*beta_mkt*) that is slightly less than 1, a small, negative exposure to HML (*beta_hml*), and small positive exposures to SMB (*beta_smb*) and UMD (*beta_umd*). After 1999, funds load more on the market, and less on SMB, HML, and UMD, consistent with an aggregate strategy shift to

[7] The averages reported in Table 1 are equal-weighted. If we value weight the expenses, we find a slight decrease from pre to post-1999. In this case, the corresponding values are 87 bps for pre-1999 expenses and 78 for post-1999 expenses.

market indexing. The four-factor model works well on average in explaining fund returns, yielding R^2 of 78% and 87%, in the pre and post-1999 periods, respectively.

Panel B (pre-1999 sample) and Panel C (post-1999 sample) of Table 1 report summary statistics by expense ratio deciles. Each year we split all funds into deciles by their expense ratios and then report contemporaneous means and standard deviations of fund characteristics.

Average expense ratios (*expense ratio*) of decile 10 exceed those of decile 1 by roughly 230 bps, in both the pre-1999 and post-1999 periods. In the pre-1999 sample, average expense ratio changes are most negative (-14 bps) in decile 1 and most positive (42 bps) in decile 10. These mean changes become smaller in the post-1999 sample: funds in the bottom expense ratio decile decrease their expenses on average by 1 bps in the same year, while funds in the top decile increase their expenses on average by 6 bps in the same year.

All of the fund performance variables decrease by expense ratio deciles. For example, the spread in yearly four-factor Carhart alphas between expense ratio decile 1 and decile 10 is 2.0% pre-1999 and 1.8% post-1999, which is comparable to the spread in expense ratios, especially post-1999. Thus, these simple descriptive statistics suggest that funds with higher expense ratios on average underperform their cheaper competitors by approximately their expense ratios.

We also find that average fund size (*TNA*) in decile 1 is much larger than average size in decile 10, suggesting that economies of scale play a role for expense ratios. The average size in decile 1 is approximately 1.5 (1.9) Billion USD larger in the pre-1999 (post-1999) periods than the average size fund in expense ratio decile 10. We also find a greater concentration of fund families (*Family dummies*) in the lower expense deciles, although there are a non-trivial number of funds that belong to large fund families that reside in the higher expense deciles. For example, in Panel C, 57% of the funds in expense decile one are funds that belong to a fund family with more than 100 funds and 46% in expense decile ten are funds that belong to a fund family with more than 100 funds.[8] Moreover, we also find a greater concentration of institutional funds (*Institutional dummy*) and ETFs (*ETF dummy*) in the lower expense deciles.

Finally, Panel D of Table 1 shows pooled correlations between fund characteristics. These correlations are consistent with our previous interpretations of patterns between expense ratio deciles and other fund characteristics. In general, none of these correlations seem to be high enough to cause worries about multi-collinearity problems in the subsequent multivariate analysis.

[8] In later cross-sectional tests we find that large families charge greater expenses.

Of course, the most important limitation of this univariate analysis from Table 1 is that it ignores that expense ratios may reflect different fund strategies and characteristics. This is something that we will explore in more detail in later sections of the paper. These simple summary statistics, however, already suggest that, to some extent, expense ratios can be explained by economic determinants. For example, funds' risk characteristics seem to be correlated with expense ratios: more expensive funds tend to exhibit higher absolute loadings on standard risk factors (i.e., on MKT, SMB, and UMD). Similarly, the average R^2 of the four-factor model decreases as we move from decile 1 to decile 10, suggesting that the managers of the higher expense funds may be following "unique" strategies, likely in an attempt to outperform. However, these managers also trade much more (i.e., the *turnover* is much higher for the high expense funds relative to the low expense funds), which may contribute to their low return performance. Overall, these patterns between risk characteristics and expense ratios are intuitive and suggest that expensive funds do follow, at least to some extent, more active strategies, load more aggressively on individual risk factors, and implement strategies that go beyond the standard risk factors.

3. The Pricing of Mutual Funds (Regression-Based Approach)

Our goal is to compare prices (total expense ratios including management expenses and 12b-1 fees) across funds. Of course, not all funds are the same and differences in fund characteristics might justify price differences. In the regression-based approach, we follow Lach (2002) and Sorensen (2000) to control for fund heterogeneity using cross-sectional regressions. As controls we use a comprehensive set of fund characteristics that have been shown to be important in determining fund expenses (e.g., see Gil-Bazo and Ruiz-Verdu (2009) and Wahal and Wang (2011)). An alternative approach that is discussed in detail in Section 4 is to match funds using holdings data.

3.1 Residual Expense Estimation and the Pricing of Individual Fund Characteristics

We regress fund expenses on lagged fund characteristics including performance and risk characteristics. As our set of explanatory variables changes over time (e.g., fund family information is only available after 1998), we estimate a cross-sectional regression each year. Another advantage of this specification is that it allows for changing relationships (i.e., time-varying coefficients) between fund characteristics and expenses. The residuals of these regressions can be interpreted as deviations of fund expenses from expected expenses given the set of characteristics used in the regression. Thus, using the residuals, we can compare prices across "identical" funds, under the assumption that we have controlled for the correct fund characteristics.[9]

[9] We are careful to include fund characteristics that we believe should matter to the average investor. Many of these characteristics are related to fund performance – items that should be the first order determinants of fund expenses.

In Panel A of Table 2 we present the details of the yearly cross-sectional regressions used to estimate the residuals. The reported coefficients are time series averages of cross-sectional regression coefficients obtained from the annual cross-sectional regressions. We estimate these models separately for three samples: (i) S&P 500 index funds[10], (ii) the full sample of all US-equity mutual funds, and (iii) for the largest quintile of annually-ranked TNA funds. We also standardize the independent variables to have a mean of zero and a standard deviation of one (see Panel B of Table 2) using each variable's yearly cross-sectional mean and SD. The standardized coefficients, thus, allow us to discuss a fund fee price estimate for a one standard deviation change in each independent variable, and also allow us to rank the fund characteristics in terms of economic importance.

The last column of Table 2, Panel A, summarizes coefficient estimates for S&P 500 index funds. Elton, Gruber, and Busse (2004) and Hortacsu and Syverson (2004) claim that these funds are very homogeneous and, thus, can directly be compared in terms of fees. Interestingly, in the period of 1999 to 2014 we find that our set of fund characteristics is still able to explain, on average, 59% of the variation in expense ratios across these funds. Thus, it seems that according to these fund characteristics they are not as homogeneous as expected. The most important characteristic in terms of fund pricing is the institutional dummy: institutional S&P 500 index funds charge, on average, 28 basis points less than retail funds. Fund size also plays an important role, as a one standard deviation increase in the fund size yields a reduction in fees of 23 basis points indicating large economies of scale in this fund segment. Interestingly, we also observe that index funds that are part of large management companies charge, on average, 12 basis points more. Overall, we conclude that the subset of S&P 500 index funds is certainly homogenous in terms of strategy but not in terms of fee-relevant fund characteristics.

For the full sample of US-equity funds in the pre-1999 period, the model explains approximately 25% of the variation in expenses; in the post-1999 period, the model explains 27%. The signs of the coefficients are mostly consistent with the literature: e.g., we observe that better performing funds (*Annual return*), less volatile funds (*sdmret*), larger funds (*TNA*), younger funds (*fund age*), lower turnover funds (*turnover*), institutional funds (*Institutional dummy*), ETFs (*ETF dummy*), and funds with higher R^2 from the Carhart

However, we also include service and other non-performance related characteristics in the expense models. We also perform extensive robustness tests (see Section 3.4 for details) to show that our results are not sensitive to the specific choice of variables.

[10] We identify S&P500 index funds by the following steps: (1) we first extract all index funds using CRSP's index fund flag; (2) for each index fund identified in step 1, we regress the full-sample monthly returns of the fund on the S&P 500 index monthly returns, and (3) using the beta of each fund from the regression we retain the funds that have a beta within the range [0.995, 1.005] and a R-squared within the range [0.995 and 1.005]. This results into total 185 S&P 500 index funds from 1999 to 2014.

four-factor model have lower expenses. Across the pre- and post-1999 periods, we essentially see the same relationships.

In terms of economic importance or "pricing" of individual fund characteristics, we observe substantial variation across variables. For example, consider the most important fund characteristics according to the full sample results in the post-1999 period. The coefficient on the institutional dummy is -51.65, suggesting that fund investors pay an extra 52 bps for non-institutional funds on average. Also, investors pay 44 bps to be in non-ETF funds and 35 bps to purchase an extra unit of fund standard deviation of return, which can probably be viewed as the price of buying a more active and less diversified fund. A one standard deviation smaller fund charges, on average, 25 bps less while funds that are part of the largest management companies (with more than 100 funds) charge an additional 14 bps, on average.

As far as the price that investors pay for gross-of-expenses fund performance are concerned, the results seem counter-intuitive because of the negative coefficient on lagged returns (*Annual return*) in the fee regressions. Fund investors pay 8 bps for a unit standard deviation of lower annual gross returns. The negative sign, however, is consistent with Christoffersen and Musto (2002) and Gil-Bazo and Ruiz-Verdu (2008) who argue that as fund returns go down, performance sensitive investors exit the fund, leaving a majority of performance insensitive investors, for whom fund management then raises the fees. Even more importantly, past fund performance does not seem to be of first order importance for fund fees. At least, in terms of the absolute magnitude of its price it consistently does not rank among the top priced variables.

Finally, we look at the annually ranked largest (top TNA quintile) funds separately. Even though we control for fund size in our full sample estimates, our results might still be driven by tiny funds if, for example, the relationship between size and fees is nonlinear. Furthermore, from an economic point of view, these are hugely important funds that cover, at the end of our sample period, more than 90% of capital invested in our sample of funds.

Our regression-based approach explains a bit more of fee variation in the sample of large funds – 31% pre-1999 and 34% post-1999 – than in the full sample. In terms of coefficient estimates and relative importance of fund characteristics for fund pricing, large funds largely behave similarly to all funds. It is, however, interesting that the discount for institutional funds is cut down to 37 bps in the case of large funds. Similarly, we do not find any premium associated with the standard deviation of returns of large funds. In contrast, a one standard deviation increase in turnover is associated with a rather large fee premium of 11 bps. We also observe that large funds within the largest management companies charge a premium of 19 basis points. Last but not least, expense ratios of large funds seem to be rather independent from past performance.

Overall, we view these results to be largely consistent with economic theory and intuition regarding the pricing of mutual funds. They help us build confidence that the regressions control for important aspects of fund heterogeneity. In a next step, we will focus on the residuals of these regressions, what we deem the "residual expenses," and analyze them in the time-series and cross-section.

3.2 Detailed Analysis of Fee Dispersion

Our main point of interest, the distribution of residual expenses, is summarized in Table 3 and visualized over time in Figure 1. In the figure, each year we plot the reported (left column) and the residual (right column) expense spread between the 25th and 75th, 10th and 90th, and 1st and 99th percentile points of the distribution (note that the mean residual is zero by construction).

For the sample of S&P 500 index funds, we find that the average interquartile residual fee spread is 32 bps and the average 10-90th percentile residual fee spread is 61 bps. For reported expenses, the average interquartile spread is 50 bps and the 10-90th percentile spread amounts to 116 bps. Considering the homogeneity of these funds in terms of strategy and the large number of fund characteristics controlled for in the regression, these values are stunningly large. Even more interestingly, Figure 1 shows that spreads in reported and residual fees did not at all shrink over time – they are nearly as big at the end of the sample as they were towards the beginning. These results are consistent with the numbers reported in Elton, Gruber and Busse (2004). Importantly, however, we find that the situation has not changed at all over the last 15 or so years.

If we extend our analysis to the full sample of US equity funds, we find an average interquartile residual fee spread of 62 bps and an average 10-90th percentile spread of 123 bps. Figure 1 illustrates that these large numbers are not driven by the early years of our sample. Rather the opposite is the case, as we observe a gradual increase in these spreads up until the early 90ties and constant levels afterwards. If we split the full sample of funds into the first and second halves using 1990 as the midpoint, we find that the residual fees are higher in the second half. Specifically, we find an interquartile (10-90th percentile) spread in residual fees of 41 bps (104 bps) in the first and 64 bps (125 bps) in the second half (the first and second half numbers are not reported in the tables). This pattern seems surprising and rather counter-intuitive given the immense growth – in terms of dollars invested and funds competing for capital – experienced by the fund industry, and given the increase in transparency due to improved regulation and easier information dissemination. To illustrate the dramatic growth of the fund industry, the graphs in the top row of Figure 1 also show the total TNA covered by the sample over time.

Even more surprisingly, we find the same pattern in a more pronounced way for the sub-sample of largest funds. On average, as reported in Table 3, for these funds we document an interquartile residual fee spread of 41 bps and a 10-90th percentile spread of 92 bps. These economically very large numbers are to a large extent driven by the last 25 years of the sample period. If we split the sample into halves, the interquartile (10-90th percentile) spread in residual fees is 20 bps (46 bps) for the first half and 43 bps (96 bps) for the second half. Thus, the residual fees for large TNA funds have more than doubled from the first to the second half of the sample. Examples of individual year residual fees illustrate this pattern of the growth in fees over time. For example, in 1970 the corresponding spreads are very tight at 15 (interquartile spread) and 24 bps (10-90th percentile spread). During the 80ties and 90ties, however, we observe a pronounced increase in spreads with the peak in 2003 at levels of 55 bps (interquartile spread) and 110 bps (10-90th percentile spread). Surprisingly, spreads then stay at these high levels; for example, in 2014, the last year in our sample, the interquartile spread is 44 bps and the 90-10th spread is 81 bps. Given that in this sample we only consider the 20% largest funds in a given year, we find these results to be economically relevant and very surprising. Importantly, they also clearly show that the phenomenon of fee dispersion is not at all driven by small funds. Finally, it is noteworthy to point out that this increase in spreads for the largest funds corresponds to an increase in the concentration in the fund management industry. The graphs in the middle row of Figure 1 show how the share of total TNA associated with the largest funds evolves over time (the solid line in the graphs; the y-axis on the right hand side shows the corresponding scale). While largest funds represent, on average, 83.6% of the market value of our sample, we see a pronounced increase starting from less than 75% in the mid 80ties to more than 90% around 2008.

3.3 Economic Magnitude of Fee Dispersion

An important question that still remains to be answered is whether the documented fee spreads are economically large. In this section we present two types of analysis to explicitly quantify the economic importance.

First, we implement a simple ex-ante trading strategy that trades funds based on the residual expense distribution, illustrating the negative wealth effects of investing in similar, but higher expense funds.[11] We assume no taxes. For comparison purposes, we also report a similar strategy using reported expenses. We compute the returns to a trading strategy that buys funds in the bottom decile and sells funds in the top decile of expenses. We rebalance these portfolios every year and compute the cumulative four-factor alphas over the 49 year sample period to equally-weighted portfolios.[12]

[11] Of course, this is not an implementable strategy since one cannot short sell open-ended mutual funds.
[12] We estimate the cumulative four-factor model alpha as follows. Using monthly returns from the annually rebalanced low-fee minus high-fee portfolio, we estimate the monthly 4 factor alpha each year and multiply it by 12 to obtain an estimate of the annual alpha. We then compound the annual alphas over time and report the cumulative alphas.

For index funds (Figure 2, bottom row), we find that the cumulative performance of our trading strategy increases monotonically. Over the sample period of 14 years, an investor purchasing the lowest residual expense index funds would have earned compounded abnormal returns 13.4% higher than an investor purchasing the most expensive index funds.

For the full sample of funds (top row), we interestingly find that during the first few years, until the late 70ties, our hypothetical trading strategy would have not earned positive cumulative abnormal returns suggesting that managers were able to "earn their keep." Starting from the early 80ties, however, we find a nearly monotonically increasing pattern of outperformance resulting in cumulative abnormal returns over the full sample of slightly more than 70% higher for an investor purchasing lowest residual fee funds than for an investor purchasing the most expensive funds.

We perform a similar trading strategy for the sub-sample of largest funds (middle row of Figure 2). The results are very similar: large TNA low-residual fee funds outperform large TNA high-residual fee funds by a cumulative four factor abnormal return of 50% over the full sample. Consistent with the earlier discussion of time-series dynamics of residual fee spreads, this outperformance is earned almost entirely during the last 25 years of the sample period. During the first twenty years, in contrast, the cumulative performance of our trading strategy was variable, with low residual fee funds strongly outperforming high fee funds from 1966 to the middle of the 1970s, followed by a 15 year or so period of high fee funds outperforming low fee funds.

In the left column of Figure 2, we also report results for a trading strategy that conditions on reported, rather than residual, expenses. As one would expect, the patterns are similar across columns in Figure 2 but magnitudes are larger for the reported fees. We also report compounded differences in reported and residual expenses (the dashed lines) to benchmark our performance results. For all funds, over the full sample period, these compounded differences amount to 193% for the reported expenses and 160% for the residual expenses. Thus, while the difference in abnormal returns between high expense and low expense funds is less than the cumulative difference in expenses, investors bear significant costs from investing in high expense mutual funds that are not recouped through higher performance of these funds.[13]

As a second strategy to quantify the economic magnitudes, we calculate the industry's cumulative positive residual revenue. To calculate the cumulative positive residual revenue in a given year, we look at all funds with positive residual fees, then multiply those positive residual fees with the funds' TNA, CPI-adjusted to 2014 dollars, and sum everything up across funds. If we interpret positive residual fees as

[13] In contrast to our results, Ramadorai and Streatfield (2011) find little difference in performance across high and low management fees (i.e., the non-performance fee part of hedge fund expenses) for hedge funds. They conclude that high management fees are "money for nothing" in the hedge fund industry.

excessive or inefficient fees, then the cumulative positive residual revenue can be interpreted as an industry-level measure of excessive fee-taking in dollar terms. Of course, the exact numbers from this analysis should be interpreted with caution since they are conditional on our choice of fund characteristics included in our base fee model of Table 2. Nevertheless, we think that this analysis is informative, as it explicitly links fee residuals to dollars invested.

In the case of the sample of largest funds, for example, cumulative positive residual revenue amounts to USD 2.10 billion per year, on average over the sample period. This is a large number in economic terms and corresponds to slightly less than 18% of total fee revenue earned by the sample of largest funds. For comparison, in the most recent year in the sample, 2014, cumulative positive residual revenue equals nearly USD 7 billion or 22% of total revenues earned.[14]

To summarize, the levels of fee dispersion that we document for the full sample of US equity funds, for the sub-sample of largest funds, and for the sub-sample of index funds are economically important. Consistent with our analysis of fee dispersion measures across time but contrary to our ex-ante expectations, we find that the issue of fee dispersion has gained substantially in importance during the last 25 years or so. Put differently, fee dispersion was much less of an issue, in statistical and economic terms, in the 70ties and early 80ties than it has been in more recent times.

3.4 Robustness Tests

Obviously, controlling for heterogeneity among the funds in our sample is crucial for our analysis. To make sure that the results are not driven by the specific set of fund characteristics employed in the pricing regressions, we run a comprehensive set of robustness tests in which we either add more fund characteristics (e.g., a performance persistence measure or style fixed effects) to our main model or replace existing fund characteristics by alternative measures (e.g., using risk-adjusted performance measures instead of simple gross-of-fee returns).[15] In each of these tests, we rerun the cross-sectional regressions for the updated set of fund characteristics and then analyze the residual fee dispersion.

[14] The cumulative positive residual revenues for the full sample of funds are only slightly larger than these amounts for the largest quintile of annually ranked TNA funds because of the fact that the largest TNA funds represent over 90% of the dollar value of the entire fund sample.

[15] Specifically, we include the following fund characteristics in the robustness tests (details on the construction of these variables can be found in the data appendix): (i) measures of abnormal gross returns such as four-factor alphas, the t-statistics of the four-factor alphas and Carhart alphas using alternatively rolling-window, expanding-window and full sample estimates, (ii) all previously employed performance measures but net-of-fees, (iii) a measure of performance persistence, (iv) a filtering approach that sets all alphas equal to zero whose t-statistics are below three, and (v) style fixed effects using a combination of Lipper codes, Wiesenberger codes and Strategic Insight codes. In total, we evaluate 15 different specifications of the pricing regressions.

All of these robustness tests show that our results are insensitive, qualitatively and quantitatively, to the specific set of fund characteristics included in the pricing regressions. The levels of the 10-90th percentile spreads discussed in the previous section vary by less than 3 bps across all these robustness tests. To not overwhelm the reader with pages of repetitive tables, we abstain from reporting the detailed results of these robustness tests – they are available from the authors upon request.

Another set of robustness tests deals with the issue of funds that have multiple share classes. In our main results we treat each share class as an individual fund. If share classes proxy for different distribution channels (Bergstresser, Chalmers and Tufano (2009)) or different investor clienteles, then different share classes of the same fund could (and often times do) have different expense ratios. Thus, we evaluate whether our levels of expense dispersion are driven by different share classes. Note that we report and discuss additional results regarding the role of distribution channels in Section 5.2.

Share classes are not automatically identified within the CRSP Mutual Fund Database. We use the MFLINKS tables that are provided by WRDS for this purpose. The original idea of these tables is to link the funds in the CRSP Mutual Fund Database with the ones covered in the Thomson Reuters Mutual Fund Ownership Database. Our analysis in this section begins in March of 1980 since that is when the share class data starts. After identifying the individual share classes for a given fund, we aggregate the share classes (i.e., the expenses, returns, and other characteristics) into a common fund using equal and value weighting (using the total net asset values to determine weights). To avoid potential expense dispersion from share class aggregation, we also perform tests using the largest share class only for each fund. We re-estimate our main tests on these new, aggregated samples. Before discussing detailed expense dispersion results, it is interesting to look at some descriptive statistics regarding the use of different share classes in the mutual fund industry. First, we find that before 1995 it was very uncommon to have multiple share classes. Second, after aggregating multiple share classes into funds, we have on average more funds (798) with than funds without multiple share classes (542) each year. Third, the average size of funds without multiple share classes (approximately USD 548 million) is smaller than the size for aggregated funds with multiple share classes (approximately USD 759 million using value-weighted aggregation).

Table 4 summarizes expense dispersion results for the full sample of funds (Panel A) and the top size quintile (Panel B) of the aggregated funds.[16] Row 1 of each panel reports no share class aggregation results as a base case (using the basic expense models of Table 2 on the post 1980 sample) and rows 2-4 report results for the three aggregation methods. For the full sample, we find that the interquartile (10-90th

[16] Note that we do not look separately at the sample of S&P 500 index funds in this robustness test due to the small size of that sample. The residual fees used to calculate the spreads reported in Table 4 are derived from re-estimating the basecase regression model, as described in Table 2, separately for each fund-level sample. Coefficients in these regressions are qualitatively similar to those reported in Table 2 and, thus, not reported for reasons of brevity.

percentile) spread in residual fees drops from 63 to 58 (125 bps to 121 bps) when we use value-weighted aggregation and to 54 (116) bps if we focus only on largest share classes. The reduction in residual fee spreads is somewhat more pronounced in the case of largest funds in which case the 10-90th percentile spread drops, for example, from 101 bps to 76 bps when we only consider the largest share class.

Overall, we conclude that our results on the pervasiveness and magnitude of fee dispersion in the mutual fund industry are robust to the issue of multiple share classes. However, as one would have expected, we do find that the tendency of funds to issue several share classes is related to the phenomenon of fee dispersion, in particular so for the largest funds.

Finally, Figure 3 compares the time-series dynamics of the residual expense distribution for our base-case (share class-level, 1st column) and the fund-level analysis (2nd column). In general, the graphs look very similar across columns, documenting a minor impact of share classes on the time-series dynamics of residual expense dispersion. Recall that one of our key results is that expense dispersion does not decline over time; quite in contrast, it actually increases. The graphs clearly show that share class aggregation does not have a noticeable impact on this result.

4. The Pricing of Mutual Funds using a Holdings-Based Approach

So far, we have documented that seemingly similar funds charge very different fees using multivariate regressions to control for fund heterogeneity. In this section, we explore a different approach for identifying similar funds that matches funds using their holdings following Wahal and Wang (2011). One important advantage of this approach is that it is completely model-free; i.e., it does neither depend on the linear pricing framework nor on specific fund-expense models.

4.1 Methodology

For each fund in our sample we obtain holdings information from the Thomson Financial CDA/Spectrum holdings database. This holdings database is linked with the CRSP mutual fund files using the MFLINKS file provided by Wharton Research Data Services. The sample starts in March of 1980 when the holdings information becomes available. To match funds in terms of holdings we develop a pair-wise measure of fund overlap. We use a simple and intuitive measure, namely the sum, across all holdings, of absolute differences in weights for a given pair of funds. We deem this measure "uniqueness." The measure is bounded between zero (perfect overlap) and two (no overlap).[17] It is symmetric in the sense that the ordering of the funds does not matter.[18]

We calculate this measure yearly for all fund pairs (at the fund level, not at the share class level; to aggregate, we use the share class with largest TNA for a given fund). In total, the uniqueness measure is estimated for approximately 2.9 million pairs per year. For each fund, its matched fund is defined as the fund with the lowest pair-wise uniqueness measure (i.e., the largest overlap in terms of holdings) in a given year. We refer to this sample of matched fund pairs as the "full pairs sample." We perform all our analysis for the full pairs sample and for pair sub-samples based on quintiles of the uniqueness distribution of matched fund pairs; i.e., based on the similarity in terms of holdings of the matched pairs. Thus, fund pairs in quintile one (five) are "most similar" ("least similar") fund pairs. Note that "least similar" fund pairs are still relatively similar compared to the average of randomly drawn fund pairs. Finally, we also define "very similar funds" as the bottom decile of the uniqueness measure for the full pairs sample.

In Panel A of Table 5 we provide summary statistics on pair characteristics to provide a sense of how well the holdings-based algorithm performs in identifying similar funds. For each sample, we report the mean and interquartile range (IQR) for the uniqueness measure and differences in average yearly returns (*Annual return*), four-factor model adjusted R-squares (R^2), and beta loadings on the four factors. The average uniqueness value for the full pairs sample is 1.05 with an IQR of 0.58. As we move from quintile five (i.e., "least similar funds) of uniqueness to the bottom decile (i.e., "very similar funds") the mean of the uniqueness sorting variable decreases from 1.49 to 0.15. The differences in R-squares, returns, and betas across fund pairs suggest that the uniqueness measure does a decent job identifying similar funds; all three difference metrics decrease as we move from less to more similar fund pairs.

4.2 Fee dispersion across funds with similar holdings

Next, we examine if funds with similar holdings charge similar expenses.[19] In Panel B of Table 5, we report the absolute difference in reported expense ratios and residual expenses for matched pairs. The residual expenses are from our base-case expense regression models as described in Table 2. The expense differences are large: for the full pairs sample, the average reported expense difference is 49 bps, 54 bps

---

[17] For example, consider two funds with holdings in only two stocks, A and B. If fund 1 holds 100% in A and 0% in B, and fund 2 holds 0% in A and 100% in B, then the uniqueness measure (the sum, across all holdings, of absolute differences in weights) is the absolute value of (1-0) plus (0-1) which is 2, resulting in the funds having no overlap. In contrast, if fund A and B both hold 100% in A and 0% in B, then the uniqueness measure is 0 (i.e., (1-1)+(0-0) = 0) signifying the same holdings.

[18] In contrast, the overlap measures used in Wahal and Wang (2011) are not symmetric: i.e., in their framework it matters which fund is the incumbent fund and which fund is the newly entering fund.

[19] We do not perform the holdings based analysis on the S&P 500 index fund subsample, since as one would expect, the holdings-based approach to control for fund heterogeneity is not effective in creating a spread in the uniqueness measure for S&P 500 index funds since these funds are all very similar to each other in terms of their holdings.

for the interquartile range, 103 bps for the 10th to 90th percentile spread, and 219 bps for the 1st to 99th percentile spread. Expense spreads decrease monotonically from less similar funds to very similar funds, but are still economically large even for the very similar funds. For example, at the 10th and 90th percentile points, the quintile one uniqueness fund pairs have a 101 bps spread, and the very similar funds have a spread of 94 bps in reported expenses.

Thus, matching on holdings gives us qualitatively similar expense spreads as we get from the regression-based residual expense spreads of Table 3. In fact, when we examine the model-based residual spreads for these matched pairs (as reported in the right-hand side of Panel B of Table 5), we see that the spreads decrease to some extent relative to the reported expenses (e.g., for quintile one pairs, the 10-90th percentile spreads drop from 101 bps to 83 bps), consistent with the idea that controlling for fund characteristics has explanatory power on top of holdings, but including characteristics along with holdings still leaves a large unexplained spread in expenses.

In Figure 4 we plot the time-series of the annual distributions of reported and residual expense differences for the full pairs sample, most similar funds (i.e., the quintile one sample of uniqueness), and least similar funds (i.e., the quintile five sample of uniqueness). In addition, the plots also include the yearly average uniqueness value of the pairs included in each figure (solid line). Similar to the time series plots of the residual spreads in Figure 1, there is a lot of time series variation in these plots but only a slight drop in average expense differences in more recent years. In fact, for the most similar funds, we see evidence that despite becoming much more similar in terms of holdings (i.e., the average uniqueness represented by the solid line is decreasing), there is no commensurate drop in expense differences.

5. Drivers of Fee Dispersion

So far, we have documented and, in detail, analyzed the large and economically important levels of fee dispersion that prevail among US equity funds. In the remaining parts of the paper we dig deeper into several mechanisms, such as competition and varying distribution channels, and investigate whether they help us understand fee dispersion. Given that several of the proxies considered in this section depend on data that is sparse or missing for some funds, we exclude the sub-sample of index funds from the analysis and also focus on the post-1999 time period.

5.1 Price Competition

One potential interpretation of large levels of fee dispersion in the mutual fund industry is lack of price competition among funds (see among others Haslem, Baker and Smith (2006), Gil-Bazo and Ruiz-Verdu (2009), and Barras, Scaillet and Wermers (2010) to support this view). This interpretation, however, is at odds with other papers that argue that competition works well among funds. Most prominently, Wahal and

Wang (2011) conclude that the mutual fund industry behaves like a competitive industry, as incumbent funds decrease their expenses when new funds with similar holdings enter the industry. To investigate these conflicting conclusions further, we extend their idea to all funds and construct a measure of competition per fund per year, aggregated from each fund's holdings overlap with all other funds available in a given year.[20]

More specifically, our competition measure is based on the pair-wise uniqueness measure introduced in section 4. To come up with a fund-level uniqueness measure (*Fund Average Uniqueness*) we calculate the simple average of a fund's pair-wise uniqueness measures with all other funds.[21] This measure is constructed so that as it increases (i.e., average holdings with other funds become less similar), competition is assumed to decrease. A fund whose average uniqueness is close to two (the measure is bounded between zero and two), has completely unique holdings and, thus, presumably faces little competition. In contrast, a fund with a low average uniqueness measure has holdings that are similar to the holdings of many other funds and, thus, it is most likely exposed to substantial competition.

Funds, however, also have mechanisms at their discretion to reduce price competition. For example, they can establish provisions that potentially reduce investor exit from or switching across funds (e.g., rear loads or other switching costs such as prohibitions on frequent trading or high transaction fees). To capture the existence of such provisions we create a dummy variable that identifies funds that are potentially attractive to investors because of salient advertising (i.e., funds with high marketing expenses) but difficult to exit (i.e., have high back-end loads). We call these funds *easy-in-hard-out funds.*

In some cases, investors might stay invested in a fund not because of fund provisions but because of inattention or lack of understanding. Christoffersen and Musto (2002) and Gil-Bazo and Ruiz-Verdu (2008) show that performance-sensitive investors withdraw assets from poorly performing funds leaving only performance-insensitive investors as holders of the funds' shares. Funds respond to the fact that the fund flows of the remaining investors are not sensitive to fund performance by raising expenses. To evaluate whether this mechanism can help explain fee dispersion, we estimate each fund's flow-performance sensitivity (*Flow-perf sensitivity*) by regressing monthly flows on lagged monthly net-of-expense returns using an expanding window (with a minimum of 12 monthly observations).

[20] As an alternative to matching on holdings, we identify competing funds as funds that have similar betas to a given fund. To estimate fund betas, we regress the time series of monthly returns for the fund against an intercept, MKT, SMB, HML and UMD using 3 years of data from year t to t-2. We require a minimum of 12 monthly returns to estimate the betas. Then we determine each beta's quartile and match funds if all four betas are in the same quartile of their respective distributions. Results from this strategy are not reported in the paper for reasons of brevity and are available from the authors upon request. However, these results are similar to the matching on holdings results.

[21] This measure of uniqueness is constructed at the fund-level. Note, however, that the analysis in this section is performed at the share-class level. Thus, we use the same uniqueness value for all share classes of the same fund.

In Table 6, we report the results of tests using these three measures of fund competition. In Table 6, Panel A, we report pooled averages and standard deviations of fund average uniqueness and flow-performance sensitivity across deciles of reported expense ratios. It also shows the overall (i.e., across all years) fraction of easy-in-hard-out funds per expense decile. There is a monotonically decreasing relation between fund competition and reported expense ratios: as we move from low to high reported expense deciles, the fund average uniqueness increases. This is consistent with the notion that more unique funds are able to charge higher fees, on average.

We also find, however, that the fraction of easy-in-hard-out funds increases monotonically from 0.41% for expense-decile 1 to 7.65% for expense-decile 10. Thus, fund provisions that prohibit fund exit or switching certainly play an important role in the fee setting process. Interestingly, we do not find support for a negative link between fund fees and the performance-sensitivity of flows. The patterns are non-monotonic in this case, and if we compare the most extreme expense deciles, we observe that the most expensive funds show higher flow-performance sensitivity estimates than the cheapest funds.

In a next step, reported in Table 6, Panel B, we add each of these additional fund characteristics separately to our regression-based pricing approach to investigate how they affect fees in a multivariate context. Interestingly, the coefficient on fund average uniqueness turns out to be negative and weakly significant for the full sample suggesting that competition actually works in the wrong way: i.e., more unique funds tend to charge lower fees. One potential explanation for this result is that it is driven by small funds that offer relatively unique strategies but also face fiercer competition for flows to grow. Consistent with this interpretation, the coefficient of fund average uniqueness becomes positive, economically larger and statistically very significant for the sample of large funds. Among these funds, we observe that uniqueness in terms of holdings is associated with larger expense ratios – we find that a one standard deviation change in uniqueness is associated with an extra 7 bps in fees - consistent with the notion that among these large funds competition works to some extent.

If we control for the easy-in-hard-out dummy, we find that it is associated with an approximate 20 bps increase in expected fees in both fund samples. This is consistent with the summary statistics and suggest that fund provisions that make exit and switching more costly are, as one would expect, associated with higher expected fees. Note also, that data quality on loads is relatively sparse and, thus, we end up with a much smaller sample size if we include this dummy in the analysis.

The results for flow-performance sensitivity are mixed. In the full sample, we surprisingly find a positive coefficient suggesting that funds with flows that are more sensitive to performance charge higher fees. In the case of large funds, we also find a positive, albeit insignificant, coefficient. Overall, our evidence does not support the mechanisms suggested in Christoffersen and Musto (2002) and Gil-Bazo and Ruiz-Verdu (2008).

Ultimately, however, we are interested in analyzing whether these variables lead to a noticeable reduction in fee dispersion. Panel C of Table 6 shows the corresponding results. We find that controlling for fund average uniqueness has a noticeable impact on fee dispersion. For example, in the case of large funds the average interquartile spread drops from 42 to 33 bps and the 10-90th percentile spread drops from 93 to 68 bps. Similarly, controlling for the easy-in-hard-out dummy leads to a reduction in fee dispersion in the case of the full sample. Both of these drops, however, have to be interpreted with caution, as the construction of these variables results in a reduction of the sample. Importantly, however, the remaining levels of fee dispersion are still large.

Bottom line, we find some support for the competitive mechanism documented in Wahal and Wang (2011), at least for the sub-sample of large funds. In other words, some of the spread in residual fees across similar large TNA funds is due to large funds that are more unique in terms of holdings. We also find that provisions that increase exit and switching costs help explain variation in observed fund fees. In contrast, the link between flow-performance sensitivity and fees is weak and counter-intuitive, as more expensive funds seem to also feature more flow-performance sensitivity. Importantly, these proxies help reduce fee dispersion only to a limited extent.

5.2 Distribution Channels

Another mechanism that could drive our fee dispersion results is that funds are sold through different distribution channels. To better understand the importance of this aspect, we first rerun our empirical analysis separately for institutional and retail funds, which obviously use very different distribution channels. Indeed, the literature (see Christoffersen and Musto (2002), Bris, Gulen, Kadiyala, Rau (2007) and others) has shown that institutional funds tend to have lower expenses and are presumed to be held by more sophisticated investors relative to retail funds. Thus, if holders of institutional funds are more educated about funds and have a greater influence on prices, it is possible that our results do not hold for them.

Table 1, Panel A, shows that, on average, 23.2% of all share classes are institutional share classes during the post-1999 sample period. Panel C shows, as one would expect, that the fraction of institutional share classes drops across total expense ratio deciles: while the fraction is 47.1% in the lowest deciles, it is only 1.4% in the top decile. The average expense ratio of institutional funds is 95 basis points in our sample while the one of retail funds is 152 basis points.

In Figure 5, we plot reported expenses and residual expenses separately for both retail and institutional funds. The reported and residual spreads are indeed higher for retail funds, but we still see evidence of relatively large spreads in residual expenses within the sample of institutional funds with no clear trend of decreasing expense spreads in more recent years. More specifically, the time-series averages of the 10^{th}-90^{th} percentile residual spread, for example, amount to 85 (133) basis points (see Table 7) for institutional (retail) funds. Thus, the level of fee dispersion, after controlling for a large range of fund characteristics, remains large and puzzling even *within* the separated samples of retail and institutional funds.

In a next step, we follow Bergstresser, Chalmers and Tufano (2009) and split the sample of retail funds into *directly-sold* and *broker-sold* funds based on reported 12b-1 fees. 12b-1 fees are part of the total expense ratio but are explicitly identified fees for marketing and distribution costs. Average 12b-1 fees of retail funds amount to 0.6% in the post-1999 sample. Bergstresser, Chalmers and Tufano (2009) show that, as one would expect, 12b-1 fees are substantially larger for broker-sold than for directly-sold funds. Thus, we classify retail funds to be directly-sold (broker-sold) when reported 12b-1 fees are below 0.25% (above 0.75%). On average, we observe 12b-1 fees for 74% of all retail funds during the post-1999 sample period. Out of these funds, our classification scheme determines 39%, on average, to be directly-sold and 48% to be broker-sold retail funds. For funds that we classify as directly-sold (broker-sold) funds we find average 12b-1 fees of 0.2% (0.97%).

Figure 6 shows the empirical distributions of reported and residual fees for directly-sold and broker-sold retail funds over time while Table 7 provides summary statistics of fee spreads. In the case of reported fees, we confirm our expectation that broker-sold funds have higher levels (average expense ratios of directly-sold (broker-sold) funds amount to 125 (204) basis points) and more pronounced dispersion in fees. Rather surprisingly, however, the distributions of residual fees look very similar. The time-series averages of the 10^{th}-90^{th} percentile spread, for example, are 83 (70) basis points for directly-sold (broker-sold) funds. This implies that our empirical model explains more of the variation in fees across broker-sold than across directly-sold funds. Bottom line, however, is that, within both samples, the remaining fee dispersion is economically large. Thus, our results do not seem to be driven by distribution channels.

6. Conclusion

In this paper we examine how mutual funds price their services for a large cross-section of funds (i.e., all mutual funds that focus on investing in US equities) and a long time-series of 49 years. Surprisingly, after we control for a variety of fund characteristics related to performance, service, and other features that investors are likely to care about, we find that the unexplained portion of fund expenses exhibits considerable dispersion. Fee dispersion in this context implies some sort of pricing inefficiency in the sense that funds with similar characteristics charge different fees. Importantly, we find this result also among the subset of S&P 500 index funds and among the subset of largest funds. Furthermore, we find similar levels of fee dispersion when using a model-free matching approach based on fund holdings.

The most notable result, in our opinion, is that this dispersion has not declined over time. In contrast, it actually seems to have been much less of an issue during early years of the sample, increased dramatically during the 80ties and 90ties, and stayed at elevated levels in recent years. These time-series dynamics are particularly pronounced for the subset of large funds.

The level of dispersion that we find is substantial in economic terms. For example, the costs for getting it wrong – investing in high expense funds when close-to-identical low expense funds are available – are large; we show that a low-expense fund investor would have earned approximately 71 to 145% more in cumulative abnormal returns than a high-expense fund investor over our sample period. Similarly, a simple back-of-the-envelope calculation suggests that in 2014, approximately USD 7 billion, or 22% of total fee revenues earned in our sample, are associated with fees that exceed expected fees according to our simple pricing model.

We also show that competition does work, at least to some extent, as there is evidence that large funds, which operate in crowded strategies, having holdings that are similar to many other funds, tend to charge lower fees. We also document that, not surprisingly, funds that attract investors with more advertising than other funds, and then put in place potential exit barriers, what we refer to as easy-in-hard-out funds, charge higher fees and that distribution channels matter for fees. However, none of these mechanisms result in a substantial reduction of fee dispersion.

Overall, our results pose an important and multi-dimensional puzzle regarding the fees charged in the mutual fund industry. Potential explanations of our results are, of course, that we do not control for the complete set of fund characteristics that affect fund fees[22] or that we do not capture relevant characteristics of the fund industry such as frictions accurately. While we are unable to completely rule these out, we also find it implausible to expect them to substantially reduce the large spreads in fees, particularly given the comprehensive set of robustness tests that we employ in the paper.

One explanation that would be consistent with the above patterns is the existence of multiple investor clienteles with varying levels of sophistication and access to information. For example, retail investors often times have limited knowledge of financial products. Thus, issues such as financial literacy and

[22] One specific example for such a fund characteristic is trust in the fund manager. Gennaioli, Shleifer, and Vishny (2015) develop a model in which investors pick portfolio managers on performance and trust. Investor trust in the manager lowers an investor’s perception of the portfolio's risk, and allows managers to charge higher expenses to investors who trust them more.

advising of households should be of first order importance for regulators. Of course, it is not obvious that enabling (retail) investors with the basic tools to select funds would solve the issue of fee dispersion.[23] As pointed out by Carlin and Manso (2011) funds may optimally react to investor learning by increasing the level of obfuscation (i.e., by making it harder for investors to learn). They argue, however, that an increase in competition should lower the incentives for obfuscation and, thus, should enable investors to learn more quickly.[24] Thus, from a regulator's perspective it is also important to increase transparency and comparability in the industry.

[23]Perhaps smarter investors are more fee aware. In a study of investor heterogeneity based on intelligence, Grinblatt, Ikaheimo, Keloharju, and Knüpfer (2013) find that high IQ investors in Finland prefer low fee funds.

[24] Ellison and Wolitzky (2012) develop a static model of obfuscation and find that competition might actually lead to more confiscation, increased search costs and more price dispersion.

References

Bakos, Y., 2001, The Emerging Landscape of Retail E-commerce, *Journal of Economic Perspectives,* 15, 69-80.

Barras, L., Scaillet, O. and R. Wermers, 2010, False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas, *The Journal of Finance*, Vol. LXV, No. 1, 179-216.

Bergstresser, D., J. M. R. Chalmers and P. Tufano, 2009, Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry, *The Review of Financial Studies*, 22, 4129-4156.

Berk, J. B., and R. C. Green, 2004, Mutual Fund Flows and Performance in Rational Markets, *Journal of Political Economy*, 112, 1269-1295.

Bessler, W., D, Blake, P. Luckiff, and I. Tonks, 2010, Why is Persistent Mutual Fund Performance so Difficult to Achieve? The Impact of Management Turnover and Fund Flows, working paper.

Bravin, Jess and Moyer, Liz, High Court Ruling Adds Protections for Investors in 401(k) Plans, *Wall Street Journal*, May 18, 2015.

Bris, A., H. Gulen, P. Kadiyala, and P. Raghavendra Rau, 2007, Good Stewards, Cheap Talkers, or Family Men? The Impact of Mutual Fund Closures on Fund Managers, Flows, Fees, and Performance, *The Review of Financial Studies*, 20, 953-982

Brown, J. and A. Goolsbee, 2002, Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry, *Journal of Political Economy*, 110, 481-507.

Brynjolfsson, E. and M. Smith, 2000, Frictionless Commerce? A Comparison of Internet and Conventional Retailers, *Management Science,* 46(4), 563-585.

Carhart, M., 1997.On persistence in mutual fund performance. *Journal of Finance*, 52, 57-82.

Carlin, B. I. and G. Manso, 2011, Obfuscation, Learning, and the Evolution of Investor Sophistication, *The Review of Financial Studies*, 24, 754-785.

Christoffersen, S. and D. Musto, 2002, Demand Curves and the Pricing of Money Management, *Review of Financial Studies, Vol. 15, No. 5.*

Cremers, M., M. Ferreira, P. Matos, and L. Starks, 2015, Indexing and Active Fund Management: International Evidence, *Journal of Financial Economics*, forthcoming.

Das, S. R. and R. K. Sundaram, 2002, Fee Speech: Signaling, risk-sharing, and the impact of fee structures on investor welfare, *The Review of Financial Studies*, 15, 1465-1497.

Editorial Board, 2015, Protecting Fragile Retirement Nest Eggs, *New York Times*, February 28, 2015.

Ellison, G. and A. Wolitzky, 2012, A Search Cost Model of Obfuscation, *Rand Journal of Economics*, 3, 1756-2171.

Elton, E., M. Gruber, and J. Busse, 2004, Are Investors Rational? Choices Among Index Funds, *Journal of Finance*, Vol. 59, No. 1.

Elton, E., M. Gruber, and J. C. Rentzler, 1989, New public offerings, information, and investor rationality: the case of publicly offered commodity funds, *Journal of Business*, Vol. 62, 1-15.

Gennaioli, N., A. Shleifer, R. W. Vishny, 2015, Money Doctors, The *Journal of Finance*, 70, 91-114

Gil-Bazo, J., and P. Ruiz-Verdu, 2008, When Cheaper is Better: Fee Determination in the Market for Equity Mutual Funds, *Journal of Economic Behavior and Organization*, pp. 871-885.

Gil-Bazo, J., and P. Ruiz-Verdu, 2009, The Relation between Price and Performance in the Mutual Fund Industry, *Journal of Finance*, 64, 2153-2183.

Grinblatt, Mark and Ikaheimo, Seppo and Keloharju, Matti and Knüpfer, Samuli, IQ and Mutual Fund Choice (May 14, 2013). AFA 2013 San Diego Meetings Paper, San Diego Jan. 4-6, 2013. Available at SSRN: http://ssrn.com/abstract=2021957 or http://dx.doi.org/10.2139/ssrn.2021957

Haslem, J. A., Baker, H. K. and D. M. Smith, 2006, Are retail S&P 500 index funds a financial commodity? Insights for investors, *Financial Services Review*, 15, 99-116.

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

Khorana, A., H. Servaes, and P. Tufano, 2009, Mutual Fund Fees around the World, *Review of Financial Studies*, 22 (3), 1279-1310.

Lach, S., 2002, Existence and Persistence of Price Dispersion: an Empirical Analysis, *The Review of Economics and Statistics*, 84 (3), 433-444.

Light, Joe, 2013, Don't Pay High Fees for Index Funds, *Wall Street Journal*, May 3, 2013.

Nakamura, L., 1999, The Measurement of Retail Output and the Retail Revolution, *Canadian Journal of Economics*, 32(2), 408-425.

Nanda, V., M. P. Narayanan and V. A. Warther, 2000, Liquidity, investment ability, and mutual fund structure, *Journal of Financial Economics*, 57, 417-443.

Pástor, L. and R. Stambaugh, 2002, Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics*, 63, 315-349.

Pástor, L. and R. Stambaugh, 2012, On the Size of the Active Management Industry, *Journal of Political Economy* 120, 740–781

Porter, Eduardo, 2015, Americans Aren't Saving Enough for Retirement, but One Change Could Help, *New York Times*, March 3, 2015.

Pratt, J., D. Wise and R. Zeckhauser, 1979, Price Differences in Almost Competitive Markets, *Quarterly Journal of Economics*, 939, 189-211.

Ramadorai, T. and M. Streatfield, 2011, Money for Nothing? Understanding Variation in Reported Hedge Fund Fees, Working Paper.

Scholten, P. and S. Smith, 2002, Price Dispersion Then and Now: Evidence from Retail and E-tail Markets, *Advances in Microeconomics: Economics of the Internet and e-Commerce* 11, 63-88.

Sorensen, A., 2000, Equilibrium Price Dispersion in Retail Markets for Prescription Drugs, *Journal of Political Economy*, 108(4), 833-862.

Wahal, S., and A. Wang, 2011, Competition among mutual funds, *Journal of Financial Economics*, 99, 40-59.

Data Appendix

Table A. Sample Selection

We follow Bessler et al. (2010) who use a combination of Lipper codes, Wiesenberger codes and Strategic Insight codes to identify domestic equity funds. Specifically, we include funds in our sample with the following classification codes:

1. Lipper: CA, EI, EIEI, G, GI, I, LCCE, LCGE, LCVE, MC, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, SCCE, FS, H, NR, S, SESE, TK, TL, UT.
2. Wiesenberger: AGG, G, G-I, G-I-S, G-S, G-S-I, GCI, GRI, GRO, IG, I-G-S, I-S, I-S-G, IEQ, ING, LTG, MCG, S-G, S-GI, S-I-G, S-I, SCG, ENR, FIN, HLT, TCH, UTL.
3. Strategic Insight: AGG, GMC, GRI, GRO, ING, SCG, ENV, FIN, HLT, NTR, SEC, TEC, UTI.

Table B. Variable Construction and Definitions

Variable Name	Variable Definition	Source
Δ*(expense ratio)*	Yearly change in the expense ratio.	Calculated
12b-1 fees	Ratio of the total assets attributed to marketing and distribution costs. Available since 1992.	CRSP MF Database
Alpha	Four factor alpha. For each December and each fund, we estimate a monthly four-factor alpha using three years of monthly after-expense, excess fund returns. *Alpha* is the estimated monthly alpha (i.e., the constant in the time-series regression) multiplied by 12. *Before-expense alphas* are estimated in the same way using monthly before-expense fund returns, which are calculated by adding the total expense ratio to monthly after-expense returns. *Expanding window alphas* are estimated from an expanding estimation window rather than a rolling window of three years. *Filtered alphas* replace alpha estimates by zero when the corresponding t-statistic of the alpha estimate is below three in absolute terms. *Full sample alphas* are unconditional alphas estimates exploiting all available data per fund.	Calculated
Annual flow	Annual fund flow. It is estimated as $Flow_t = \left(TNA_t - TNA_{t-12}\left(1+return_t\right)\right)/TNA_{t-12}$ and is winsorized at the 1% level. $Return_t$ is the return over the prior 12 months.	Calculated
Annual return	Annual fund return, gross of expenses. We first compounding monthly after-expense returns within the previous 12 months. This 12-month return is then added to the annual total expense ratio. Monthly return values are calculated as a change in NAV including reinvested dividends from one period to the next. NAVs are net of all management expenses and 12b-fees. Front and rear load fees are excluded. Annual return is in decimal form, that is 0.01 is 1%.	CRSP MF Database and Calculated
Back-end load	The rear load is a fee charged by the fund when an investor withdraws funds. The rear load typically varies by investment level and duration of the investment. The rear load value is the equal weighted average across all reported rear load values across these dimensions.	CRSP MF Database
B*eta_mkt* *Beta_hml* *Beta_smb* *Beta_umd*	Fund betas from the four-factor model. Each December and for each fund, we estimate the monthly four-factor model betas using 3 years of monthly after-expense excess return. Refer to the information on the calculation of *Alpha* for details regarding *before-expense, expanding window, filtered* and *full sample* estimates of betas.	Calculated
Carhart alpha	For each month, and for each fund, we first estimate a monthly after-expense alpha as the difference between the fund's after-expense excess return in month *t* and the realized risk premium, defined as the vector of betas times the vector of contemporaneous factor realizations in month *t* (see Carhart (1997) and Gil-Bazo and Ruiz-Verdu (2009)). Betas are estimated from 3 years of monthly after-expense fund returns and lagged by one month. The Carhart alpha that we use in the analysis is yearly and is calculated by compounding monthly Carhart alphas estimated over the previous 12 months. Refer to the information on the calculation of *Alpha* for details regarding *before-*	Calculated

Variable Name	Variable Definition	Source
	expense, expanding window, filtered and *full sample* estimates of Carhart alphas. Carhart alpha is in decimal form, that is 0.01 is 1%.	
Easy-in-hard-out funds	Dummy variable equal to 1, 0 else, for funds whose back-end load is in the top decile and whose 12b-1 expense ratio is in the top quartile.	Calculated
Expense ratio	Annual ratio of total investment that shareholders pay for the fund's operating expenses, which include 12b-1 fees.	CRSP MF Database
Family1 dummy *Family10 dummy* *Family100 dummy* *Family250 dummy*	A dummy variable equal to 1, 0 else, if a fund is part of a management company with more than 1 (10) [100] {250} funds associated with it. The standard case in our analysis is *Family100 dummy*.	Calculated.
Flow-perf sensitivity	For each fund and each year, we estimate the fund's flow-performance sensitivity as the coefficient of lagged monthly performance in a regression that explains monthly flows. The regression starts with 1 year of monthly data and uses an expanding window.	Calculated
Fund age	Age of fund calculated as the difference between current year and year of fund initiation.	Calculated
Fund average uniqueness	For each fund, this is the average of its *Uniqueness* with all other funds, estimated yearly.	Calculated
Institutional dummy *Open dummy* *ETF dummy*	A dummy variable equal to 1, 0 else, if a fund is an institutional fund, or open to new investment, or is an ETF.	CRSP MF Database
ln(MgmtComp TNA)	The natural log of each December's sum of total net assets of all funds belonging to the same management company.	Calculated
ln(TNA)	The natural log of total net assets per fund as of December-end.	Calculated
Persistence dummy	A dummy variable equal to 1, 0 else, for a given fund in year t if the fund is among the top-20% funds with respect to yearly net performance in years t-1 and t-2. The term "*before-expense persistence dummy*" refers to a persistence dummy that is based on gross returns rather than net returns.	Calculated
R^2	For each December and each fund, we estimate the four-factor model using 3 years of monthly fund returns. Then we collect the adjusted R^2 of these models.	Calculated
Sdmret	Standard deviation of monthly net-of-fee returns calculated from 3 years of monthly fund returns. Sdmret is in decimal form, that is 0.01 is 1%.	Calculated
Style Fixed Effects	Fund styles are defined using the classification codes described in Table A of the Data Appendix.	CRSP MF Database
TNA	Total net assets as of December-end in millions of USD.	CRSP MF Database
Tstat alpha	The t-statistic associated with *alpha*. Refer to the information on the calculation of *Alpha* for details regarding *expanding window* and *full sample* estimates of Tstat alpha.	Calculated
Turnover	Annual fund turnover is calculated as the minimum of aggregated sales or aggregated purchases of securities, divided by the average 12-month total net assets of the fund.	CRSP Mutual Fund Database

Variable Name	Variable Definition	Source
Uniqueness	It is the sum, across all holdings, of absolute differences in weights for a given pair of funds, estimated yearly. The measure is bounded between zero (perfect overlap) and two (no overlap). The uniqueness variable is based on fund holdings which are available at the fund rather than the share class level. As a consequence, all results that involve this variable are at the fund level. For each fund, its uniqueness value is defined as the value of this sum for the fund pair that results in the maximal holdings overlap.	Calculated

Table 1. Summary Statistics

The table reports summary statistics and a correlation table of our sample of domestic equity mutual funds (see Table A in the Data Appendix for a detailed description of the sample). The standard deviations in Panels A and B are computed as the mean of yearly cross-sectional standard deviations. The data covers the period of 1966 to 2014 and is a yearly panel. Variables are defined in Table B in the Data Appendix. The table focuses on the variables used in our base-case fund pricing model. Some information is only available after 1999 (e.g., information on management companies) and, thus, we split the sample into a pre-1999 and a post-1999 subset. Panel A presents full period summary statistics. Panel B and C summarize the sample by expense ratio deciles in the pre- and post-1999 periods, respectively. Panel D contains correlations. The last column in Panel B and C reports the difference between decile 1 and decile 10. Stars indicate significance at the 1% (***), 5% (**) and 10% (*) level. *Annual return, Carhart alpha, the R-Square,* the standard deviation of monthly returns (*Sdmret*) and all summary statistics of dummies are in decimal form, that is 0.01 is 1%. *Annual Flow* and *Turnover* are in percentages. *Expense ratio* and Δ(*Expense ratio*) are in basis points. TNA is total net assets as of December-end in millions of USD.

Panel A. Full Sample

	Pre - 1999		Post - 1999	
	Mean	**SD**	**Mean**	**SD**
Number of funds per year	1051	968	7931	1157
Expense ratio	121.000	99.000	138.000	98.000
Δ(Expense ratio)	7.000	81.000	0.000	56.000
Annual return	0.122	0.182	0.075	0.226
Carhart alpha	-0.014	0.101	-0.012	0.092
Beta_mkt	0.857	0.309	0.978	0.331
Beta_smb	0.226	0.440	0.161	0.361
Beta_hml	-0.012	0.420	-0.008	0.378
Beta_umd	0.055	0.289	0.018	0.195
R^2	0.779	0.241	0.870	0.155
Annual flow	1.008	4.696	0.972	4.878
ln(TNA)	3.699	2.293	3.537	2.592
TNA	361.379	1679.748	530.320	2849.820
Fund age	8.088	8.612	8.985	8.169
Sdmret	0.047	0.032	0.051	0.025
Turnover			1.009	2.551
ln(MgmtComp TNA)			8.864	2.657
Family1 dummy			0.882	0.323
Family10 dummy			0.802	0.399
Family100 dummy			0.524	0.499
Family250 dummy			0.250	0.433
Institutional dummy			0.232	0.422
Open dummy			0.727	0.446
ETF dummy			0.020	0.139

Panel B. Summary Statistics by Expense Ratio Deciles --- Pre - 1999 Sample

	Decile 1		Decile 3		Decile 5		Decile 7		Decile 10		Decile 1 - 10
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Diff**
Expense ratio	31	28	75	29	104	21	136	28	268	154	-237***
Δ(Expense ratio)	-14	58	-3	33	2	27	4	39	42	149	-56***
Annual return	0.171	0.216	0.138	0.157	0.136	0.169	0.136	0.183	0.119	0.209	0.052***
Carhart alpha	-0.010	0.080	-0.010	0.072	-0.010	0.084	-0.016	0.109	-0.030	0.137	0.020***
Beta_mkt	0.856	0.407	0.842	0.287	0.853	0.279	0.851	0.282	0.858	0.346	-0.002
Beta_smb	0.072	0.446	0.113	0.330	0.197	0.392	0.273	0.432	0.384	0.540	-0.312***
Beta_hml	0.034	0.469	0.002	0.316	-0.018	0.383	-0.013	0.406	-0.041	0.536	0.075***
Beta_umd	0.008	0.254	0.034	0.209	0.053	0.245	0.072	0.288	0.085	0.396	-0.077***
R^2	0.810	0.268	0.809	0.235	0.797	0.225	0.765	0.242	0.702	0.256	0.108***
Annual flow	1.213	5.292	0.568	3.382	0.533	3.026	0.769	3.744	1.203	4.709	0.010
ln(TNA)	4.916	2.433	5.113	1.892	4.477	1.638	3.890	1.686	2.370	1.740	2.545***
TNA	1550.352	5290.461	691.041	1775.102	285.858	766.173	198.393	669.415	46.796	133.216	1503.556***
Fund age	9.778	10.120	12.047	9.861	11.178	9.026	8.490	7.369	6.753	7.154	3.024***
Sdmret	0.037	0.032	0.039	0.017	0.042	0.018	0.046	0.020	0.051	0.023	-0.014***

Panel C. Summary Statistics by Expense Ratio Deciles --- Post - 1999 Sample

	Decile 1		Decile 3		Decile 5		Decile 7		Decile 10		Decile 1 - 10
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Diff**
Expense ratio	37	19	95	9	123	11	157	17	261	238	-224***
Δ(Expense ratio)	-1	9	-1	9	-1	10	-1	13	6	173	-7***
Annual return	0.071	0.198	0.080	0.219	0.081	0.225	0.076	0.230	0.066	0.253	0.005*
Carhart alpha	-0.004	0.069	-0.008	0.087	-0.010	0.089	-0.013	0.096	-0.022	0.116	0.018***
Beta_mkt	0.899	0.285	0.995	0.268	0.993	0.246	0.983	0.358	0.987	0.497	-0.088***
Beta_smb	0.058	0.315	0.123	0.334	0.187	0.352	0.201	0.376	0.257	0.405	-0.199***
Beta_hml	0.048	0.306	-0.001	0.364	0.008	0.359	-0.023	0.398	-0.090	0.471	0.138***
Beta_umd	-0.005	0.155	0.017	0.191	0.024	0.184	0.021	0.205	0.025	0.237	-0.029***
R^2	0.878	0.195	0.880	0.144	0.878	0.132	0.867	0.148	0.823	0.185	0.055***
Annual flow	1.019	4.872	1.215	5.688	0.935	4.548	0.946	4.564	0.833	4.550	0.186**
ln(TNA)	5.147	2.628	4.292	2.582	3.723	2.443	3.041	2.430	1.919	1.982	3.23***
TNA	1958.091	6605.952	674.114	2390.214	321.450	978.862	185.910	652.200	36.535	124.755	1921.556***
Fund age	8.851	8.706	10.431	9.907	9.424	8.724	8.207	7.146	7.703	6.071	1.148***
Sdmret	0.044	0.022	0.050	0.024	0.051	0.023	0.053	0.025	0.059	0.030	-0.015***
Turnover	0.405	0.512	0.815	0.793	0.936	1.472	1.180	2.619	1.738	4.861	-1.333***
ln(MgmtComp TNA)	10.134	2.486	8.947	2.618	8.595	2.559	8.611	2.765	8.000	2.714	2.134***
Family1 dummy	0.888	0.316	0.860	0.347	0.883	0.321	0.884	0.320	0.889	0.315	-0.001
Family10 dummy	0.861	0.346	0.780	0.414	0.776	0.417	0.787	0.410	0.792	0.406	0.069***
Family100 dummy	0.573	0.495	0.510	0.500	0.482	0.500	0.517	0.500	0.461	0.499	0.112***
Family250 dummy	0.243	0.429	0.272	0.445	0.240	0.427	0.240	0.427	0.173	0.379	0.070***
Institutional dummy	0.471	0.499	0.417	0.493	0.221	0.415	0.142	0.349	0.014	0.117	0.458***
Open dummy	0.744	0.437	0.741	0.438	0.737	0.441	0.732	0.443	0.680	0.467	0.064***
ETF dummy	0.128	0.334	0.013	0.112	0.000	0.009	0.001	0.024	0.001	0.022	0.127***

Panel D. Pooled Correlation

	Expense ratio	Δ(Exp. ratio)	A. Ret.	Carhart alpha	Beta mkt	Beta hml	Beta smb	Beta umd	R^2	Annual Flow	ln(TNA)	Fund age	Sdmret	Turnover
Exp. ratio														
Δ(Exp. ratio)	0.54													
Annual return	-0.04	-0.03												
Carhart Alpha	-0.08	-0.04	0.37											
Beta_mkt	0.03	-0.01	0.00	-0.07										
Beta_smb	0.10	0.01	0.08	-0.02	0.17									
Beta_hml	-0.03	0.00	-0.04	-0.06	-0.16	0.01								
Beta_umd	0.02	-0.01	-0.01	-0.08	0.07	0.10	-0.09							
R^2	-0.09	-0.02	0.00	-0.02	0.37	0.02	-0.10	-0.05						
Annual Flow	0.00	0.01	0.05	0.07	0.00	0.01	0.01	0.02	-0.02					
ln(TNA)	-0.25	-0.03	0.10	0.05	-0.04	-0.05	0.02	0.00	-0.01	-0.06				
Fund age	-0.05	-0.01	0.05	-0.01	-0.06	-0.07	-0.02	-0.03	-0.01	-0.16	0.42			
Sdmret	0.13	0.02	-0.07	-0.05	0.47	0.32	-0.21	0.05	0.11	0.00	-0.09	-0.08		
Turnover	0.11	0.00	-0.02	-0.01	0.02	0.03	-0.04	0.03	-0.10	0.02	-0.09	-0.05	0.13	
Ln(MgmtCompany TNA)	-0.09	-0.07	0.01	0.03	0.09	-0.10	0.00	-0.04	0.17	-0.04	0.31	0.20	0.00	-0.02

Table 2. Base-Case Fund Expense Regressions

The table reports results of Fama-MacBeth regressions in which yearly expense ratios are regressed on lagged fund characteristics (see Table A in the Data Appendix for a detailed description of the sample). We standardize all the independent variables to mean 0 and standard deviation 1 using each variable's yearly cross-sectional mean and SD. The data covers the period of 1966 to 2014 and is a yearly panel. Variables are defined in Table B in the Data Appendix. All variables are lagged by one year. Expense ratio (the dependent variable in all regressions) is in basis points. We split the sample into pre-and post-1999 sub-periods since several variables are only available after 1999. The specifications reported in this table represent the base-case specifications. Coefficient estimates are time series averages of cross-sectional regression coefficients obtained from annual cross-sectional regressions. Values in parentheses are t-statistics. We perform the regressions on the full sample of mutual funds, for the largest quintile of annually ranked *TNA* funds, and for S&P 500 index funds (post-1999 only for index funds). Panel A reports the regression results. In Panel B, for the independent variables used in the fee regressions, we report in column (1) the time-series average of the cross-sectional mean (μ), in column (2) the time-series standard deviation (σ_μ) of the cross-sectional mean, and in column (3), the time-series average of the cross-sectional SD (σ).

Panel A. Base-case Fama-MacBeth Regressions of Yearly Expense Ratios on Lagged Fund Characteristics

	Pre-1999		Post-1999		
	Full	**Large**	**Full**	**Large**	**S&P500**
Intercept	119.01	114.38	143.43	146.09	74.17
	(31.3)	(18.13)	(44.68)	(43.40)	(14.88)
Annual return$_{t-1}$	-5.87	-1.76	-8.11	-2.50	6.96
	(-4.04)	(-1.52)	(-2.81)	(-1.75)	(0.93)
Beta_mkt $_{t-1}$	-1.09	-6.89	-3.78	-0.80	-12.68
	(-0.64)	(-1.21)	(-1.45)	(-0.18)	(-2.91)
Beta_hml $_{t-1}$	6.36	0.32	17.49	14.77	-4.38
	(3.87)	(0.06)	(1.73)	(2.23)	(-2.58)
Beta_smb $_{t-1}$	-0.49	6.32	-0.04	-4.06	4.34
	(-0.39)	(2.68)	(-0.02)	(-2.45)	(1.11)
Beta_umd $_{t-1}$	2.95	5.08	0.22	-4.47	-0.46
	(2.33)	(3.66)	(0.12)	(-2.29)	(-0.18)
R^2_{t-1}	-3.07	-0.13	-10.93	-5.83	2.49
	(-2.45)	(-0.06)	(-9.61)	(-10.31)	(-0.82)
Flow $_{t-1}$	0.20	4.41	-1.38	-2.41	1.95
	(0.17)	(1.30)	(-3.74)	(-3.28)	(2.12)
LN(TNA) $_{t-1}$	-30.86	-18.30	-24.54	-24.81	-22.72
	(-24.93)	(-22.59)	(-28.40)	(-23.46)	(-11.22)
Fund age $_{t-1}$	1.32	-7.46	3.15	-1.59	-1.48
	(1.32)	(-4.71)	(2.81)	(-1.65)	(-2.16)
Sdmret $_{t-1}$	16.31	26.41	34.49	4.37	1.04
	(3.09)	(1.26)	(3.34)	(1.00)	(0.43)
Turnover $_{t-1}$			2.12	10.83	2.89
			(3.98)	(6.42)	(1.73)
LN(MgmtComp TNA) $_{t-1}$			-11.43	-19.03	0.12
			(-8.12)	(-16.79)	(0.07)
Family100 dummy			13.71	19.11	11.59
			(7.44)	(7.58)	(4.55)
Institutional dummy			-51.65	-36.58	-28.05
			(-15.59)	(-9.90)	(-24.51)
Open dummy			7.49	2.41	-13.80
			(1.57)	(0.79)	(-4.75)
ETF dummy			-44.19	-37.10	19.91
			(-30.27)	(-22.38)	(5.61)
Avg. R-Squared	0.25	0.31	0.27	0.34	0.59
# of Obs.	26231	5232	107722	21538	1941

Panel B. Time-series Average of the Cross-Sectional Mean (μ), the Time-Series Standard Deviation (σ_μ) of the Cross-Sectional Mean, and the Time-Series Average of the Cross-Sectional SD (σ) of Fund Characteristics.

	Pre - 1999						Post - 1999								
	Full Sample			Largest Funds			Full Sample			Largest Funds			S&P 500 Index Funds		
	μ	σ_μ	σ	μ	σ_μ	σ	μ	σ_μ	σ	μ	σ_μ	σ	μ	σ_μ	σ
Annual return$_{t-1}$	0.12	0.15	0.12	0.12	0.14	0.10	0.08	0.19	0.14	0.09	0.19	0.14	0.08	0.19	0.04
Beta_mkt $_{t-1}$	0.84	0.06	0.32	0.83	0.10	0.29	0.99	0.04	0.28	0.97	0.04	0.25	0.99	0.03	0.01
Beta_hml $_{t-1}$	0.23	0.09	0.44	0.11	0.07	0.28	0.17	0.04	0.36	0.12	0.03	0.33	-0.16	0.02	0.01
Beta_smb $_{t-1}$	-0.02	0.05	0.40	-0.02	0.05	0.30	0.00	0.07	0.37	0.01	0.09	0.36	0.01	0.02	0.01
Beta_umd $_{t-1}$	0.08	0.07	0.31	0.07	0.05	0.19	0.02	0.05	0.19	0.02	0.05	0.17	-0.02	0.03	0.01
R^2 $_{t-1}$	0.77	0.04	0.24	0.81	0.09	0.24	0.87	0.03	0.14	0.88	0.03	0.15	1.00	0.01	0.00
flow $_{t-1}$	0.57	0.56	2.71	0.27	0.33	1.44	0.95	0.38	4.59	0.35	0.15	2.27	-0.78	2.71	9.51
ln(TNA) $_{t-1}$	3.83	0.49	1.84	6.39	0.54	0.77	3.66	0.33	2.52	7.04	0.28	1.02	4.95	0.21	2.55
Fund age $_{t-1}$	9.74	3.21	6.10	13.08	4.29	5.94	8.90	1.84	7.92	14.54	1.38	10.91	8.25	3.55	4.92
Sdmret $_{t-1}$	0.05	0.01	0.02	0.04	0.01	0.02	0.05	0.02	0.02	0.05	0.01	0.02	0.05	0.01	0.00
Turnover $_{t-1}$							1.01	0.19	2.28	0.66	0.11	0.69	0.11	0.03	0.19
ln(MgmtComp TNA) $_{t-1}$							8.62	1.19	2.55	10.13	0.78	1.80	8.36	1.46	3.13

Table 3. Residual Expense Spreads

This table presents time-series averages of the cross-sectional residual (using the base-case pricing models of Table 2) and reported fee spreads. The unit of observation is basis points (bps). See Table A in the Data Appendix for a detailed description of the sample. The data covers the period of 1966 to 2014 and is a yearly panel (for index funds, the sample period is 1999 to 2014). The variables are defined in Table B of the Data Appendix. We report these statistics separately for the full sample of US-equity mutual funds, for the largest funds (top TNA quintile) and for S&P 500 index funds.

	Mean Reported Fee Spread (bps)		
	25^{th} to 75^{th} Percentile	**10^{th} to 90^{th} Percentile**	**1^{th} to 99^{th} Percentile**
Full Sample	81	149	283
Largest-TNA-Funds	52	119	200
S&P 500 Index Funds (Post-1999 Only)	50	116	150

	Mean Residual Fee Spread (bps)		
	25^{th} to 75^{th} Percentile	**10^{th} to 90^{th} Percentile**	**1^{th} to 99^{th} Percentile**
Full Sample	62	123	246
Largest-TNA-Funds	41	92	175
S&P 500 Index Funds (Post-1999 Only)	32	61	111

Table 4. Share Class Aggregation and Fee Dispersion

This table presents time-series averages of the cross-sectional residual fee spreads, in basis points (bps), for our sample of mutual funds (see Table A in the Data Appendix for a detailed description of the sample) after individual share classes are aggregated to the fund level. Panel A reports mean residuals for the full sample, Panel B reports residuals spreads for the top quintile of annually ranked TNA funds. The sample period covers 1980 to 2014, as share classes cannot be identified in earlier years. Row 1 in each panel reports the results from the base-case specification at the share class level. Rows 2-4 report results for samples in which we aggregate share classes using three different aggregation schemes: equal-weighting (specification 2), value-weighting (specification 3) and selection of the largest share class (specification 4). In each case, we re-estimate the base-case fund pricing regressions for each sample.

Panel A. Full Sample

	Mean Residual Spread (bps)		
	25^{th} to 75^{th} Percentile	**10^{th} to 90^{th} Percentile**	**1^{th} to 9^{th} Percentile**
No Aggregation	63	125	242
Aggregated Share Classes (EW)	63	126	266
Aggregated Share Classes (VW)	58	121	263
Aggregated Share Classes (Largest)	54	116	265

Panel B. Top Quintile of Funds

	Mean Residual Spread (bps)		
	25^{th} to 75^{th} Percentile	**10^{th} to 90^{th} Percentile**	**1^{th} to 9^{th} Percentile**
No Aggregation	50	101	183
Aggregated Share Classes (EW)	45	87	163
Aggregated Share Classes (VW)	39	79	157
Aggregated Share Classes (Largest)	37	76	158

Table 5. Differences in Reported Expense Ratios and Residual Expense Ratios for Holdings-matched Fund Pairs

The table reports mean absolute differences in reported expense (residual expense) ratios and spreads of absolute differences in reported expense (residual expense) ratios for matched fund pairs. The matching is based on holdings' overlap measured in terms of "uniqueness." For each possible fund pair, we estimate, yearly, the sum, across all holdings, of absolute differences in weights. For each fund, its uniqueness value is defined as the value of this sum for the fund pair that results in the maximal holdings overlap. The uniqueness measure is bounded between zero (perfect overlap) and two (no overlap). We refer to this sample of matched fund pairs as the "full pairs sample." We rank uniqueness into quintiles, where quintile 1 contains the most similar fund pairs and quintile 5 contains the least similar funds pairs. We also define "very similar funds" as the bottom decile of the uniqueness measure. In Panel A we present mean absolute differences and interquartile ranges (IQR) for fund characteristics of the matched pairs. In Panel B we present mean absolute differences, interquartile ranges and inter-percentile spreads for reported expense ratios and residual expense ratios, in basis points. The residual expense ratios are from the base-case expense regressions of Table 2. All variables are contemporaneous to the matching of fund pairs.

Panel A. Differences in Fund Characteristics for Matched Pairs

		Avg. # of fund pairs/year	Uniqueness (Matching Criterion)		Annual return		R^2		Beta_mkt		Beta_smb		Beta_hml		Beta_umd	
			Mean	IQR	Mean	IQR	Mean	IQR	Mean	IQR	Mean	IQR	Mean	IQR	Mean	IQR
Full pairs sample		1145.4	1.05	0.58	0.07	0.07	0.08	0.08	0.16	0.18	0.23	0.25	0.2	0.23	0.12	0.13
Very similar funds		114.94	0.15	0.12	0.03	0.03	0.03	0.02	0.15	0.15	0.08	0.08	0.09	0.09	0.05	0.05
Uniqueness Quintiles	**1**	229.41	0.32	0.42	0.03	0.03	0.03	0.03	0.13	0.15	0.09	0.1	0.1	0.11	0.06	0.07
	2	229.12	0.9	0.34	0.05	0.05	0.05	0.05	0.13	0.14	0.14	0.14	0.16	0.17	0.1	0.1
	3	229	1.18	0.2	0.06	0.06	0.06	0.06	0.16	0.16	0.18	0.19	0.2	0.21	0.12	0.12
	4	229.12	1.36	0.18	0.08	0.08	0.09	0.09	0.18	0.18	0.27	0.29	0.25	0.26	0.14	0.15
	5	228.71	1.49	0.25	0.11	0.12	0.13	0.14	0.22	0.23	0.45	0.52	0.31	0.33	0.18	0.18

Panel B. Differences in Expenses for Matched Pairs

		Reported Expense Ratio				Residual Expense Ratio			
Spreads		Mean	IQR	10^{th} to 90^{th}	1^{st} to 99^{th}	Mean	IQR	10^{th} to 90^{th}	1^{st} to 99^{th}
Full pairs sample		49	54	103	219	46	47	94	206
Very similar funds		41	48	94	174	38	41	84	167
Uniqueness Quintiles	**1**	42	49	101	176	40	41	83	172
	2	43	51	98	192	40	43	84	176
	3	47	52	98	204	45	48	93	190
	4	51	55	104	240	50	52	104	214
	5	63	57	111	321	56	53	105	274

Table 6. Fund Competition

The table summarizes results for competition-related proxies. Variables are defined in Table B in the Data Appendix. All results are based on the post-1999 sample. Panel A reports means and standard deviations across reported expense ratio deciles. The last column in Panel A reports the difference between average characteristics in deciles 1 and 10. Stars indicate significance at the 1% (***), 5% (**) and 10% (*) level. Panel B reports the coefficients these variables receive when added separately to the base-case specifications described in Table 2. We distinguish the full sample of funds and the sub-sample of large funds (i.e., top TNA quintile). We standardize all independent variables to mean 0 and standard deviation 1 using each variable's yearly cross-sectional mean and SD. Panel C reports mean residual expense spreads for the full sample and for the top quintile of annually ranked TNA funds based on the corresponding fee-pricing models described in Panel B. All variables are measured at the share class level.

Panel A. Summary Statistics of Competition-Related Proxies by Reported Expense Ratio Decile

	Decile 1		Decile 3		Decile 5		Decile 7		Decile 10		Decile 1-10
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Diff.
Fund average uniqueness	1.845	0.112	1.865	0.088	1.883	0.085	1.879	0.086	1.901	0.074	0.056***
Flow-perf sensitivity	-0.088	14.620	0.080	1.592	0.147	1.548	0.076	1.693	0.189	1.910	0.277***
Fraction of Easy-in-hard-out funds	0.41%		2.12%		3.65%		6.50%		7.65%		7.24%

Panel B. Separately Estimated Coefficients in Expense Regressions

	Full Sample				Large Funds			
Controlling for…	Coeff.	t-stat	# of Obs.	Avg. R^2	Coeff.	t-stat	# of Obs.	Avg. R^2
Fund average uniqueness	-2.32	-1.85	91813	28.75%	7.08	6.62	19670	40.12%
Easy-In-Hard-Out	20.71	20.17	36713	26.46%	22.21	4.26	4669	29.30%
Flow-perf sensitivity	7.19	2.44	104548	27.99%	12.42	1.00	22015	34.21%

Panel C. Average Residual Fee Spreads in Basis Points

	Full Sample			Largest Funds		
	25th to 75th	10th to 90th	1st to 99th	25th to 75th	10th to 90th	1st to 99th
Base Case	62	124	247	42	93	177
Fund average uniqueness	54	113	260	33	68	144
Flow-Perf Sensitivity	63	123	232	44	95	180
Easy-In-Hard-Out	53	100	210	46	89	187

Table 7. Residual Expense Spreads in Different Distribution Channels.

For reported and residual fees (using the base-case model, as described in Table 2, but estimated separately for each distribution channel group) this table presents the time-series averages of the cross-sectional spreads, in basis points (bps), for institutional funds, all retail funds, directly-sold retail funds and broker-sold retail funds. Our sample consists of domestic equity mutual funds (see Table A in the Data Appendix for a detailed description of the sample). The data in this analysis covers the period of 1999 to 2014 and is a yearly panel. The variables are defined in Table B of the Data Appendix.

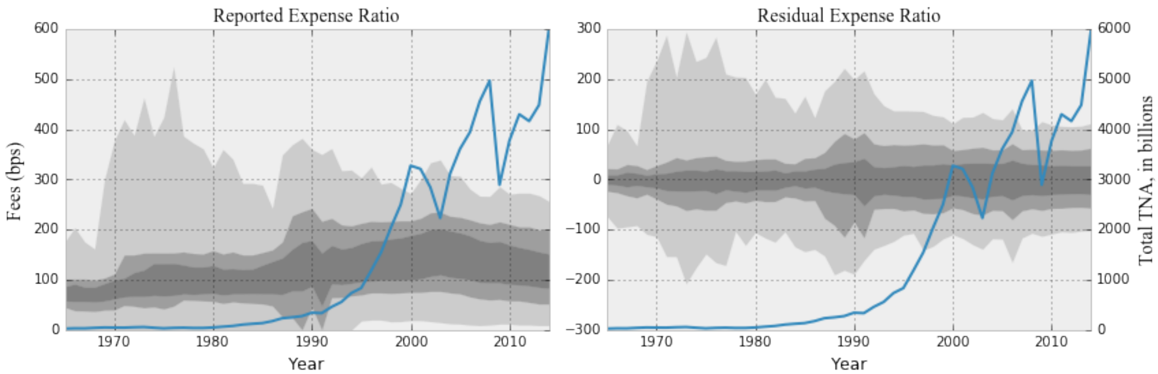
Sample	**Mean Reported Fee Spread (bps)**		
	25^{th} to 75^{th} Percentile	*10^{th} to 90^{th} Percentile*	*1^{th} to 99^{th} Percentile*
Institutional Funds	50	108	190
Retail Funds	83	142	280
Retail Funds: Directly-Sold	41	106	235
Retail Funds: Broker-Sold	38	88	212

Sample	**Mean Residual Fee Spread (bps)**		
	25^{th} to 75^{th} Percentile	*10^{th} to 90^{th} Percentile*	*1^{th} to 99^{th} Percentile*
Institutional Funds	44	85	170
Retail Funds	71	133	247
Retail Funds: Directly-Sold	40	83	178
Retail Funds: Broker-Sold	34	70	170

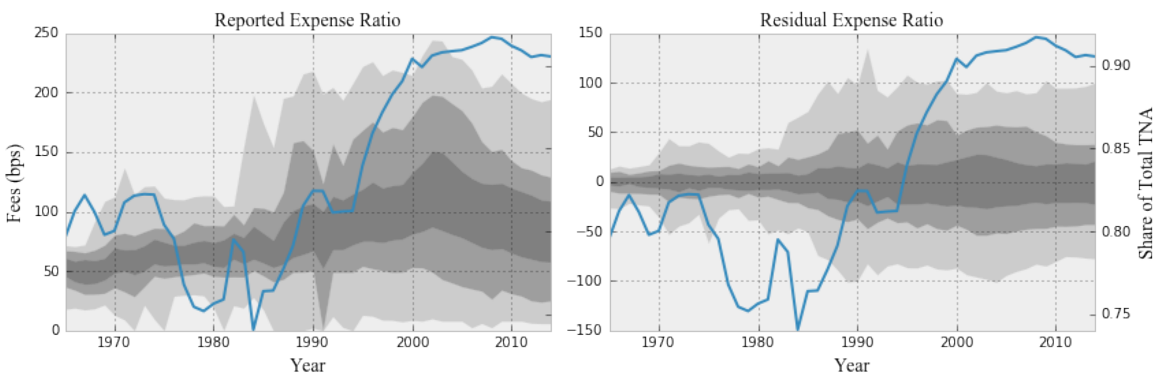
Figure 1. Fund Expense Dispersion

These figures show the dispersion of reported expense ratios (left column) and residual expense ratios (right column) across funds and over time. The graphs show the ranges between the 25th and 75th (darkest grey), 10th and 90th (medium dark grey) and 1st and 99th percentile (light grey) points of the distributions. Graphs in the top row, for the full sample, also plot the aggregate TNA of all funds in the sample in Billions of USD (blue line). In rows 2 (top TNA quintile funds) and 3 (S&P 500 index funds), the solid line represents the fraction of aggregate TNA represented by funds in those respective sub-samples. The residual expenses are defined as the regression residuals of the expense models specified in Table 2. Our sample consists of domestic equity mutual funds (see Table A in the Data Appendix for a detailed description of the sample). The data covers the period of 1966 to 2014 and is a yearly panel. The left column y-axis is annual reported fees in basis points (bps) and the right column y-axis is annual residual fees in basis points (bps).

All Funds



Large Funds



S&P 500 Index Funds

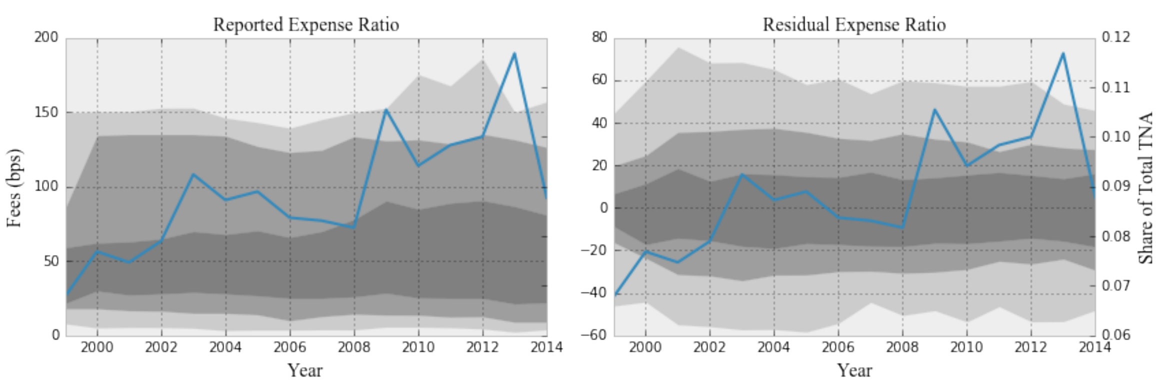
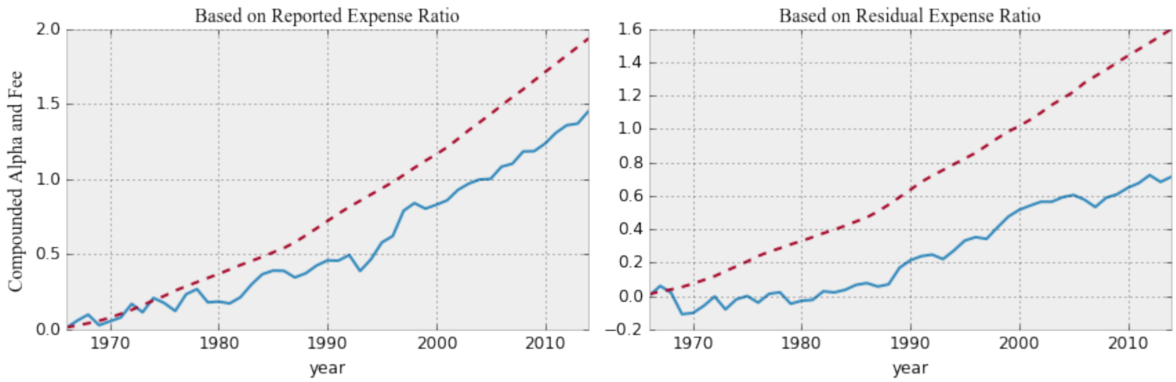


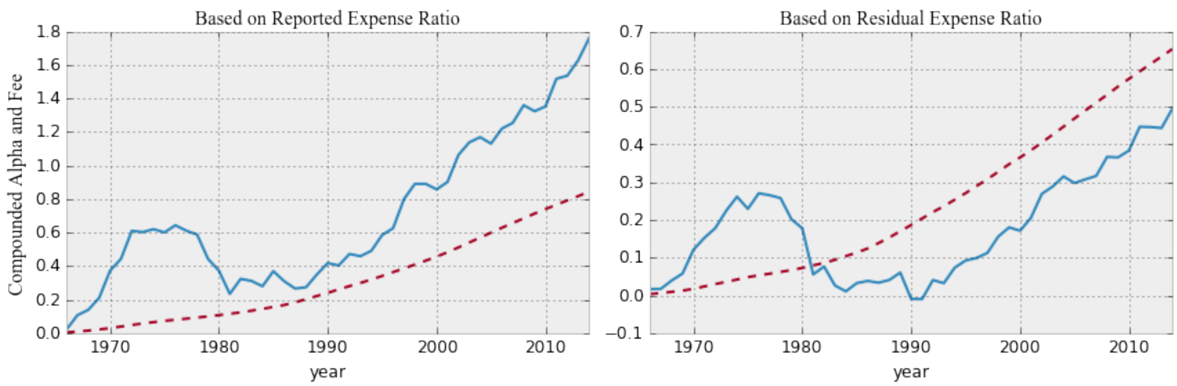
Figure 2. Evaluation of Trading Strategy

These figures show the cumulative Carhart alpha (the solid line) of a strategy that buys funds in the bottom decile of reported expense ratios (residual expense ratios) and shorts funds in the top decile of reported expense ratios (residual expense ratios). The figures also report the cumulative spread between average reported expense ratios (residual expense ratios) of funds in the top and the bottom decile (the dashed line). The residual expenses are defined as the regression residuals of the expense models specified in Table 2. Our sample consists of domestic equity mutual funds (see Table A in the Data Appendix for a detailed description of the sample). The data covers the period of 1966 to 2014 and is a yearly panel.

All Funds



Large Funds



S&P 500 Index Funds

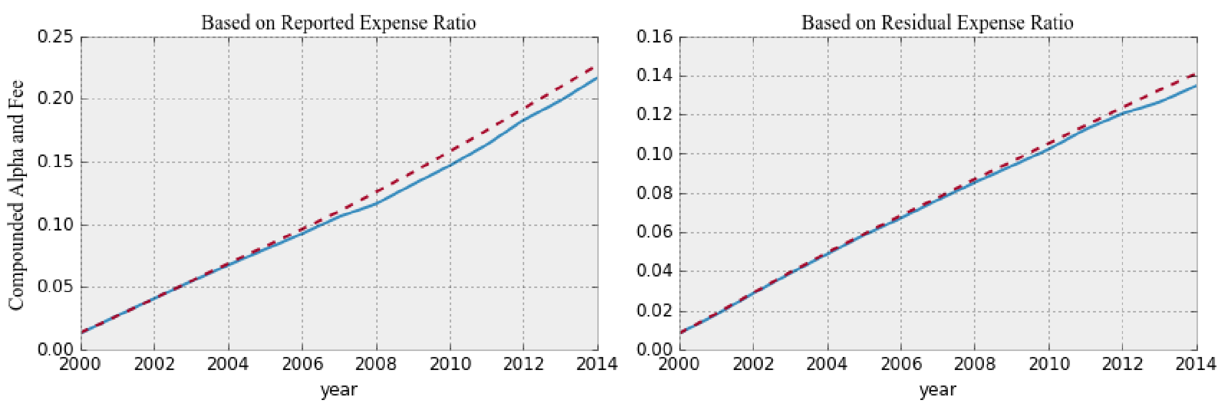
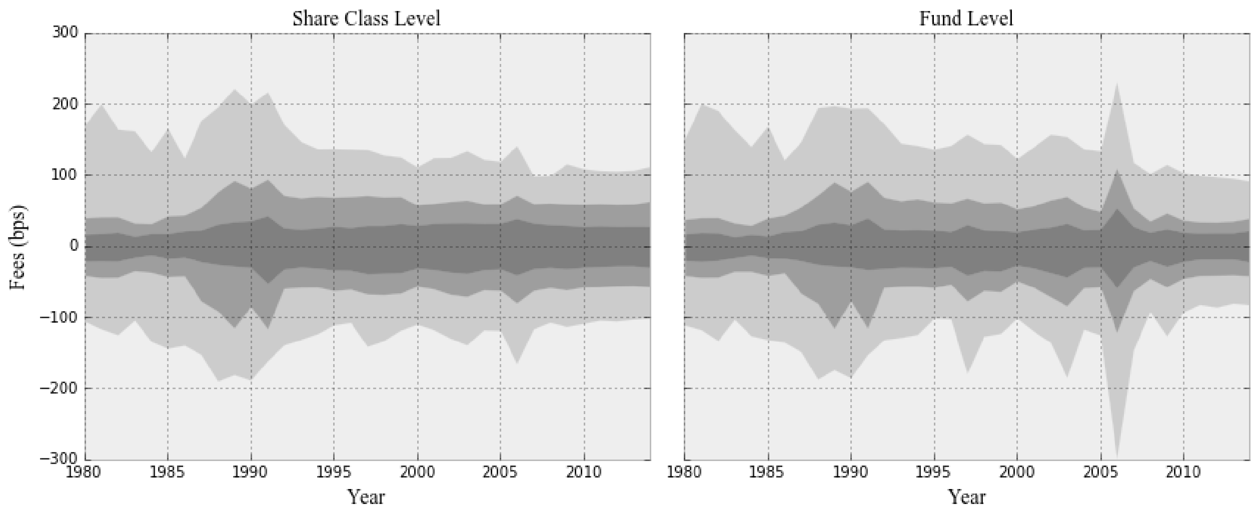


Figure 3. Fund Expense Dispersion at the Share Class and Fund Level

These figure show residual expense ratios across funds and over time for aggregated and non-aggregated share classes. In the 1st (2nd) column, the sample consists of observations at the share class (fund; if a fund has multiple share classes we select the largest share class) level. The graphs reported in the 1st row are based on all observations, while the graphs in the 2nd row only consider the top-20% share classes/funds in terms of fund TNA at the beginning of each year. The graphs show the ranges between the 25th and 75th (darkest grey), 10th and 90th (medium dark grey) and 1st and 99th percentile (light grey) points of the distributions. The residual expense is defined as the regression residual of the expense models specified in Table 2. Our sample consists of domestic equity mutual funds (see Table A in the Data Appendix for a detailed description of the sample). The data covers the period of 1966 to 2014 and is a yearly panel. The y-axes in both columns are the same scale and are the annual residual fees in basis points (bps).

All Funds



Large Funds

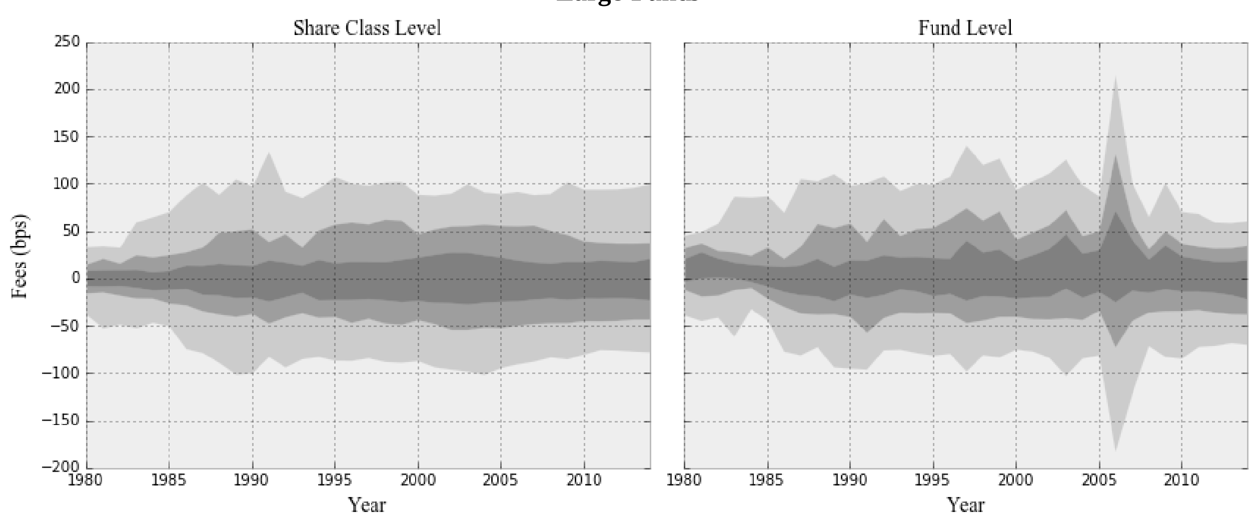
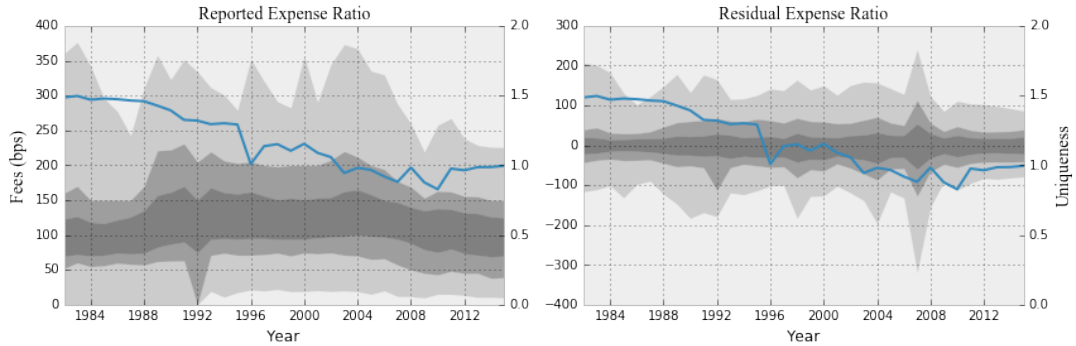


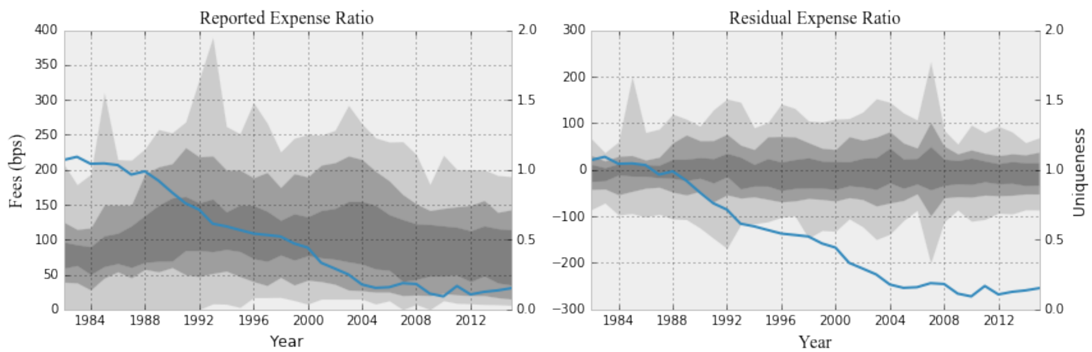
Figure 4. Fund Expense Dispersion for Holdings-matched Fund Pairs

These figures show the interquartile spread (dark grey), 10th to 90th percentile spread (medium grey) and 1st to 99th percentile spread (light grey) for the absolute differences in reported expense ratios (1st column) and residual expenses (2nd column) for holdings-matched fund pairs. The matching is based on holdings' overlap measured in terms of "uniqueness." For each possible fund pair, we estimate, yearly, the sum, across all holdings, of absolute differences in weights. For each fund, its uniqueness value is defined as the value of this sum for the fund pair that results in the largest holdings overlap. The uniqueness measure is bounded between zero (perfect overlap) and two (no overlap). The first row shows reported expense ratio spreads for the full sample of pairs, the second row shows the most similar pairs (i.e., quintile 1 of uniqueness) and the third row shows least similar pairs (i.e., quintile 5 of uniqueness). The solid lines show the samples' mean uniqueness measures each year. The residual expense ratio is defined as the regression residual of the expense models specified in Table 2. Our sample consists of domestic equity mutual funds (see Table A in the Data Appendix for a detailed description of the sample). The data covers the period of 1980 to 2014 and is a yearly panel. The left column y-axis is annual reported fees in basis points (bps) and the right column y-axis is annual residual fees in basis points (bps).

Full Pairs Sample



Most Similar Pairs



Least Similar Pairs

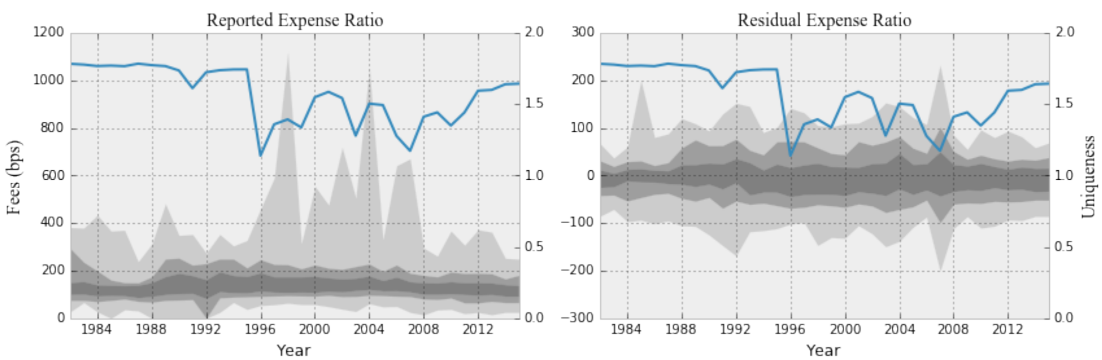
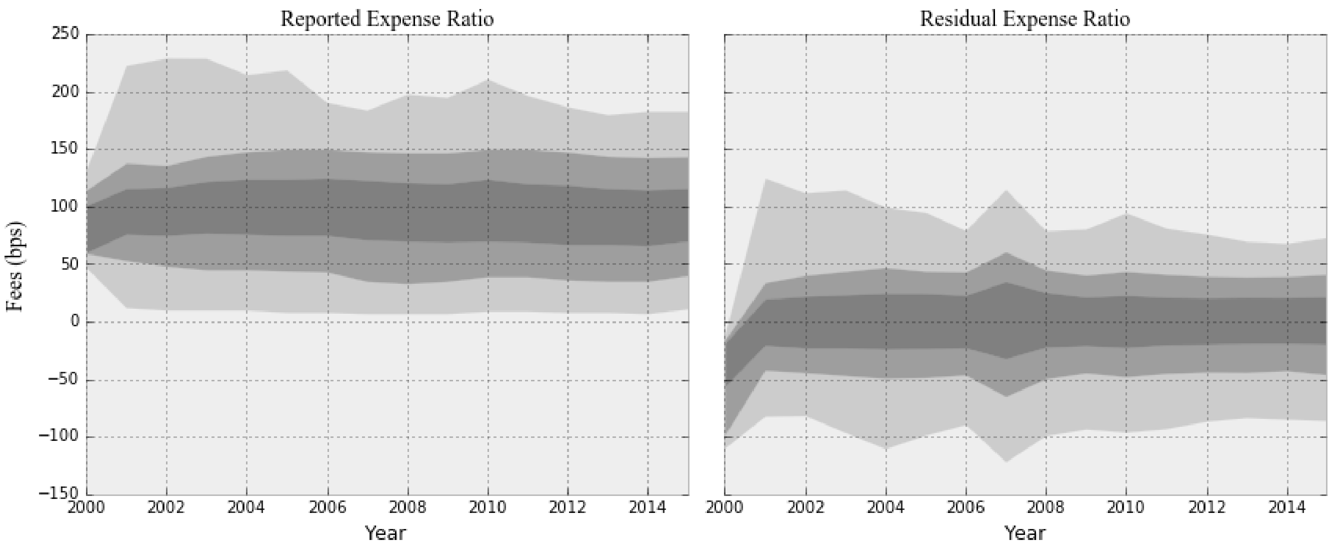


Figure 5. Fund Expense Dispersion of Institutional and Retail Funds

These figures show the dispersion of reported expense ratios (left column) and residual expense ratios (right column) across funds that are institutional funds (top row) and funds that are retail funds (bottom row). The graphs show the ranges between the 25th and 75th (darkest grey), 10th and 90th (medium dark grey) and 1st and 99th percentile (light grey) points of the distributions. The residual expenses are defined as the regression residuals of the expense models specified in Table 2 but separately estimated for each subset of funds. Our sample consists of domestic equity mutual funds (see Table A in the Data Appendix for a detailed description of the sample). The data in this analysis covers the period of 1999 to 2014 and is a yearly panel. The left column y-axis is annual reported fees in basis points (bps) and the right column y-axis is annual residual fees in basis points (bps). The y-axes in both columns are the same scale.

Institutional Funds



Retail Funds

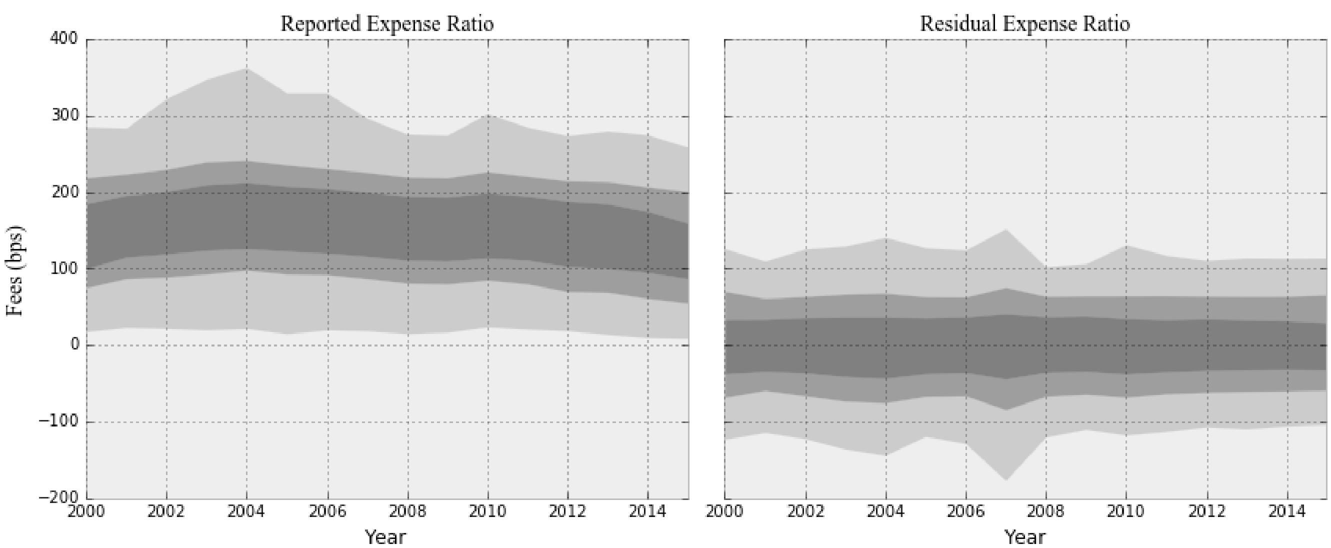


Figure 6. Fund Expense Dispersion of Directly-Sold and Broker-Sold Retail Funds

These figures show the dispersion of reported expense ratios (left column) and residual expense ratios (right column) across directly-sold retail funds (top row) and broker-sold retail funds (bottom row). The graphs show the ranges between the 25th and 75th (darkest grey), 10th and 90th (medium dark grey) and 1st and 99th percentile (light grey) points of the distributions. The residual expenses are defined as the regression residuals of the expense models specified in Table 2 but separately estimated for each subset of funds. Our sample consists of domestic equity mutual funds (see Table A in the Data Appendix for a detailed description of the sample). The data in this analysis covers the period of 1999 to 2014 and is a yearly panel. The left column y-axis is annual reported fees in basis points (bps) and the right column y-axis is annual residual fees in basis points (bps). The y-axes in both columns are the same scale.