Text-Based Mutual Fund Peer Groups*

Simona Abis and Anton Lines

Columbia University

December 31, 2020

PRELIMINARY DRAFT

Abstract

In this paper we ask whether active equity mutual funds differentiate their product offering to match preferences of heterogeneous investors. We then study the equilibrium allocation between the supply of differentiated funds and the demand by different investor types. We use unsupervised machine learning to categorize US active equity mutual funds into Strategy Peer Groups (SPGs) based on their strategy descriptions in prospectuses. We find rich variety in funds' self-described strategies that cannot be fully accounted for by differences in risk-adjusted returns. SPGs, though, display significant and interpretable differences in characteristics of stocks held. Funds in different SPGs have a different likelihood of targeting retail, institutional or retirement investors who, in turn, self-allocate differently across SPGs. Likely indicating differential preferences over investment characteristics.

Keywords: Institutional Investors, Mutual Funds, Assets Demand, Machine Learning, Regulatory Disclosures.

*Correspondence: Simona Abis, e-mail: simona.abis@gsb.columbia.edu; Anton Lines, e-mail: anton.lines@gsb.columbia.edu

1 Introduction

> Mutual fund prospectuses provide you with important information so you understand how the fund works and can easily compare it with other funds. If you wish to make an informed investment decision, you should read the prospectus before buying or selling shares in a mutual fund. SEC (2016)

The Securities and Exchange Commission (SEC) requires all mutual funds in the United States to regularly publish prospectuses in which, among other pieces of information, they must provide descriptions of their investment strategies. These documents provide a unique view on how funds present their value proposition.

In this study we use these self-disclosed strategy descriptions as a proxy for what funds promise their current and prospective clients. We use unsupervised machine learning to group funds into strategy peer groups (SPGs) based on similarities in the text, which correspond to distinct investment styles or philosophies. Having categorized funds in this manner, we then ask whether managers behave in a way that is consistent with their stated goals and methodologies, and how this relates to fund flows across different investor types. This approach allows us to better understand the drivers of capital allocation within the mutual funds industry and consequently of mutual funds' demand for financial assets.

Understanding the demand for financial assets of mutual funds is of particular relevance as at the end of 2019 this industry managed roughly 24% of the total US market capitalization.

In a perfectly efficient market, understanding the drivers of demand of single investors or even of a group of investors is generally not crucial. Indeed investors are assumed to act atomistically, following rational expectations about future cash-flows. Under those assumptions, equilibrium prices are perfectly efficient. Even in the presence of large investors who trade for non-fundamental reasons (e.g. for liquidity or behavioral reasons) prices remain efficient as long as they are sufficiently elastic.

An increasing body of evidence, though, shows that some of those assumptions might be violated. In particular Koijen and Yogo (2019) show that prices are not perfectly elastic, as a large portion of institutional investors are subject to mandates which limit their ability to absorb non-fundamental shocks. Additionally, investors might hold heterogeneous beliefs about future cash-flows and/or have non-fundamental preferences. These different beliefs and preference are correlated across investors types generating price pressure. They propose a model of asset prices which allows incorporating these features of demand.

Approaching asset pricing from a demand perspective has the clear benefit of allowing to study the role played by different institutions in the determination of asset prices, with relevant regulatory implications. The main criticism, though, is that demand is imputed from commonality in holdings so the drivers of demand are not well understood. Indeed, even though a large portion of variation in asset prices can be explained with demand preferences over a small set of characteristics, a sizable residual remains unexplained.[1]

Systematically analyzing funds' self-disclosed strategy descriptions allows us to better characterize mutual funds' demand for financial assets. Indeed, it allows us to better understand what are the priced and non-priced asset characteristics that different investors care about and the consequent clientele effects. We explore the drivers of both fundamental and non-fundamental demand in the US active equity mutual fund industry from 2000 to 2017. We decided to restrict the scope of our analysis to active funds for two reasons. First, active funds have more discretion in determining their investment strategies, hence they are better able to differentiate their products to target different clienteles. In this context, how they decide to communicate their strategy to investors becomes particularly relevant, hence we believe that prospectuses' content might be more valuable. Second, Koijen et al. (2020) show that, together with hedge funds, active investment advisors are the group of investors with the highest price elasticity. This, coupled with their greater freedom to deviate from market weights, makes their impact on prices larger.

Our first step in understanding the drivers of active equity mutual funds' demand is to systematically explore the content of their strategy descriptions in prospectuses: the "Principal Investment Strategies" (PIS) section. This is the first paper to systematically analyze the full content of the PIS section, for that reason we start by providing some descriptive statistics. We show that there exists a large variation in word counts, textual complexity, sentiment scores and content of these descriptions in the panel of fund-date observations. Word counts range from approximately 50 to $1,600$. Complexity, measured in approximate years of schooling required to understand the text, varies from 5 to 25 years. Positive, negative, and finance-specific sentiment scores also show high cross-sectional variation, refuting the commonly-held belief that these descriptions are mainly boilerplate.

[1]For instance Koijen et al. (2020) are able to explain 52% of the variation in valuation ratios in the US using preferences over a commonly used set of priced characteristics.

Next, we assess differences/similarities in the full content of these descriptions using unsupervised machine learning to categorize them into linguistically distinct groups. We show that the universe of strategy descriptions can be categorize into 15 groups (henceforth *Strategy Peer Groups* – SPGs). To arrive at these SPGs, we use a simple and intuitive machine learning algorithm: *k-means*. The algorithm is unsupervised, meaning it does not rely on any pre-defined categories and simply searches for patterns in the data. The only hyperparameter required is the desired number of clusters. There is no agreed upon solution on an ex-ante methodology to chose the correct number. Hence we develop two criteria, which we label *density* and *stability*, that allow us to determine the optimal number of clusters in our setting.[2]

Clustering thus provides a way of categorizing funds into peer groups based on the similarity of their strategy descriptions. Each SPG is characterized by a distribution over words and bi-grams (*features*). These can be represented using word clouds, in which the size of each feature indicates its frequency. This visualization aids interpretability and allows us to inspect and label each SPG.[3] Some groups, such as the "Small Cap" and "Mid Cap" SPGs, appear to correspond to standard factor exposures, but many of the groups go beyond these factors. Some are associated with specific asset classes (e.g. Fixed Income; Derivatives),[4] some with stock characteristics (Dividend; Products & Services), some with investing mechanics (Quantitative; Defensive), and some with international markets (Foreign (ADR), Foreign (EM)). The first six most popular strategies among funds in our sample are "Undervalued" (44,849 fund-month observations), "Sector" (28,953), "Dividends" (22,174), "Derivatives" (21,787), and "CompetitiveAdvantage" (21,402), "Products & Services" (20,855).

Despite the heterogeneity in narrative descriptions that we document, this might not translate into significant differences in funds' actions. Hence, the second step of our analysis is to relate textual similarity to similarities in funds' asset allocation choices and performance.

[2]The density criterion is satisfied if, when increasing the number of clusters, each new cluster contains a sufficiently large number of funds while being linguistically distinct from each other new cluster. The stability criterion is satisfied if when decreasing/increasing the number of clusters, most observations are jointly classified into the same group across specifications.

[3]None of our empirical results depend on the specific labels we choose for the clusters.

[4]While all funds in our sample have over their lifetime on average at least 80% of holdings in common stock, this percentage might vary over time and what they do with the remaining holdings is often their most distinguishing feature.

We start by showing that fund-level characteristics vary across groups. For instance, funds in the “Sector”, “Dividends”, “Derivative”, “Quantitative”, “Defensive” and “Foreign (ADR)” SPGs charge significantly lower fees than the average fund, whereas funds in the “Competitive Advantage”, “Products & Services”, “Small Cap”, “Mid Cap” and “Intrinsic Value” SPGs charge significantly higher fees. We report significant differences also in the size, age and turnover ratio of funds belonging to different SPGs. Given these systematic differences, we control for demeaned fund-level characteristics in all our subsequent analyses.

Next, we show that funds belonging to the same SPG have significantly more similar raw returns and holdings.

The total amount of capital managed by funds in the different SPGs also greatly differs. Since strategies proliferation and their relative size must be an equilibrium outcome that reconciles the supply of funds by fund families with investors’ demand. To rationalize this product differentiation we ask whether funds in different SPGs provide a significantly different level of risk-adjusted performance. We don’t find support for this hypothesis. Indeed, none of the SPGs provides greater risk-adjusted performance with respect to the average fund.

These results lead us to hypothesize that the difference in assets allocated across different SPGs must be driven by investor preferences over either the unique combination of risk exposures that they provide or other “services” they offer that, even if not priced, provide different clienteles with additional utility.

We test this hypothesis in two steps. First, we show that funds in the same SPG have significant similarities in risk factor exposures and in other characteristics of stocks held. Next, we show that different SPGs are offered to and attract investors with different preferences (retail, institutional or retirement).

Expanding on the results of step one. We start by showing that funds in the same SPG have similar exposures to known asset pricing factors (market, size, value, momentum, investment and profitability). These exposures are interpretable in light of the information disclosed in prospectuses. For instance, as expected, funds in the “Small Cap” SPG have a significantly higher exposure to the size factor while funds in the “Dividends” SPG have a significantly lower exposure to it. More subtle exposures also emerge. For instance, funds in the “Defensive”, “Dividends”, “Derivatives” and “Fixed Income” SPGs have a significantly lower exposure to the market portfolio. While funds in the “Quantitative”, “PE-Ratio” and

"IntrinsicValue" SPGs have a significantly higher exposure to the value factor. These results further suggest that SPGs offer a different and complex exposure to the various factors. For instance, the "Dividends" SPG has a significantly higher exposure to the "Value", "Investment" and "Profitability" factors, while it has a significantly lower exposure to the "MarketBeta", "Size" and "Momentum" factors. This is consistent with funds in this group investing in larger, profitable firms which are conservative in their investments and are further along in their life cycle, as they are more likely to pay dividends. High dividend paying firms also often tend to have a lower market beta, like for instance utility companies.

These results indicate that a first level of differentiation across these funds is their exposure to asset pricing factors, what is commonly defined as styles. Other papers have shown that preferences over styles drive demand by mutual funds, for instance Ben-David et al. (2020) show that flows purely motivated by changes in Morningstar's methodology to compute ratings exert price pressure on underlying stock and style returns. Looking at SPGs adds to this analysis by showing that funds differentiate themselves beyond the classical 3×3 matrix provided by Morningstar (value/blend/growth, small/mid/large) but offer their clients more sophisticated combinations of these factors, likely to satisfy different investor preferences.

We additionally construct a fund-level measure of similarity in characteristics of stocks held with respect to other funds in the same SPG. Having identified SPGs through self-disclosed descriptions, gives us a clear lens through which to examine similarities in holdings characteristics across funds. To avoid relying on subjective interpretations, we assume that the average fund in each SPG exemplifies the SPG's core strategy. We calculate the weighted sum of squared differences between the stock characteristics of each fund's portfolio and those of the exemplar, calling this measure Characteristic Dispersion ($CDisp$). A higher $CDisp$ indicates that a fund holds less similar stocks to the other funds in its peer group. The selected characteristics fall into eight categories: firm assets, firm liabilities, income statement, security market, industry, information availability, sentiment and strategy characteristics. The categories each carry equal weight in the measure, and variables within each category are also equally weighted. We further split this measure into two parts, using priced characteristics associated with the Fama-French five factors plus momentum ($CDispP$) and other characteristics ($CDispNP$) and into the eight components, which we evaluate separately. In a rational market, investors should be indifferent between funds with different characteristics, provided those characteristics carry an appropriate risk premium. However, if non-priced characteristics interact with investors' personal preferences or constraints, a fund's deviation from its core strategy may impose a cost without the associated risk compensation, and they may no longer be indifferent. Finally, we compute a version of Characteristic Dispersion relative to funds outside of the current peer group ($\widetilde{CDispP}$ and $\widetilde{CDispNP}$).

We find that, on average, funds are more similar to other funds in the same peer group than they are to those outside of the group, along both priced and other characteristics of stocks held. This result confirms in a very general way that managers mostly do behave according to the specifications or limitations set forth in their prospectuses. We also explore individual strategy groups in a more descriptive manner as a complement to the general measure. For example, we find that "Dividend" funds invest in the highest dividend yield stocks among all peer groups, and concentrate their industry exposure in high-payout sectors such as Utilities, Telecoms, and Consumer Staples.

Second, to further explain differences in preferences, we split the Total Net Assets (TNA) managed by each fund into three groups containing all share classes targeted towards retail, institutional and retirement investors respectively. The categorization is obtained using the "retail" and "institutional" indicator variables in the CRSP Mutual Fund dataset, complemented with textual analysis of fund names in order to further distinguish the retirement share classes (usually categorized with institutional).

We report that there has been a shift in the total amount of retail, institutional and retirement capital in this market. Indeed, at the beginning of 2000 institutional and retirement capital accounted respectively for 12.58% and 0.15% of the overall TNA managed by these funds, while at the end of 2017 it amounted to 37.04% and 10.06% respectively (Figure 8).

Changes in investor type, and hence in their preferences, could partly drive differences in size across SPGs. In this paper, we do not try to disentangle whether these effects are driven by supply or demand shocks, i.e. whether funds are creating/marketing products to attract different investors types or if changes in investor demand prompted the creation of differentiated products. Instead, we are interested in the equilibrium outcome of the intersection of these demand and supply forces. We document that there exists a sorting of different investor types into strategies (SPGs) that display distinctive characteristics in their underlying asset selection.

In order to formally show this result we proceed in three steps. First, we show that investor of different types allocate a significantly greater (smaller) proportion of their net flows to funds in different SPGs. For instance, we show that, corresponding to a change in net flows, retail and institutional investors allocate proportionally less of their capital to the "Quantitative" SPG while retirement investors allocate proportionally more to it. Second, we show that the likelihood that funds belonging to different SPGs will offer a retail, institutional or retirement share class differs. For instance "Dividend" funds have a lower likelihood of offering institutional or retirement share classes, while "Tax" funds have a lower likelihood of offering a "Retirement" share class. Third, we show that, conditionally on a share class being offered, the percentage of total retail, institutional and retirement capital managed by funds in different SPGs differs. For instance, funds belonging to the "Dividend" SPG have a significantly greater proportion of capital in their retail share classes and a significantly lower proportion in the institutional ones relative to other funds; while the opposite is true for funds in the "Small Cap" SPG. In all results, to address potential omitted variables and endogeneity issues we control for a number of fund-level characteristics and we add month and fund family fixed effects. This essentially allows us to compare funds in the same month which belong to the same fund family but differ in their SPG assignment.

Overall, in this paper we show that there exist clusters of funds which offer a similar value proposition to investors. These funds choose to hold similar securities and attract similar investors. These correlated preferences at the investor-type level translate into similarities in how they allocate net flows across funds; which could have relevant asset pricing implications. Indeed, correlated inflows/outflows into funds which hold similar securities could introduce fragility in the stock market imposing non-fundamental price pressure to the underlying securities held by these funds. In future work we plan to study the impact of flow-induced-trading at the SPG level on the price of underlying securities.

Finally we add a number of robustness tests. First we construct pairwise similarities in the returns and holdings of funds and show that our text-based measure has incremental impact in explaining these similarities relative to other characteristic based measures (e.g. the Daniel et al. (1997) characteristic portfolios). Second, we repeat clustering based on two different sub-periods (before and after the Great Recession). We show that in recent years a few new clusters have emerged and these particularly target institutional investors. Primarily this is true for the "Long-Short" cluster.

The remainder of the paper is organized as follows: Section 2 highlights the contribution of the paper relative to the existing literature; Section 3 describes the data and the creation of strategy peer groups; Section 4 contains the empirical analysis; and Section 5 concludes. Details about the pre-processing of fund prospectuses and about the K-means algorithm can be found in the Appendix.

2 Related Literature

This paper contributes to the literature by providing a detailed analysis of the drivers of fundamental and non-fundamental demand in US active equity mutual funds. In that respect, it is most closely related to studies which shows clientele effect in asset pricing. Among others (Koijen and Yogo (2019), Koijen et al. (2020), Ben-David et al. (2020), Hong and Kacperczyk (2009))

A key contribution of the paper is being able to extract peer groups of active equity mutual funds from their textual disclosures to investors. Hence our work relates to studies which use textual analysis in a finance context and in particular those utilizing regulatory disclosures by mutual funds. In that respect, the two most closely related papers to ours are Kostovetsky and Warner (2020) and Abis (2020). Kostovetsky and Warner (2020) use short excerpts from prospectus strategy descriptions to examine distinctiveness of fund styles.[5] They find that small and start-up fund families have more distinctive text, and that the flow-performance relationship is weaker for more distinctive funds. In this paper rather than focusing on distinctiveness we highlight similarities across funds and their implications for flows from different investor types. We also offer a wealth of descriptive insight into the mutual fund strategy landscape, and specifically investigate whether funds' behavior lines up with their stated objectives. Abis (2020) utilizes the same full strategy descriptions from mutual fund prospectuses to identify Quantitative funds. She shows that differences in the learning abilities of Quantitative and Discretionary funds determine differences in their strategy and performance. This paper provides a more comprehensive analysis of the full

[5]Morningstar provides short excerpts of the full strategy text available in EDGAR filings. The average length of Principal Investment Strategy descriptions in the Morningstar dataset is 70 words, while the average length of this section in our dataset is 317 words.

content of mutual fund strategy descriptions, rather than focusing on a single dimension (quantitative vs. discretionary). Other papers that utilize mutual funds prospectuses are Hillert et al. (2016), Beggs et al. (2019).

Another closely related stream of research is the rich literature of funds benchamarking. The paper which is most closely related to ours is Hunter et al. (2014). They assign funds to peer groups ex-post, based on the best fit of fund returns to the returns of nine Russell indexes along size and value dimensions—similar to Morningstar's equity style boxes. They find that fund selection is improved when estimating alphas with a FFC factor model that includes the peer group index. By contrast, we use the text of fund prospectuses to construct peer groups ex-ante, then examine the behavior of funds according to which group they are part of.

3 Novel Mutual Fund Industry Mapping

3.1 Data

In this paper we study active equity US mutual funds from January 2000 to December 2017. To do so, we combine information about mutual fund characteristics, returns and holdings with a novel textual dataset of their mandatory disclosures to the SEC (prospectuses). We start the analysis in the year 2000, which is when reliable prospectuses data starts being available.

Characteristics and Return: We obtain fund characteristics and returns from the CRSP Survivorship-Bias-Free Mutual Fund dataset. We restricts the sample to equity funds and further exclude international funds, sector funds, index funds, and underlying variable annuities. We account for incubation bias by excluding observations dated after the fund's first offer date (Elton et al. (2001)). We also exclude funds with less than $5 million in Total Net Assets – TNA (Kacperczyk et al. (2008)). We further aggregate all information relative to all share classes of the same funds for each time period. We do so by: keeping the first offer date of the oldest share class, summing the TNA of all share classes and weighting all other variables (e.g. fees, returns, turnover, ...) by lagged TNA. Following Abis (2020), we identify share classes of the same fund by constructing a comprehensive fund identifier using the CRSP Class Group identifier, the WFICN identifier in the MFLinks linking table, and fund names. This choice is particularly relevant for matching return and characteristic to funds' holdings. In fact, the MFLinks linking table excludes many new funds in recent years (Zhu (2020), Shive and Yun (2013)). We finally exclude funds for which we have less that 12 months of observations.

Holdings: Holdings are obtained by merging the Thomson CDA/Spectrum Mutual Fund Holdings dataset from January 2000 to August 2008 and the CRSP Mutual Fund Holdings dataset from September 2008 to December 2017. The date of switch is chosen to maximize coverage for the funds of interest. We finally remove other small funds that either hold fewer than 10 stocks or that on average dedicate less than 80% of their assets (excluding cash) to holding common stocks (Kacperczyk et al. (2008)). We then forward fill holdings to the monthly frequency.[6]

Prospectuses: We obtain fund prospectuses through the Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) of the SEC. The EDGAR system has been active since 1994, subsequently experiencing a 2 years phase-in period. Progressive reforms on mandated disclosure then led to greater standardization in disclosures formatting. In particular in 2000, with Release $No.33-7684$, the SEC started accepting disclosures in HTML format (SEC (2006)). Due to the smaller coverage and the lack of standardization in earlier years, we are able to obtain reliable coverage for the funds of interest as of the year 2000, which is when our sample starts. We require that all observations in the final dataset be matched to a strategy description (more details in section 3.2).

Other: To construct some of the variables of interest we also use the CRSP stock-level monthly database, Compustat, Fama–French factors and industry portfolios and macro economic series from FRED.

Our final dataset consists of $2,995$ unique funds and $320,750$ fund–month observations. Table 1 provides descriptive statistics for the final dataset.

[6]For 42% of the final sample monthly holdings were already available. 90% of the data is forward filled for at most 1 quarter and 99% is forward filled for at most 2 quarters. Maximum forward filling is restricted to 1 year.

3.2 Principal Investment Strategies

Collection The SEC requires all mutual funds families to publish quarterly prospectuses about all of their funds. These prospectuses are divided into sections, each addressing a different regulatory question. In this paper we will focus on a specific section: Principal Investment Strategies (PIS). This corresponds to Item 9(c) of the N-1A mandatory disclosure form.[7] This item mandates that funds disclose their principal investments strategies including the types of securities they tend to hold and the main criteria utilized in selecting those securities. Funds provide narrative descriptions of their strategy which are only constrained by the above requirements and to be written in "plain English" under rule 421(d) of the Securities Act.[8]

In this paper we utilize a comprehensive panel dataset of strategy narrative descriptions by fund–month, which we merge to the traditional mutual fund dataset. This dataset is similar to the one utilized by Abis (2020). She uses these sections to look for one specific piece of information: whether funds follow a quantitative approach. This paper is the first to provide a comprehensive analysis of the full content disclosed by funds about their strategies.

We are able to match $31,695$ prospectuses to our funds of interest. Prospectuses may be published in any day of the year and are often published less than once per quarter. Since any material change to the management of the fund must be reported to both the SEC and fund investors, for any month in which a prospectus is not available we fill that information forward using the latest available prospectus.

Description The regulatory requirement is somewhat vague, so one could imagine that funds might converge to providing short and uninformative strategy descriptions. Contrary to this hypothesis, Figure 1 shows that there exists significant cross-sectional dispersion in the length of Strategy sections across fund-month observations. We observe significant variation also in the Sentiment and Complexity of these sections.

Panel 1 of Figure 2 shows the distribution across fund–month observations of the Flesch-Kincaid grade level complexity measure (Kincaid et al. (1975)) for Strategy sections. Panel 2 displays the same distribution for the Flesch readability ease measure (Flesch (1948)). These measures are based on the relative number of total words to total sentences (average

[7] https://www.sec.gov/files/formn-1a.pdf

[8] https://www.sec.gov/rules/final/33-7497.txt

sentence length) and the relative number of total syllables to total words (average word length) present in a given text. They are calibrated to indicate respectively: the number of years of schooling required to comprehend the text (Panel 1) and a standardized readability index on a range of $[0, 100]$ (Panel 2). Figure 2 shows that there exists a large dispersion in the complexity (readability) of these sections, with most sections requiring between 5 to 20 years of training (achieving an ease of reading score between 10 and 80).

Finally, we use the Loughran and McDonald dictionaries of *positive*, *negative*, *uncertainty* and *litigious* words to measure sentiment. These dictionaries are adapted to account for specific characteristics of financial language (Loughran and McDonald (2011)). Figure 3 shows the distribution of the frequency of *positive* (Panel 1), *negative* (Panel 2), *uncertainty* (Panel 3) and *litigious* (Panel 4) words for all pooled fund–month strategy descriptions. Also here we observe large cross-sectional differences in the sentiment of strategy sections across fund–month observations.

These descriptive statistics suggest that the narrative descriptions of Strategy descriptions provided by funds through their prospectuses might contain relevant and heterogeneous information about funds strategies.

3.3 Strategy Peer Groups

Given the above descriptive findings, we utilize a machine-learning algorithm, applied to all extracted Strategy narrative descriptions, to group funds into peer groups based on similarities in their stated Strategies. This allows us to map the whole active equity mutual fund industry over time into interpretable "Strategy Peer Groups" (SPGs).

Pre-Processing In order to make the sections machine readable we *first* pre–process them using the "bag of words" approach. This procedure yields for each section a list of all stemmed words and bi-grams.[9]

The *second* step is to aggregate all unique words and bi-grams found in any of the Strategy sections into a unique corpus. We then identify boilerplate language by computing the frequency of all 4-grams (4-word combinations) present in the entire corpus. We remove

[9]We first remove symbols and stop-words (e.g. is, the and, etc.), we additionally remove a list of context specific stop-words (e.g. advisor, fund, ecc.). Next we stem each word to its root using the Porter stemmer algorithm (thus, e.g. "company", "companies", ... = "compani"). Finally we tokenize each section into words and bi-grams (i.e. all consecutive two words combinations).

the most frequent combinations at the corpus level (boilerplate) from each document. We further restrict the total list of words and bi-grams considered based on their within-corpus frequency.[10] This step reduces classification noise by removing all items that are less likely to be informative in differentiating the descriptions. The remaining words and bi-grams are the "features" utilized in the clustering algorithm.

The *third* step consists in representing each corpus with a "features matrix", whose columns are all the selected features and whose rows are all available strategy descriptions. The elements of each sparse matrix are either 0 or 1, where 1 indicates that a specific feature is present in a given section.

The *fourth* and final step is to substitutes the 1*s* in each sparse matrix with a weight representing the frequency in which that feature appears in the specific document relative to the frequency in which it appears in the whole corpus (namely, term frequency - inverse document frequency weighting – henceforth, features matrix = $tfidf$ matrix).

Clustering We cluster funds based on their stated Strategy description in order to generate "Strategy Peer Groups" (SPGs). We use the K-Means algorithm to achieve this goal. Results are similar using other clustering algorithms such as Gaussian Mixture. We chose K-Means for its simplicity, as in our setting this does not seem to trade-off with the quality of the categorization.

The K-Means algorithm takes as its only inputs the $tfidf$ matrix, the desired number of clusters and a tolerance parameter. The goal of the algorithm is to minimize the within-cluster euclidean distance between all elements assigned to that cluster and a cluster centroid. Euclidean distance is computed as follows:

$$\sum_{r=1}^{R} ||x_r - x_r^C||^2 \tag{1}$$

where x_r is the frequency assigned to feature r in a specific document and x_r^C is the frequency assigned to feature r in a cluster's centroid. R is the total number of features.

Centroids are initialized in an uninformed manner, then an iterative process is initiated

[10]We remove all words and bi-grams that appear in more than 30% of the sections and in less than 5%.

to optimize the centroids selection and the allocation of all documents to those centroids. In each iteration, each centroid is redefined as the average feature-frequency vector of all documents assigned to that centroid in the previous iteration, then all documents are reassigned to the closest centroid. Convergence is reached when the euclidean distance between cluster centroids in two consecutive iterations is smaller than the user-defined tolerance level (for a detailed description of the K-Means algorithm see Appendix A).

The key parameter to be specified is the number of clusters to be estimated. The optimal number varies depending on the true structure of the data. In order to choose this parameter we run the K-Means algorithm for different numbers of clusters, specifically all numbers between 10 and 20. Each run is independent.

We then compare the obtained categorizations in terms of two criteria:

- *Cluster stability.* For the approach to be robust it should not be very sensitive to the exact number of chosen clusters. In particular, the algorithm should categorize the majority of sections into the same group across different specifications. What we would expect to see from a robust approach is that, by decreasing the number of chosen clusters, funds belonging to more detailed categories should be homogeneously grouped into the same broader one.

- *Cluster density.* If exceeding the correct number of clusters, we would expect the algorithm to split homogeneous groups into smaller categories with very low density and a large overlap in identifying features. This would also make them difficult to distinguish from an interpretability perspective.

We choose the optimal number of clusters which satisfies both of these criteria.

To facilitate the discussion, we labeled all obtained clusters by inspecting their word clouds and reading a random sample of prospectuses from each identified group. Note that labeling is not needed in order to assess the above criteria or in any of the subsequent analysis.

Figure 4 visually displays the *stability* and *density* criteria for the identification of the correct number of clusters in constructing SPGs. More specifically, the heat maps display the joint frequency of cluster assignments when going from a smaller (rows) to a larger (columns) number of clusters. The color scale on the right hand side of the Figures maps colors to the number of underlying classified sections.

Analyzing the joint assignment from the perspective of the two outlined criteria, we can see that the *cluster stability* criterion is satisfied. Indeed both heat maps display a few high density combinations and many low to no density ones. This indicates that the majority of observations are jointly classified into the same category across different specifications. If looking at the assigned labels we observe that stability is preserved also from an interpretability perspective. Indeed most observations are assigned across specifications to clusters with an identical or semantically related label.

The *density* criterion is satisfied when going from [15] to [16] strategy clusters; while it is not satisfied when going from [16] to [17]. This indicates that in order to accurately map the data more than [16] SPGs are required.

Panel 1 of Figure 4 shows that the *density* criterion is satisfied when going from [15] to [16] SPGs. In fact we observe that the "Sector" cluster emerges by reassigning observations from the "Undervalued", "Competitive Advantage" and "Quantitative" clusters. This cluster has a high density and it is characterized by a distinct wordcloud.

Conversely, we can see that the *density* criterion is **not** satisfied when going from [16] to [17] SPGs, by analyzing Panel 2 of Figure 4. Here we observe that the key addition is a second "Mid Cap" cluster with very similar identifying features as the previously existing one.

In our main specification we utilize [16] strategy clusters, as this satisfies both selection criteria and provides the most interpretability. Appendix A.2 details the exact methodoplogy implemented, translating these two criteria into quantitative measures.

Clusters Interpretation The SPGs identified through this methodology should be interpreted as groups of peer funds which place a similar emphasis on specific aspects of their Strategy. The assumption here being that the time spent speaking about different aspect of strategy is proportional to their relevance for the fund. The labels and the key prominent features are only a short-hand to indicate those characteristics that are most distinctive for any given peer group. Clusters, though, are identified by a full distribution over words.

Strategy Peer Groups With this methodology we are able to construct a panel dataset which, for all $2,995$ funds of interest, provides us with a cluster assignment into a Strategy

Peer Group (SPG), for every month that they are alive between January 2000 and December 2017. Figure 6 displays the number of funds assigned to each SPG each month and their cumulative TNA.

Figure 7 shows the frequency of assignment of all fund-month observations into the [15] identified Strategy Peer Groups. The "Undervalued" peer group is the largest by number of fund-month observations, it also counts 971 funds being assigned to it at some point in their lives. The second largest peer group by number of observations is the "Sector" (678 funds); followed by the "Dividends" (417 funds) and the "Competitive Advantage" (436 funds) peer groups. Figure 7 additionally shows that Strategy sections belonging to unique funds tend to be assigned to the same SPG over time. In fact we observe that $1,057$ funds are assigned to only one SPG throughout their lives. The vast majority of funds are assigned to a maximum of 5 SPGs throughout their lives. We observe a very small number of funds that are assigned to more than 5 SPGs, this might be attributed to estimation noise. Note that no estimation constraint imposes that sections belonging to the same fund be classified into the same SPG, these are treated independently. Hence the stability of these assignment over time further points to the robustness of the methodology.

Finally Table 2 shows differences in size, age, expense and turnover ratios across all estimated SPGs. Each Panel of the table (rows) represents estimates from separate regressions in which the cluster assignment is a dummy equal to 1 if the observation is assigned to the cluster of interest, 0 otherwise. All regressions include fund-family and month fixed effects and clustering at the fund and month level. They additionally include controls for the following fund characteristics: the natural logarithm of age and total net assets, the wisorized expense ratio and turnover ratio, monthly growth in net fund flows, and monthly fund flow volatility. Each model excludes from the controls the dependent variable. All controls are de-meaned, hence the estimated coefficient should be interpreted as the increment/decrement in average value of the dependent variable for the group of interest with respect to the average across all funds. The average effect across all funds is displayed in the last row of table 2.

As described in this section, this text-based clustering provides rich and interpretable strategy peer groups assignment for all funds of interest. The remainder of this paper will explore whether this textual similarity also translate into significant similarity in fund characteristics and in the characteristics of the implemented strategies and on whether investors with different preferences allocate differently across these strategies.

4 Empirical Analysis

4.1 Hypotheses Development

Given the increasing evidence of clientele effects in asset pricing, we hypothesize that managers' narrative descriptions of their funds' strategies can shed light on funds' mandates and the clientele they attract.

We begin by testing whether similarities in textual descriptions of funds' strategies, correspond to similarities in funds' actions. Funds that misrepresent themselves in prospectuses are liable to be sued by the SEC. Despite that, one might imagine that funds could use prospectuses as marketing material. In that case, the rich variation in textual descriptions we document might not translate into significant differences in funds' actions. This leads us to formulate our first hypothesis, that similarities in text do not translate into similarities in actions.

Hypothesis 1. *Funds use prospectuses as marketing material: similarities in the "Principal Investment Strategies" section do not translate into significant similarities in funds' actions.*

To test this hypothesis, we utilize the developed strategy peer groups. This allows us to identify groups of funds with similar strategy descriptions. We then ask whether funds that belong to the same SPG have more similar returns and/or hold more similar stocks. Section 4.2 further elaborates on the empirical strategy and results. We reject hypothesis 1, as we observe that belonging to the same SPG translates into significantly more similar fund characteristics, raw returns and holdings.

As displayed in Figure 6, SPGs also differ in size. This must be an equilibrium outcome in which the demand for each SPG meets its supply and the relative size of the SPGs is determined. Given this consideration, we ask whether these differences in size are driven by significant differences in risk-adjusted performance across SPGs.

Hypothesis 2. *There exist significant differences in risk-adjusted performance across SPGs.*

We test this hypothesis by checking if different SPGs display significant differences in 6-factor alphas or value added. Section 4.3 elaborates on the empirical strategy and results. We find small differences in risk-adjusted performance but not enough to justify the documented proliferation in investment strategies. Hence, we reject this hypothesis.

Given the above mentioned general equilibrium considerations, we hypothesize that these SPGs must deliver different "services" that might appeal to investors with different preferences.

Hypothesis 3. *SPGs provide different "services" which might appeal to investors with different preferences.*

We test this hypothesis in two stages. First, we check whether each SPG is characterized by similarities in risk exposures and/or in the characteristics of stocks held. We look at both characteristics that are known to be priced (size, book-to-market, momentum, investments and profitability) as well as other characteristics (e.g. industry allocation, accounting metrics and dividend yield). We build a generic measure of characteristic dispersion and we also conduct a more descriptive analysis, which compares the compatibility of strategy descriptions to the characteristics of stocks held. Section 4.4 further elaborates on the empirical strategy and results. We find support for this hypothesis. Indeed, we find that the stocks held by funds inside each SPG have more similar risk exposures and characteristics, we also find support for this hypothesis from an interpretability perspective.

Next, we check if investors with similar preferences are attracted to similar SPGs. We proxy for investor preferences by looking at three different investor types: retail, retirement and institutional. We then test whether investors of different types allocate a different proportion of their net flows to each SPG. We additionally look at whether SPG assignments can explain: (1) the likelihood of launching different share classes at the fund level; (2) the percentage of a fund's TNA coming from retail, retirement or institutional investors, conditionally on that share class being offered. Section 4.6 further develops on the empirical strategy and results. We find evidence in support of the hypothesis that investors with different preferences allocate to funds in different SPGs and funds in different SPGs target different investor types.

4.2 Test of Hypothesis 1: Fund-Characteristics, Holdings and Return

In assessing whether similarities in the text correspond to similarities in the underlying funds we first show some descriptive statistics which compare fund-level characteristics across SPGs. We do so by running the following regression for each fund characteristic and SPG:

$$Fund\_Characteristic_{jt} = \alpha + \beta I_{jt}^{SPG_{jt}} + \gamma X_{jt} + \eta_t + \iota_f + \epsilon_{it} \tag{2}$$

where $Fund\_Characteristicjt$ is the fund-level characteristic of interest for fund j at time t; $I_{jt}^{SPG_{jt}}$ is an indicator variable equal to 1 if fund j belongs to the SPG of interest at time t, 0 otherwise. X_{jt} are demeaned fund-level control variables namely: the natural logarithm of fund age and TNA and the winsorized expense ratio and turnover ratio, the growth in net fund flows and net fund flow volatility (excluding the dependent variable from each regression). We also include month and fund-family fixed-effects. Standard errors are clustered at the fund and month level.[11] Note that we run separate regressions for each SPG and fund characteristic of interest (using a different $I_{jt}^{SPG_{jt}}$).

Thanks to the demeaning of all control variables, α can be interpreted as the mean of the dependent variable when $I_{jt}^{SPG_{jt}} == 0$ (i.e. when fund j does not belong to the SPG of interest at time t). A significant β coefficient then indicates that the mean of the dependent variable is significantly different from α when $I_{jt}^{SPG_{jt}} == 1$ (i.e. when fund j belongs to the SPG of interest at time t). More specifically, for $I_{jt}^{SPG_{jt}} == 1$ the mean of the dependent variable is given by $(\alpha + \beta)$. Hence, β represents an incremental effect. Thanks to the inclusion of month and fund family fixed effects, a significant β indicates the incremental average value of the dependent variables for funds that in the same month belong to the same fund family but differ in their SPG assignment.

Table 2 displays the estimated β coefficient for each regression (one per characteristic and SPG). Estimates shows that the differences in these funds are not restricted to their textual similarity, but they also translate into significant differences in fund characteristics. Taking the "Quantitative" SPG as an example, we observe that funds in this group are significantly younger, smaller, have higher turnover ratio and charge a lower expense ratio. All differences are significant at the 1% level. This confirms the findings of Abis (2020), who

[11] The same control variables and regression specification are used in all subsequent regressions, unless specified otherwise.

identifies Quantitative funds using a supervised machine learning methodology. Beyond the "Quantitative" SPG, with respect to size, funds in the "Dividends" SPG (standard error: 1.85) are on average greater, while funds in the "Foreign (ADR)" SPG (s.e.: −1.77) are smaller. With respect to fund age, funds in the "Undervalued" (s.e.: 2.71), "Products & Services" (s.e.: 2.07), "PE-Ratio" (s.e.: 2.02) and "Foreign (EM)" (s.e.: 1.68) SPGs are older, while funds in the "Small Cap" SPG (s.e.: −3.48) are younger. With respect to expense ratios, funds in the "Sector" (s.e.: −1.76), "Dividends" (s.e.: −3.68), "Derivatives" (s.e.: −2.75), "Defensive" (s.e.: −1.74) and "Foreign (ADR)" (s.e.: −2.65) SPGs charge significantly lower exepense ratios; while funds in the "Competitive Advantage" (s.e.: 2.65), "Products & Services" (s.e.: 4.21), "Small Cap" (s.e.: 4.17), "Mid Cap" (s.e.: 2.14) and "Intrinsic Value" (s.e.: 2.81) SPGs charge significantly higher expense ratios. Finally, funds in the "Dividends" (s.e.: −5.88), "Defensive" (s.e.: −2.38), "Intrinsic Value" (s.e.: −6.15) and "Tax" (s.e.: −2.01) SPGs have a significantly lower turnover ratio; while funds in the "Sector" (s.e.: 2.48) and "Derivatives" (s.e.: −4.42) SPGs have a significantly higher turnover ratio.

Next, to test whether similarities in strategy descriptions also correspond to similarities in funds' actions we use two measures: return differences and holdings dispersion. More specifically, each month we compute measures of similarity in raw returns and holdings between each fund and the average fund *in* its assigned SPG and between each fund and the average fund *outside* its assigned SPG. Return differences are simply absolute differences in raw returns. Holdings dispersion, instead, is computed as follows:

$$Dispersion_{jt}^{G_{jt}} = \sum_{i=1}^{N_t^j} \left(w_{it}^j - \bar{w}_{it}^{G_{jt}} \right)^2 \tag{3}$$

for $G = [SPG,\ \widetilde{SPG}]$. Where SPG_{jt} is the peer group to which fund j is assigned at time t and $\widetilde{SPG}_{jt}$ represents all funds outside of SPG_{jt}. $\bar{w}_{it}^{G_{jt}}$ is the average weight allocated to stock i at time t by funds belonging to group G_{jt}.

We then run the following regressions:

$$D_{jt} = \alpha_1 + \gamma X_{jt} + \eta_t + \iota_f + \epsilon_{it} \tag{4}$$

$$\widetilde{D_{jt}} = \alpha_2 + \gamma X_{jt} + \eta_t + \iota_f + \epsilon_{it} \tag{5}$$

$$D_{jt} - \widetilde{D_{jt}} = \alpha_3 + \gamma X_{jt} + \eta_t + \iota_f + \epsilon_{it} \tag{6}$$

for $D = [RetDiff,\ Dispersion]$. X_{jt} are fund-specific demeaned control variables. The regressions includes month fixed-effects (η_t) and fund family fixed effects (ι_t). Standard errors are clustered at the month and fund level.

Table 3 reports the estimated α coefficients from the above regressions. Note that the high standard errors are due to the fact that the reported coefficient are regression intercepts. So, these results should be interpreted as differences in the mean of the dependent variable between funds in the same month and fund family, after controlling for fund-level characteristics. Model 3 shows that both absolute return differences and holdings dispersion are significantly lower *within* the assigned SPG than outside. This indicates that textual similarities are also reflected in similarities in funds' actions. In particular funds in the same SPG allocate their capital more similarly across assets, this translates in greater similarity in raw returns.

These results lead to reject hypothesis 2 that strategy descriptions in prospectuses are purely marketing material that is not reflected in funds' actions.

4.3 Test of Hypothesis 2: Risk-Adjusted Performance

We then test whether the different strategies implemented by the identified SPGs display significant differences in risk-adjusted performance. To test this hypothesis we ran the following regressions:

$$Perf_{jt} = \alpha + \beta I_{jt}^{SPG_{jt}} + \gamma X_{jt} + \eta_t + \iota_f + \epsilon_{it} \tag{7}$$

for $Perf = [Alpha,\ ValueAdded]$. Where $Alpha$ is obtained with 12-months rolling regressions of fund excess returns on the 5-factor Fama-French model plus the momentum factor.[12] $ValueAdded$ is obtained by multiplying $Alpha$ by fund j's TNA at time t. All

[12]Fund exposures to the 6 factors are obtained with rolling regressions using daily fund and factor returns. *Alpha* is then obtained as the difference between a fund's monthly return and the benchmark return obtained by multiplying monthly factor returns by the computed factor exposures.

other variables and regression specifications are the same as in Equation 2. Note that we run separate regressions for each SPG and performance measure (using a different $I_{jt}^{SPG_{jt}}$). I also run an additional regression excluding the indicator variable. The intercept of that regression can be interpreted as the average performance across all funds ($Overall$).

The coefficient of interest in each regression is β which, as in Equation 2, indicates then the incremental average performance of funds in the SPG of interest relative to that of funds in other SPGs (α).

Due to the potential impact of decreasing returns to scale on performance and given the differences in the average size of funds across SPGs, as reported in Table 2, the most appropriate measure of performance in this setting is value added.

Table 4 reports the estimated coefficients for the above regressions. For brevity only the βs are displayed for the regressions including the indicator variables; while only α is displayed for the regression which excludes the indicator variables. The last panel (row) of Model 2 shows that the average $ValueAdded$ across all funds in our sample is negative. No SPG displays a significantly higher risk-adjusted performance; while funds in the Dividends and Quantitative peer groups display a significantly lower risk-adjusted performance.

To further make sure that size differences are not just driven by differences in risk-adjusted performance, for each fund we also compute a measure of cumulative value added across all months of its existence. We assign each fund to their most frequent SPG assignemnt across the fund's history. We then run a similar regression to that in eq. 7, but excluding the time variation.

Table 5 reports the results of this regression. When comparing funds within the same family, the only consistent result is that the "Dividend" SPG deliver lower value added than other funds. With this specification, though there is some marginally sifgnificant evidence that the the MidCap SPG delivers a higher risk-adjusted performance at the 12-months horizon. Without adding family fixed-effects also the "Products and Services" SPG seems to deliver a higher risk-adjusted performance.

Whereas we find some evidence in favor of Hypothesis 2, these moderate differences in performance do not seem to justify the large differences in narrative descriptions, strategies and size previously reported.

4.4 Test of Hypothesis 3: Fund Strategies

Hypothesis 3 advances that funds in different peer groups might offer different "services" that appeal to different investor types. In this section we focus on the differences/commonalities of these strategies. Whereas in Section 4.6 we analyze how different investor types allocate capital across peer groups.

4.4.1 Risk Exposures

Even if we did not find substantial differences in the risk-adjusted performance of funds belonging to different SPGs, they might still differ in their exposures to risk factors. For this purpose we compute the exposure of each fund to the Fama-French 5 factors (Excess return on the market, Size, Value, Investment and Profitability) plus Momentum. We do so as follows: for each fund we run rolling 12 months regressions of daily excess returns on the fund on daily factor returns and we collect the fund's factor loadings. Then, for each SPG and factor we run a regression using the same specification as in Equation 2:

$$FactorLoading_{jt} = \alpha + \beta I_{jt}^{SPG_{jt}} + \gamma X_{jt} + \eta_t + \iota_f + \epsilon_{it} \tag{8}$$

where $FactorLoading_{jt}$ is the loading of fund j on the factor of interest in month t, computed using the prior 12 months of daily returns. The coefficient of interest is β, this indicates the incremental loading (positive or negative) of funds in the SPG of interest relative to the loading of the average fund in the sample (α).

Table 6 reports the results. Only the β coefficient is reported for each regression. The last Panel (row) displays the average loading across all funds to each of the factors (α). We observe that the average fund in each SPG presents a significantly different factor loading. For instance, in the last panel (row) of Model 1 we observe that the average loading of funds to the Market factor is slightly lower than 1 ($\alpha = 0.983$). But funds in the "Products & Services", "Small Cap" and "MidCap" clusters display a significantly higher average exposure to the market factor (βs are respectively: 0.0202, 0.0189 and 0.0119) which leads them to an almost 1 average exposure. On the contrary, funds in the "Dividends", "Derivatives", "Defensive", and "Fixed Income" SPGs have a significantly lower exposure (βs are respectively: -0.0230, -0.0216, -0.0168 and -0.0146). Differences are even more pronounced when looking at the size factor (Model 2). In fact the overall exposure of funds to the size factor is $\alpha = 0.222$

but, as expected, funds in the "Small Cap" SPG have almost three times the exposure to the size factor with a $\beta = 0.392$ (total exposure is $\alpha + \beta = 0.222 + 0.392$). Also funds in the "Products and Services" cluster have a significantly higher average loading to the size factor, albeit the difference being of smaller magnitude ($\beta = 0.0719$). On the other end funds in the "Dividends", "Derivatives', "Intrinsic Value" and Tax SPGs tend to hold relatively larger stocks loading less on the Size factor (βs are respectively: -0.167, -0.0325, -0.0470 and -0.0513). Differences of similar magnitude are observed in the loadings to the other risk-factors as well (Models 3-6).

An interesting example is to contrast the average loadings of funds in the "Dividends" and "Products & Services" SPGs relative to the average fund in the sample. In fact the "Dividends" SPG displays a significantly lower exposure to the market, size and momentum factors and a significantly higher exposure to the value investment and profitability factors. This is in line with the expected characteristics of high-dividend stocks. These are generally larger profitable firms which are further along in their life-cycle (i.e. load positively on value relatively to growth) and invest conservatively. In contrast, the average fund in the "Products & Services" cluster presents a higher exposure to the market, size and momentum factors and a lower exposure to the value investments and profitability factors. These seem to be smaller growth firms with positive momentum, which still have weak profitability and invest agressively. This is plausibly in line with what would be expected of a strategy that selects assets based on the assessed potential of their offered products and services.

Overall this analysis shows that, despite the similarity in risk-adjusted performance across SPGs, funds in different SPGs have significantly different combinations of risk exposures. These differences in risk exposures seem to be broadly in line with the disclosed strategy. Investors could achieve these exposures independently by investing in a combination of passive indices. Still, they might find it useful to invest in a single fund which provides the complex combinations of factor exposures that best matches their preferences.

4.4.2 Characteristics of Core Strategies

Next, we focus our attention to the characteristics of the assets held by funds in different SPGs. Hypothesis 3 implies that funds in the same SPG should invest in stocks with similar characteristics. In order to test this hypothesis we need to provide a more detailed analysis of the core strategies followed by each SPG and how closely each fund assigned to it follows its core strategy.

There are at least two possible ways to characterize these core strategies and whether funds are doing what they said they would. On the one hand, we can attempt to make a judgment about the typical investment behavior expected of each strategy group and test directly whether this behavior is observed for the funds in the group. This approach is narratively appealing but involves an unavoidable component of subjectivity. Reasonable observers may disagree about the core components of each strategy, which limits the generality of the insights such an approach can deliver. On the other hand, we can attempt to build a general measure of deviation from core strategies that does not depend on the particular interpretation one attaches to the strategy group or the particular expectations one might have for its investment behavior. We use a combination of these approaches, which is intended to mitigate concerns about subjectivity while allowing for more detailed descriptive insights.

4.4.3 General Measure: Characteristic Dispersion

Our general measure of deviation from core strategies is called Characteristic Dispersion ($CDisp$). This measure assumes that the core strategy of each peer group is whatever the average fund in that group is doing. The strategy is represented by a vector of normalized stock/firm characteristics, divided into eight categories: assets, liabilities, income statement, security market, information availability, sentiment, fund strategy and industry; we additionally consider "priced" characteristics. Asset characteristics are: current assets; inventories; non-performing assets; property, plant, and equipment; and intangibles – all scaled by total assets – and growth in total assets. Liability characteristics are: firm leverage (debt/assets); current liabilities; long-term debt; deferred taxes; all scaled by total assets. Income statement characteristics are: operating cash flow; R&D expenses – all scaled by total revenues – and earnings growth. Security market characteristics are: issuance and repurchases, scaled by shares outstanding; dividend yield and Amihoud illiquidity ratio. Information Availability characteristics are: number of analysts following the stock and number or recommendations, number of Dow Jones news articles, weighted by relevance – both scaled by market capitalization – and stock age. Sentiment characteristics are: average value and dispersion in analysts forecasts and news. Fund

Strategy characteristics are: percentage of American Depositary Receipts (ADR), foreign incorporated securities, cash and common stocks held; and the number of stocks held. Industry characteristics are: the percentage of TNA invested in each of the Fama-French 48 industries. "Priced" characteristics are so-labeled because they have been found in the asset pricing literature to be associated with common priced factors in stock returns, particularly by Carhart (1997) and Fama and French (2015): market beta, market capitalization, book-to-market ratio, one-year past returns, gross profitability (sales minus cost of goods sold, divided by assets), and investment (CAPEX divided by sales).[13]

Each of these characteristics is averaged across all stocks in each fund's portfolio, using as weights the percentage of total net assets invested in each stock. Hence, the average value for characteristic c in the portfolio of fund j at time t is given by:

$$Char_{c,j,t} \equiv \sum_{i=1}^{N_{j,t}} w_{i,j,t} c_{i,t} \tag{9}$$

where $N_{j,t}$ is the number of assets held by fund j at time t; $c_{i,t}$ is the value of normalized characteristics c for stock i at time t; and $w_{i,j,t}$ is the percentage of fund j's assets allocated to asset i at time t.

We further define $\left(\overline{Char_{c,t}}\right)_{SPG_{j,t}}$ to be the average $Char_{c,j,t}$ over all funds belonging to the same SPG as fund j at time t.

For each of the nine groups of characteristics defined above we construct a dispersion measure as the sum of squared differences between the average normalized characteristics for each fund and that for the average fund in their peer group.

$$CDisp^{I}_{j,t} = \sum_{c=1}^{K^I} \left[Char_{c,j,t} - \left(\overline{Char_{c,t}}\right)_{SPG_{j,t}} \right]^2 \tag{10}$$

where

$I = [assets,\ liabilities,\ income\ statement,\ industires,\ information\ availability,\ fund\ strategy,\ priced]$ and K^I is the number of characteristics in group I.

Finally, we construct our two baseline measure of characteristics dispersion: one using only "priced" characteristics ($CDispP$) and the other using only all other characteristics

[13] Stock level characteristics are normalized across all stocks in the CRSP universe before merging to the mutual fund dataset. Fund strategy characteristics are normalized after.

($CDispNP$). This last measure is obtained as an equally-weighted average of the dispersion measure defined in equation 10 for the remaining 8 categories of non-priced characteristics.

This separation allows us to distinguish between characteristics that carry aggregate risk exposure for which investors expect to be compensated, and idiosyncratic characteristics that carry no compensation but may still be important to individual investors due to their particular preferences and/or constraints.

Finally, we define alternative measures of characteristic dispersion with respect to the average fund *not* in the same strategy peer group, which we denote by $\widetilde{CDispNP}$ and $\widetilde{CDispP}$, respectively.

Using these measures, we test the following hypothesis:

$$H_1 : \mathbb{E}[CDisp_{i,t} < \widetilde{CDisp}_{i,t}],$$

A rejection of the null hypothesis indicates that the average fund is investing in a manner more similar to the other funds in its peer group than to a random fund outside of the group, which we interpret as promise-keeping.

We test this hypothesis using the following regression:

$$CDisp_{i,t} - \widetilde{CDisp}_{i,t} = \alpha + \gamma' X_{i,t} + \eta_t + \iota_f + \varepsilon_{i,t}. \tag{11}$$

The coefficient of interest is $\hat{\alpha}$, which estimates the difference between within-group ($CDisp$) and outside-group dispersion ($\widetilde{CDisp}$) when all control variables are equal to their mean values (all controls are demeaned). The control variable vector, X, contains the log of assets under management, the log of fund age, the fund's expense and turnover ratios, monthly percentage flows, and monthly flow volatility. η_t are month fixed effects, and ι_f are fund family fixed effects. Standard errors are clustered by fund and month.

Table 7 reports the estimated $\hat{\alpha}$s in the third column. Each panel (rows) shows the results for the different measures of characteristic dispersion. The first two panels display the estimated α when using only "priced" characteristic ($CDisp\_$"$Priced''$) and that built using only "non-priced" characteristics ($CDisp\_$"$NotPriced''$). The coefficients of -0.223 and -0.0755 in the first two panels, respectively, indicates that funds not in the same peer group are significantly more dispersed in their priced and non-priced characteristics compared to funds in the same peer group (all results significant at the 1% level). The same is true for the following panels in which we decompose the "non-priced" measure into its 8 components. Therefore we conclude that, on average, funds appear to be keeping their promises to investors and investing in stocks with characteristics aligned with the core strategy of the SPG they belong to.

We note that this interpretation does not depend on the average level of dispersion within any particular group; as long as funds with high in-group dispersion still have higher dispersion with respect to funds outside of the group. However, we acknowledge that our measure is limited by how well we have captured the characteristics that matter for all of the strategies. The $\hat{\alpha}$ could be biased towards zero if we miss a major axis of commonality for a particular strategy. One such example is the "Derivatives" peer group—our data does include the percentage of assets other than cash and common stock but does not include derivative position values specifically, which should certainly be a significant factor for that SPG.

4.5 Descriptive Analysis

We now turn our attention to a more detailed investigation of the behavior of funds in individual peer groups, and whether this behavior is consistent with the funds' self-described strategies. Before proceeding, we note a few important caveats about this exercise. First, the labeling of strategy groups based on word frequencies necessarily involves a subjective element, as does the assessment of what funds in a particular strategy group "should" be doing. As such, we try to focus on broad interpretations we think reasonable third-party observers would agree with, and to avoid drawing conclusions that are too specific or where potential ambiguity is high. Second, the large set of potential stock and fund characteristics carries a risk of data-mining or selective presentation, especially as there are too many results to display in the main body of the paper. To mitigate this concern, we try to focus on those characteristics for which there is strong ex-ante justification, while reproducing the unedited tables.

Importantly, none of the results elsewhere in this paper depend on the particular narrative choices we make in this section. Nonetheless, in our view the overall conclusions of the strategy-by-strategy analysis support our interpretation of characteristic dispersion as a measure of deviation from the core strategy. This analysis also has independent descriptive value regarding the active equity investment landscape in the United States.

For each characteristic of interest and each SPG we run the following regression:

$$Characteristic_{jt} = \alpha + \beta I_{jt}^{SPG_{jt}} + \gamma X_{jt} + \eta_t + \iota_f + \epsilon_{it} \tag{12}$$

where $Characteristic_{jt}$ are the average normalized characteristics described in section 4.4.3 for all stocks held by fund j at time t. Averages are computed by weighting each stock's characteristic by the percentage of TNA invested by fund j in that stock at time t. All other variables and regression specifications are the same as in Equation 2.

Tables 8 to 14 report all results.[14] A broad analysis reveals that some of the features we would have expected from these strategies are indeed present. For instance, funds in the "Products & Services" SPG are the ones to invest in stocks with the highest R&D investment, while funds in the "Quantitative" and "Intrinsic Value" SPGs are the ones to invest in stocks with the highest shares repurchases, likely a sign of undervaluation; while funds in the "Small Cap" SPG invest in stocks with the lowest shares repurchases. Funds in the "Dividends" SPG are the ones to invest in the oldest stocks with the highest dividend yield, while funds in the "Small Cap" SPG invest in the youngest stocks. Funds in the "Dividend" and "Tax" SPGs are the ones to invest in stocks with the highest deferred taxes. Funds in the "Fixed Income" SPG, instead, are those to have the lowest percentage of common stocks in their portfolios.

4.5.1 A Detailed Example: Dividends

The "Dividends" group arguably has the most straightforward strategy and the clearest ex-ante expectations for investors. The following excerpt from the prospectus of the Federated Strategic Value Dividend Fund, published in 2013, is typical:

> *The Fund pursues its investment objective by investing primarily in high dividend yielding, undervalued common stocks with dividend growth potential. The Adviser believes a strategic emphasis on high dividend yielding stocks can enhance both relative and absolute performance over time. In addition, investment results can be enhanced by focusing on stocks with both the potential for future dividend growth and strong value characteristics. The Adviser believes that this is achievable while targeting significantly less risk.*

[14] Table 8 only reports results for a selected number of industries.

Ex-ante, a reasonable investor would expect such a fund to invest disproportionately in stocks with higher dividend yields. Other characteristics typically associated with high-dividend-paying stocks are: (i) relatively low levels of cash on the balance sheet, and (ii) belonging to high-payout industries such as Utlities, Telecoms, and Consumer Staples. Tables 8 and 12 show that all of these characteristics are indeed significantly associated with the Dividend strategy group, with the expected sign. Other interpretable features of this strategy are investment in stocks with lower than average R&D investment, current assets and liabilities, inventories, intangibles and asset growth. While they invest in relatively older stocks which have higher than average non-performing assets, property plant and equipment, deferred taxes, long term debt, operating activities, issuance, repurchase and liquidity. The stocks they invest on additionally present a lower number of analysts following them and fewer news articles but a higher average analyst recommendation. Finally, these funds hold on average less cash and fewer stocks. Somewhat surprisingly, they also invest a higher percentage of their portfolios in American Depositary Receipts and a lower percentage in common stocks. Taken all together, the average Dividend fund appears to be selecting assets to keep its promise to investors of investing in high dividend paying stocks.

4.6 Test of Hypothesis 3: Investor Types

The second part of our test of hypothesis 3 consists in checking if investors with different preferences allocate capital differently across funds with different characteristics.

Given the conclusion from Section 4.4, we consider SPGs to be a good proxy for identifying funds with similar strategies and hence similar risk exposures and underlying characteristics of stocks held. In order to proxy for the behavior of investors with different preferences, we divide the assets managed by each fund every month into those coming from retail, retirement or institutional investors. We do so by categorizing all share classes offered by each fund into the three categories of interest and aggregating their TNA accordingly.[15]

First, as displayed in Figure 6, we note that the total amount of retail capital managed by these funds between 2000 and 2017 has remained virtually unchanged, whereas the amount

[15]We use both the institutional and retail identifiers provided in CRSP Mutual Fund dataset, as well as fund names to further isolate the retirement share class.

of institutional and retirement capital have been steadily increasing. Hence, while at the beginning of 2000 retail capital represented to 87.26% of all capital managed by these funds, by the end of 2017 retail capital represented only 52.89%, the remainder being split between institutional (37.05%) and retirement (10.06%) capital.

To further assess whether these changes are mostly due to inflows, outflows or capital appreciation we compute every month aggregate net flows and the growth in aggregate net flows at the investor-type level. Aggregate net flows are obtained by first computing net flows at the share class level for each fund-month as the change in TNA which is not due to returns. We then sum these net flows over all retail, institutional or retirement share classes for each month. The growth in aggregate net flows by investor-type is then obtained by dividing aggregate net flows by the total amount of TNA present in the market for each investor-type at time $t-1$. Figure 9 displays a time-series of these two measures for the three investor-types. As we can observe from Panel 1, in the time frame of our analysis retail share classes have mostly experienced outflows, whereas institutional share classes have experienced more balanced inflows and outflows. The retirement share classes, instead, has experienced more substantial inflows in recent years.

We then ask whether these different growth rates in the capital invested by investor with likely different preferences (retail, institutional or retirement) contribute to explain differences in net flows growth and in total capital allocated to the different SPGs.

In this paper we won't be able to perfectly disentangle whether these equilibrium effects are driven by supply or demand. For instance, we won't be able to say if the increase in institutional and retirement capital is due to a targeted offering and marketing efforts from fund families or whether the additional demand prompted fund families to increase their offering of certain products. What we are interested in, though, is the equilibrium outcome of this matching. Our goal is to assess if the different characteristics offered by funds in different SPGs, despite not generating different risk-adjusted performance, might appeal to investors with different preferences. So allowing us to better characterize heterogeneous investor preferences in the perspective of a demand based asset pricing.

We proceed in three steps. First we check whether increases in the aggregate total net assets of different investor types get allocated differently among funds in different SPGs. Next, we focus on fund offering and we check whether the likelihood of funds to offer a certain share class can be explained by their SPG assignment. Finally, we ask whether the percentage of TNA of each fund coming from retail, institutional or retirement investors can be explained by the fund's SPG assignment.

4.6.1 Net Flows Allocation

First, we analyze whether there are significant differences in how investor types allocate their capital across SPGs. In particular, we analyze whether net flows from different investor types are allocated more (less) than proportionally to some SPGs relative to others. We do so by running the following regression for each SPG:

$$\begin{aligned} NetFlowsGrowth_{jt} = \alpha + \beta_1 I_{jt}^{SPG_{jt}} + \\ + \beta_2 I_{jt}^{SPG_{jt}} * NetFlowsGrowth_t^R + \\ + \beta_3 I_{jt}^{SPG_{jt}} * NetFlowsGrowth_t^I + \\ + \beta_4 I_{jt}^{SPG_{jt}} * NetFlowsGrowth_t^{Rt} + \\ + \sum_{\tau=1}^{12} \delta_\tau R_{j,t-\tau} + \gamma X_{jt} + \eta_t + \iota_f + \epsilon_{it} \end{aligned} \quad (13)$$

where $NetFlowsGrowth_{jt}$ is the percentage growth in net fund flows or fund j at time t which is not due to returns. While $NetFlowsGrowth_t^R$, $NetFlowsGrowth_t^I$ and $NetFlowsGrowth_t^{Rt}$ are respectively the normalized percentage growth in aggregate retail, institutional and retirement net fund flows at time t. As in previous specifications, $I_{jt}^{SPG_{jt}}$ is an indicator variable equal to 1 if fund j belongs to the SPG of interest at time t, 0 otherwise.

The inclusion of fund family and month fixed effects allows us to essentially compare funds belonging to the same fund family in the same month, one belonging to the SPG of interest and the other not. This alleviates potential endogeneity concerns. Including demeaned controls for fund-level characteristics (X_{jt}) and fund returns in the previous 12 months ($R_{j,t-\tau}$) also alleviates concerns that the estimated coefficients are driven by differences in performance or other similarities at the SPG level. Note that aggregate retail, institutional and retirement net flows growths are not included directly in the regression as they are time-invariant, hence they are collinear with the time fixed-effects.

The coefficients of interest are β_2, β_3 and β_4. These indicate whether funds belonging to the SPG of interest ($I_{jt}^{SPG_{jt}} == 1$) experience a higher (lower) increase in net fund flows corresponding to increases in the aggregate net flows of retail, institutional or retirement capital, relative to funds belonging to any other SPGs.

The estimated β_2, β_3 and β_4 for these regressions are displayed in Table 15; Model 3 includes all fixed effects of interest.[16] We find that for half of the SPGs there are significant differences in how different investor types allocate their capital. In particular, between 2000 and 2017 outflows of retail capital disproportionally affected funds in the "Dividend", "Quantitative" and "Foreign (ADR)" SPGs, while funds in the "Competitive Advantage" SPG have been impacted significantly less. Similarly, funds in the "Quantitative" SPG have been allocated proportionally less of the net flows in institutional money, while funds in the "Competitive Advantage" SPG have been allocated significantly more. Finally, we find that funds in the "Derivatives", "Quantitative" and "Foreign (EM)" SPGs were allocated proportionally more of the total Retirement net flows, while funds in the "Fixed Income" and "Intrinsic Value' SPGs were allocated significantly less.

4.6.2 Offering Likelihood

The above result shows that, in our period of interest, the proportion of aggregate net flows that different investor types allocate to SPGs differs. Yet, there might be smaller SPGs or SPGs that did not experience much net flows in our time frame of interest which are mostly driven by capital of a certain investor type. The above analysis would not be able to capture those effects.

Hence, we first look at the extensive margin: i.e. at the likelihood of a certain share class being offered, given a fund's SPG assignment. In other words, are fund managers in certain SPGs more likely to offer retail, institutional or retirement share classes than managers in other SPGs? In order to answer that question we run the following regression for each SPG and investor type:

$$D\_ShareClass_{jt}^{Type} = \alpha + \beta I_{jt}^{SPG_{jt}} + \gamma X_{jt} + \eta_t + \iota_f + \epsilon_{it} \tag{14}$$

for $Type$ in $[Retail,\ Institutional,\ Retirmenent]$. Where $D\_ShareClass_{jt}^{Type}$ is an indicator variable equal to 1 if fund j offers a share class of type $Type$ at time t, 0 otherwise.

[16] Only estimates for which some of the coefficients of interest were significant are displayed.

All other variables and regression specifications are the same as in Equation 2. Adding fund family and month fixed effects allows us to compare the likelihood that two managers of the same fund family in the same month offer a retail, institutional or retirement share class given that one belongs to the SPG of interest and the other does not. Hence, the coefficient of interest (β) tells us the incremental likelihood relative to the average likelihood for funds not in the SPG of interest (α).

Table 16 reports the estimated coefficient of interest for all regressions. We observe that for 10 SPGs there exists a higher (lower) likelihood of a certain share class being offered. More specifically, relative to the average fund, there exists a significantly higher likelihood that funds in the "Derivatives", "Defensive", "Intrinsic Value" and "Fixed Income" SPGs will offer a retail share class; whereas there is a significantly lower likelihood that funds in the "Small Cap" and "Foreign EM" SPGs will offer a retail share class. For what regards institutional share classes, there is a significantly higher likelihood that they will be offered by funds in the "Foreign ADR" SPG, while there is a lower likelihood that funds in the "Dividends" or "Intrinsic Value" SPGs will offer them. Finally, for what regards retirement share classes, they are significantly more likely to be offered by "Small Cap" and "PE Ratio" SPGs, while they are significantly less likely to be offered by "Dividends" or "Tax" SPGs. It is worth noting that, the lowest likelihood of any share class to be offered is the retirement class for the "Tax" SPG. Indeed, taxation is usually deferred on retirement capital, while different tax policies apply to individuals and corporations. Hence, the "Tax" SPG does not cater to the preferences of retirement investors.

4.6.3 TNA Shares

In our final analysis we focus on the intensive margin; i.e. conditionally on a certain share class being offered, does the SPG assignment contribute to explain the percentage of retail, institutional or retirement capital managed by a given fund? We do so by running the following regression for each SPG and investor type:

$$Perc\_ShareClass_{jt}^{Type} = \alpha + \beta I_{jt}^{SPG_{jt}} + \gamma X_{jt} + \eta_t + \iota_f + \epsilon_{it} | D\_ShareClass_{jt}^{Type} == 1 \quad (15)$$

this specification is equivalent to that in Equation 14, with the difference that the left-hand-side variable is the percentage of TNA of fund j at time t which belongs to share classes of type $Type$. Here we condition of fund j offering at least one share class of type $Type$ at time t.

Thanks to the fixed effects, we compare funds that belong to the same fund family in the same month, one belonging to the SPG of interest and the other not. Then we ask whether belonging the SPG of interest translates into a higher (lower) percentage of TNA coming from retail, institutional or retirement investors. Hence, the coefficient of interest is β, represents the incremental percentage of capital of type $Type$ due to belonging to the SPG of interest, relative to the average percentage for funds not in the SPG of interest (α).

Table 17 reports the estimated βs for all regressions. We observe that for 11 of the SPGs the percentage of capital coming from a certain investor type is significantly more (less) prevalent. More specifically, conditionally on offering a retail share class, funds in the "Dividends" and "Intrinsic Value" SPGs manage a greater percentage of retail money than the average fund; while funds in the "Undervalued", "Small Cap" and "Fixed Income" SPGs manage a significantly lower percentage of retail capital. For what regards institutional capital, conditionally on offering an institutional share class, funds in the "Small Cap", "Foreign (EM)" and "Undervalued" SPGs manage a greater percentage of institutional capital; while funds in the "Dividends" and "Quantitative" SPGs manage a relatively lower percentage of institutional capital. Finally, conditionally on offering a retirement share class, funds in the "PE Ratio" SPG manage a higher percentage of retirement capital than the average fund, while funds belonging to the "Derivatives", "Defensive" and "Fixed Income" SPGs manage a significantly lower percentage of retirement capital.

4.6.4 Detailed Examples

Figure 10 displays graphically some examples of the cumulative amount of capital coming from different investor types managed by funds belonging to different SPGs. The SPGs considered are: "Competitive Advantage" (Panel 1), "Quantitative" (Panel 2), "Tax" (Panel 3) and "Dividends" (Panel 4).

The total amount of capital of all fours SPGs has been increasing over time. For the retail share class, this is mostly due to capital appreciation, as there has been an overall net outflow of retail capital from active equity mutual funds (Figure 9). Retirement share classes, instead, have experienced net inflows (Figure 9).

For what regards the first two examples "Competitive Advantage" and "Quantitative" we can observe that the overall growth of capital has been mostly driven by institutional and retirement capital. For the "Dividend" SPG, instead, it has mostly been driven by capital appreciation in the existing retail share class. Indeed Tables 16 and 17 shows that the likelihood of "Dividend" funds offering institutional or retirement share classes and them constituting a high percentage of TNA is low. While, Table **??** shows that the "Dividend" share class has experienced a greater outflow of retail capital than other SPGs. Finally, we can see from the "Tax" example that close to 0% of the capital in this SPG belongs to the retirement share class, as also shown in Table 16.

4.6.5 Overall

Taken together these results and examples paint a clear picture: different investor types invest more in funds belonging to different SPGs. Given the lack of differences in risk-adjusted performance across SPGs, these differences in capital allocation are likely due to different preferences and needs across investor types. Hence, these correlated investor flows likely mimic different investor preferences over complex risk-factor exposures or other non-priced characteristics.

5 Conclusion

In this paper we measure active equity mutual funds' differentiated offering and we show that different products attracts investors with heterogeneous preferences.

In order to measure funds' differentiated offering, we collect all Principal Investment Strategies sections from mutual fund prospectuses, in which funds provide a narrative description of the key features of their asset selection strategies. We analyze the full content of these descriptions and use them to categorize funds into Strategy Peer Groups; i.e. groups of funds who present high similarity in their strategy descriptions. We do so by using the K-Means algorithm, a standard tool in unsupervised machine learning. We show that funds can be clustered into [16] different strategy peer groups, which present interpretable differences in their descriptions.

Next, we assess whether these differentiated offerings translate into effective commonality in actions among funds belonging to the same SPG. We find that indeed funds that have a greater similarity in their strategy descriptions also display more similar raw returns and holdings. These differences though do not translate into meaningful differences in risk-adjusted performance. That is despite the fact that SPGs display very different sizes over time.

In order to explain the equilibrium allocation between funds supply and investors' demand we ask whether different SPGs provide additional "services" that investors might care about. First, we show that funds in the same SPG have similar and interpretable loadings to risk factors. These loadings are complex, so investors with different preferences might benefit from being exposed to a specific loadings combination. We then show that SPGs display significant differences in their core strategies and funds belonging to a certain SPG tend to conform more to those strategies. We measure core strategies using a host of different stock level characteristics and show in a descriptive section that these are mostly in line with what would be expected from the disclosed strategy.

Finally, we study how different investor types allocate across these SPGs. First we show that net flows from different investor types are not allocated evenly acorss SPGs. These differences in allocation cannot be explained by differences in performance, fund family characteristics, trends or other fund level characteristics. Next we show that, in the extensive margin, funds belonging to certain SPGs are more (less) likely to offer share classes targeting different investor types. Finally, in the intensive margin we show that SPG assignment has incremental explanatory power on the percentage of TNA coming from different investor types.

Taken together, these results show that funds specialize in different strategies or mandates and investors self-allocate to these strategies according to heterogeneous preferences. These correlated flows, combined with similarities in the characteristics of the underlying stocks held, might impose price pressure on the underlying securities, particularly if flows are driven by investor preferences and not fundamental information.

In future work, we plan to assess the asset pricing implications of flow-induced-trading at the SPG level.

References

Abis, S. (2020). Man vs. machine: Quantitative and discretionary equity management. *Working paper*.

Beggs, W., J. Brogaard, and A. Hill-Kleespie (2019). Quantitative investing and market instability. *Working Paper*.

Ben-David, I., J. Li, A. Rossi, and Y. Song (2020). Non-fundamental demand and style returns. *Working paper*.

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

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance 52*(3), 1035 – 1058.

Elton, E. J., M. J. Gruber, and C. R. Blake (2001). A first look at the accuracy of the crsp mutual fund database and a comparison of the crsp and morningstar mutual fund databases. *The Journal of Finance 56*(6), 2415–2430.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology 32*(3), 221–233.

Hillert, A., A. Niessen-Ruenzi, and S. Ruenzi (2016). Mutual fund shareholder letters: flows, performance, and managerial behavior. *Working Paper*.

Hong, H. and M. Kacperczyk (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 15–36.

Hunter, D., E. Kandel, S. Kandel, and R. Wermers (2014). Mutual fund performance evaluation with active peer benchmarks. *Journal of Financial Economics 112*(1), 1–29.

Kacperczyk, M., C. Sialm, and L. Zheng (2008). Unobserved actions of mutual funds. *Review of Financial Studies 21*(6), 2379–2416.

Kincaid, P. J., R. P. Fishburne, R. L. J. Rogers, and B. S. Chissom (1975). Derivation od new readability formulas (automated readability index, fox count and flesch reading ease formula) for navy enlisted personel. *Institute for Simulation and Training 56*.

Koijen, R. S., R. J. Richmond, and M. Yogo (2020). Which investors matterfor equity valuations and expected returns? *Working Paper*.

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

Kostovetsky, L. and J. B. Warner (2020). Measuring innovation and product differentiation: Evidence from mutual funds. *Journal of Finance 75*(2), 779–823.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? textual analysis dictionaries and 10-ks. *Journal of FInance 66*(1), 35–65.

SEC (2006).

SEC (2016). How to read a mutual fund prospectus. Investor Alerts and Bulletins.

Shive, S. and H. Yun (2013). Are mutual funds sitting ducks? *Journal of Financial Economics 107.1*, 220–237.

Zhu, Q. (2020). The missing new funds. *Management Science*.

A K-means

A.1 The Algorithm

The K-Means algorithm takes as inputs the $tfidf$ matrix, the number of desired clusters (k) and a tolerance threshold (τ). The algorithm is initialized by choosing k points in the vector space (centroids). Points are deliberately chosen to be far from each other in order to minimixe the likelihood of converging to a local minimum. Each chosen point represents a features vector of same length as the number of chosen features ($10,000$), whose elements exist in $[0, 1]$. Then the following steps are repeated until the pre-defined tolerance level is reached:

1. Calculate the euclidian distance between the vectors representing each document (rows of the tfidf matrix) and each of the k centroid vectors as follows:

$$\sum_{r=1}^{R} ||x_r - x_r^C||^2 \tag{16}$$

   where x_r is the frequency assigned to feature r in a specific document and x_r^C is the frequency assigend to feature r in a cluster's centroid. R is the total number of features.

2. Assign each document to the closest centroid (form clusters)

3. Generate new centroids (features vectors) by taking the item-by-item average of the feature vectors of all documents assigned to the same cluster

4. Calculate the euclidian distance between the centroids at iteration n and those at iteration $n + 1$.

   - If the largest distance is greater than the tolerance level τ, repeat all steps
   - Otherwise exit the loop and return the formed clusters (convergence)

We ran the above algorithm with different specifications for the user defined parameters (k and τ). All runs are independent (use different seeds). Despite the possibility of K-means reaching a local optimum, in our setting, the procedure is robust to changes in initial parameters (see discussion in Section A.2).

In the main specification we use the default value for τ in Python's scikit-learn implementation: 0.0001. We use 16 clusters for the determination of the Strategy Peer Groups (SPG).

A.2 Optimal Clusters Number

In order to choose the correct number of clusters we run independent runs of the K-means algorithm for $K = [10, 20]$. We then compare the categorization between each consecutive [K] and [K+1] optimal solutions. We base the choice of the optimal number of clusters on two criteria which we label: density and stability.

Observations Cross-Classification: Stability Define the crosstab matrix as the number of observations falling in cluster i under [K] and cluster j under [K+1]

$$CrossTab_{(i,j)} = \#\ clustered\ as\ i\ under\ [K]\ and\ j\ under\ [K+1],$$

If we treat [K+1] as the ground truth, and [K] as the predicted value, for any combination (i, j), the denominator of its precision is the sum for all j given i, and the denominator of its recall is the sum for all i given j. Formally, define precision and recall as:

$$Precision_{(i,j)} = \frac{CrossTab_{(i,j)}}{\sum_{i=1}^{K} CrossTab_{(i,j)}}$$

$$Recall_{(i,j)} = \frac{CrossTab_{(i,j)}}{\sum_{j=1}^{K+1} CrossTab_{(i,j)}}$$

Intuitively, if $Precision_{(i,j)}$ is large, it means that observations classified as i under [K], are very likely to be classified as j under [K+1], meaning that i under [K] is likely to be a subset of j under [K+1]. Similarly, if $Recall_{(i,j)}$ is large, it means that observations classified as j under [K+1], are very likely to be classified as i under [K], meaning that j under [K+1] is likely to be a subset of i under [K].

We finally combine the two above criteria into an $Fscore$ matrix indicating the harmonic mean of precision and recall. Due to the characteristic of the harmonic mean, if $Fscore_{(i,j)}$ is large, $Precision_{(i,j)}$ and $Recall_{(i,j)}$ are both expected to be large, and cluster i under [K] is likely to be in line with cluster j under [K+1].

$$Fscore_{(i,j)} = 2 \cdot \frac{Precision_{(i,j)} \cdot Recall_{(i,j)}}{Precision_{(i,j)} + Recall_{(i,j)}}$$

We currently use a threshold of 0.5 to tell whether the F1 score is large enough:

$$\widehat{Fscore}_{(i,j)} = Fscore_{(i,j)} > 0.5$$

Note that a high F1 score in this context indicates a high stability over the independent optimal allocations found using the K-means algorithm with [K] and [K+1] clusters (i.e. these are unlikely to be local minima).

Euclidean Distance: Density Define the Eucledian distance between the centroids of any pair of clusters under [K] and [K+1] as:

$$Dist_{(i,j)} = ||C_i^K - C_j^{K+1}||_2$$

where C_i^K indicates the centroid of tfidf vector of cluster i under [K], and C_j^{K+1} indicates the centroid of tfidf vector of cluster j under [K+1].

If the distance between two centroids is very small, the underlying clusters are likely to be very similar in meaning. We currently, use the threshold 0.2 to tell whether the distance is small enough:

$$\widehat{Dist}_{(i,j)} = Dist(i,j) < 0.2$$

Optimal Choicwe: Stability and Density For row i in a criterion matrix ($\widehat{Fscore}_{(i,j)}$ or $\widehat{Dist}_{(i,j)}$), if the sum of that row is 0 (i.e. it does not include '1s'), cluster i is likely to be a new cluster; if the sum of that row is 1 (i.e. the row includes only 1 '1' in column j), then cluster i under [K] is likely to be in line with cluster j under [K+1]; if the sum exceeds 1 (i.e. the row includes more than 1 '1'), cluster i is likely to split into multiple fractions under [K+1], and the new clusters under [K+1] are columns whose values are 1. A similar reasoning applies to column j. Note that matched clusters are those for which the corresponding column/row in the criteria matrices equals to 1; if the column's sum equals

0, the column cluster is a new cluster; if row sum is greater than 1, the corresponding rows are new clusters.

In essence, stability indicates that most clusters could be matched across the two independent runs [K] and [K+1]. Density indicates that any unmatched cluster is non-trivial. We choose the optimal K (K^*) such that both criteria are satisfied.

Figure 1: **Distribution of word count for Strategy Sections:** Pooled distribution of the number of words contained in of all fund–month observations, for the Strategy section.

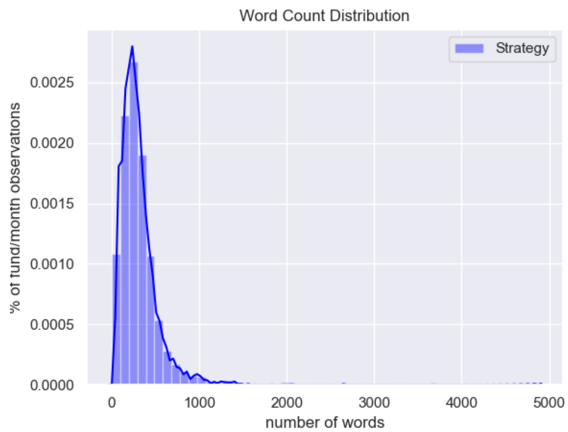


Figure 2: **Distribution of Complexity for Strategy Sections:** Panel 1 displays the pooled distribution of the Flesch-Kincaid grade level complexity measure across all fund–month observations, for Strategy sections. This measure indicates the number of years of schooling required in order to comprehend each section. Panel 2 displays the same distribution for the Flesch-Kincaid reading ease measure. This measure indicates, on a scale of $[1, 100]$, how easily a section can be read (a higher score indicates lower complexity). Both measures are based on the relative number of total words to total sentences (average sentence length) and the relative number of total syllables to total words (average word length) contained in each section.

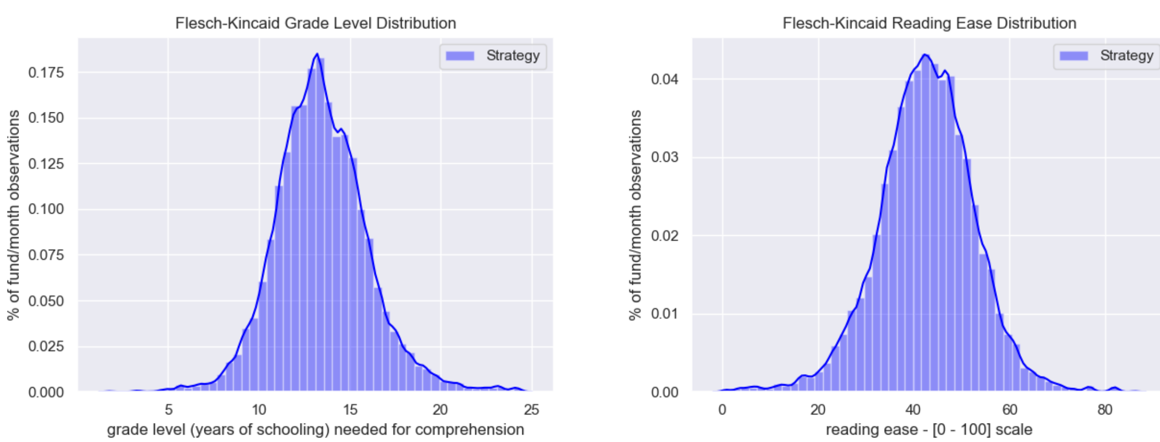


Figure 3: **Frequency of Positive, Negative Uncertainty and Litigious Words in Strategy Sections:** Panel 1 displays the pooled distribution of the frequency of *positive* words for all fund–month observations, for Strategy sections. Panel 2 display the same distribution for the frequency of *negative* words. Panel 3 for the frequency of *uncertainty* words and Panel 4 for the frequency of *litigious* words. The frequency of words per section is obtained by computing the percentage of the total number of words in each section that appears in the Loughran and McDonald *positive*, *negative*, *uncertainty* or *litigious* dictionaries. These dictionaries are adapted to account for specific characteristics of financial language Loughran and McDonald (2011)).

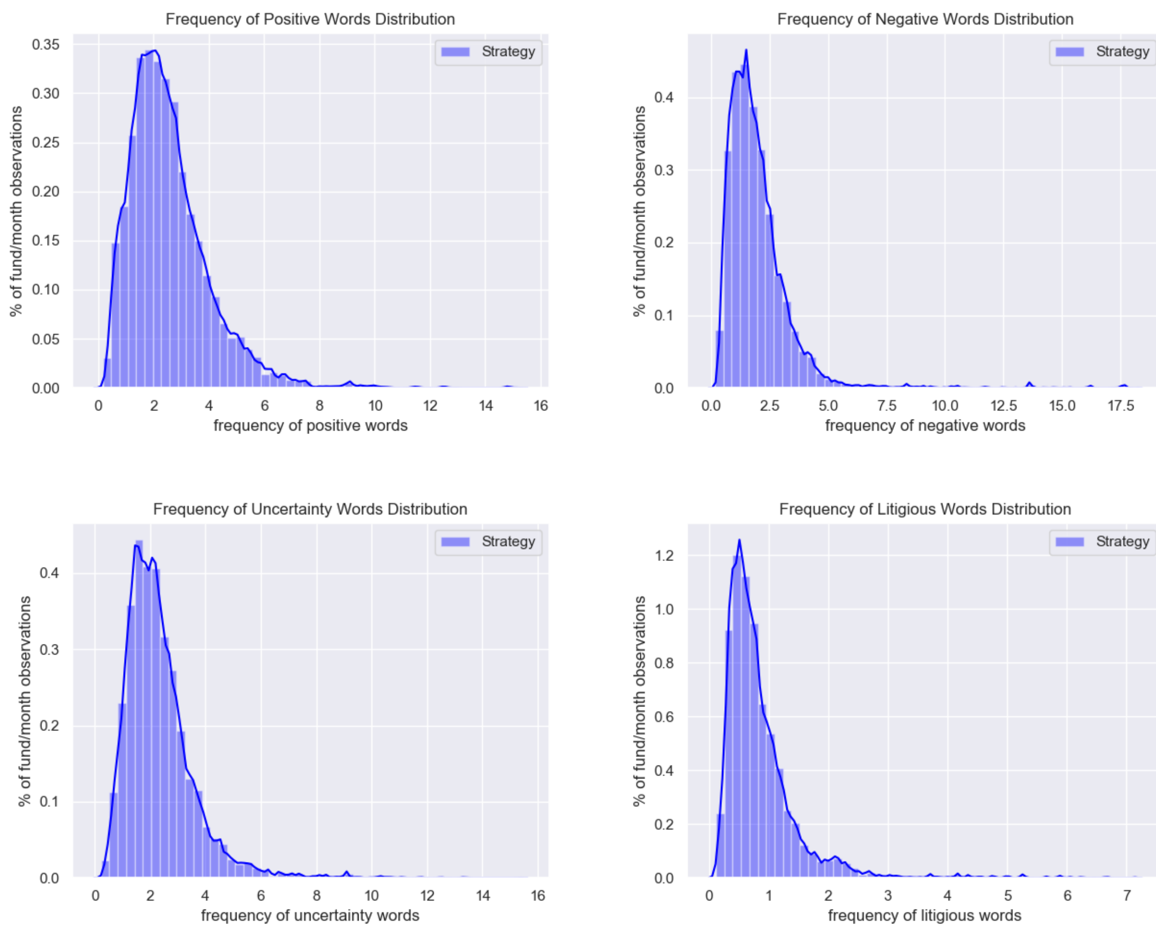
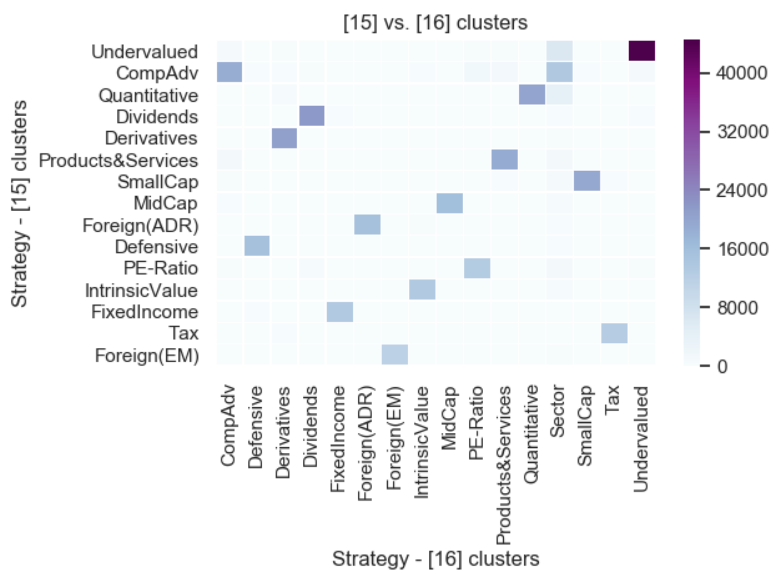


Figure 4: **Strategy - Cluster Assignment using [15], [16] and [17] clusters:** Each square in the heat maps represents the number of strategy sections assigned to a specific cluster combination using K-Means with a lower (rows) vs. higher (columns) number of clusters. A darker color indicates a higher number of sections being classified in that specific combination, according to the color map on the right hand side. Panel 1 shows the cross-allocation of strategy sections when going from [15] (rows) to [16] (columns) clusters. Panel 2 shows the cross-allocation of strategy sections when going from [16] (rows) to [17] (columns) clusters.



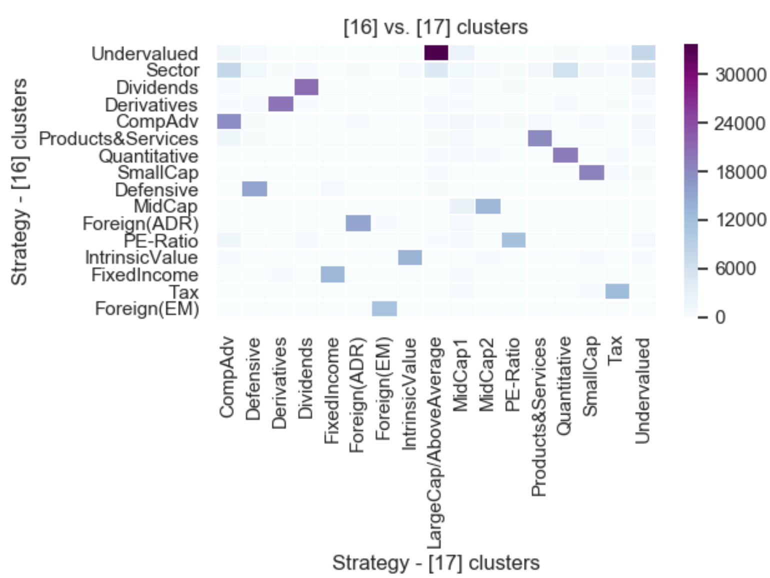
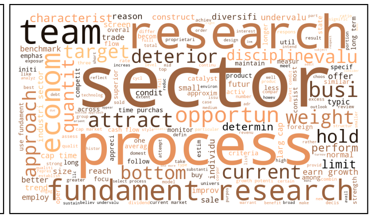


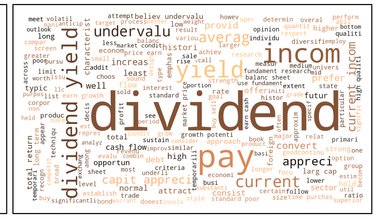
Figure 5: **Strategy word clouds – [16]:** Word clouds for all estimated SPGs using K-Means with [16] clusters. These represent the frequency of features (words and bi-grams) in the strategy sections belonging to each SPG. Words size is proportional to their frequency. The clouds are presented in order of size, with the most frequent clusters coming first.

(a) Undervalued

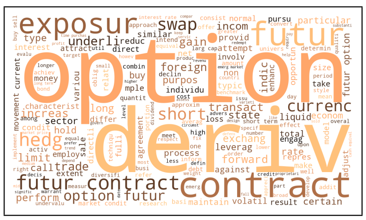
(b) Sector



(c) Dividends



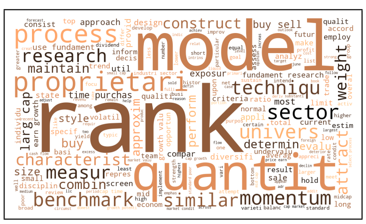
(d) Derivatives



(e) CompetitiveAdvantage

(f) ProductsServices

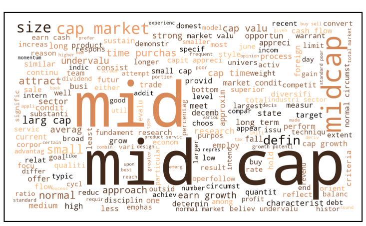
(g) Quantitative



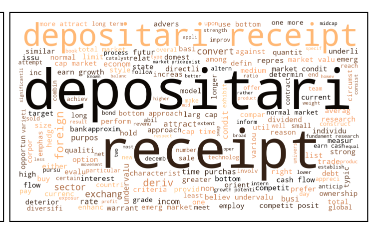
(h) SmallCap

(i) Defensive

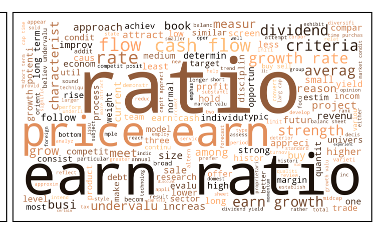
(j) MidCap



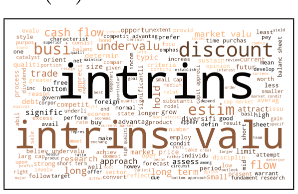
(k) ForeignADR



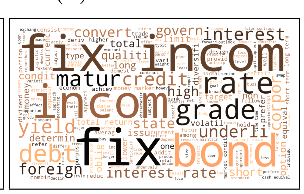
(l) PE-Ratio



(m) IntrinsicValue



(n) Fixed Income



(o) Tax

(p) ForeignEM

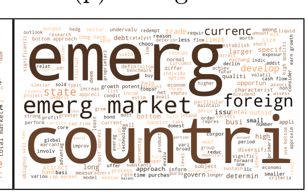


Figure 6: **TNA by SPG over time:** Panel 1 displays that cumulative TNA managed by funds in the different SPGs between January 2000 and December 2017. Panel 2 displays the number of funds assigned to the different SPGs between January 2000 and December 2017.

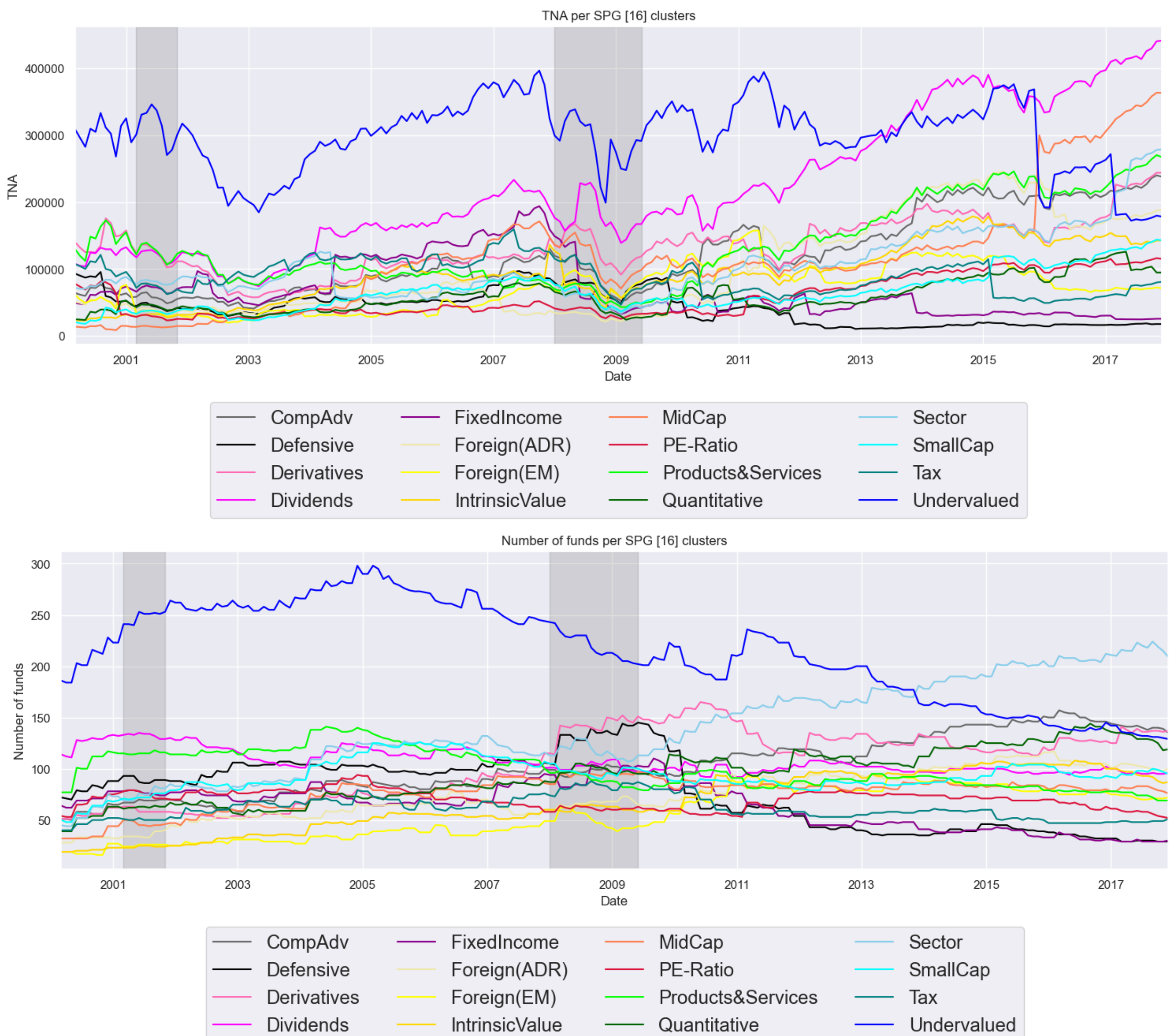


Figure 7: **Strategy clusters frequency and assignments per fund – [16]:** Panel 1 shows the number of fund/month observations assigned to each of the SPGs estimated using K-Means with [16] clusters. The number next to each bar indicates the unique number of funds that, at some point of their lives, are assigned each SPG. Panel 2 shows the number of SPG assignments that unique funds receive throughout their lives.

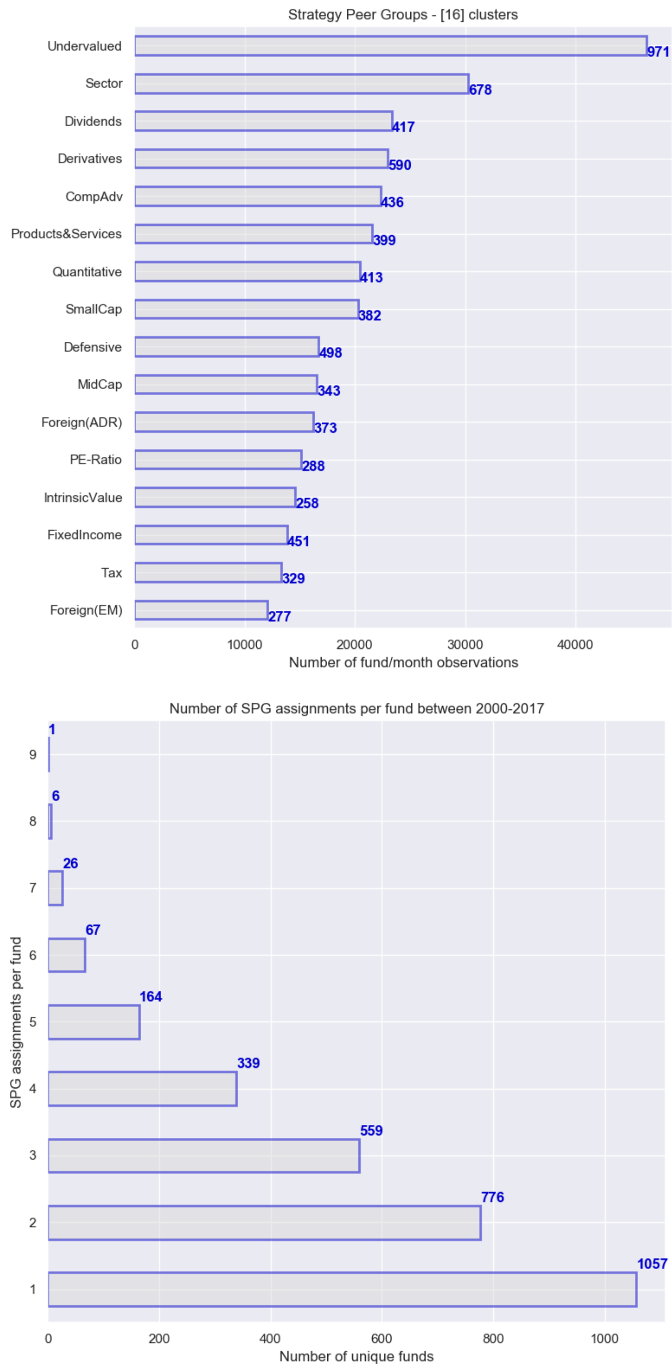


Figure 8: **TNA by investor type:** Panel 1 displays the overall TNA managed by funds in our sample by retail, retirement and institutional share classes. Panel 2 displays the same quantities in percentage of the total TNA each month.

Figure 9: **Net flows by investor type:** Panel 1 displays aggregate net flows by month and investor type (retail, retirement and institutional). Panel 2 displays the month-on-month growth in aggregate net flows by investor type.

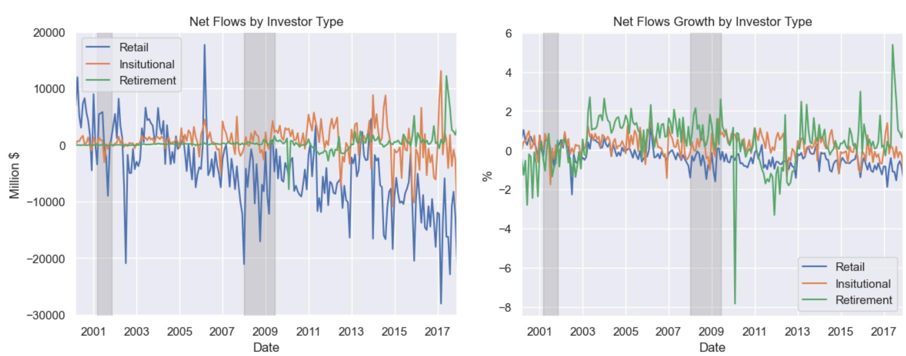


Figure 10: **TNA by investor type and SPG:** Examples of the evolution of the cumulative TNA managed by funds in a given SPG, broken down by total retail, institutional and retirement capital. The provided examples are: “Competitive Advantage” (Panel 1), “Quantitative” (Panel 2), “Tax” (Panel 3) and “Dividends” (Panel 4).

Table 1: **Summary Statistics:** Variables: fund size measured with Total Net Assets in millions of dollars ($TNA(M\$)$), funds age measured in months from inception ($Age(months)$), funds expenses ratio winsorized at the 0.1% level ($ExpenseRatio$), funds turnover ratio winsorized at the 0.1% level ($TurnoverRatio$), cash holdings in percentage of total net assets ($Cash(\%)$) and monthly returns ($Returns$). For each variable this table displays the number of available observations ($count$), the mean ($mean$), the standard deviation (sd), the minimum (min) and maximum (max) values, and the 25^th ($p25$), 50^th ($p50$) and 75^th ($p75$) percentiles.

	count	mean	sd	min	p25	p50	p75	max
TNA (M$)	320750	1187.66	4277.64	5.00	68.30	246.00	907.40	177462.59
Age (months)	320738	178.32	157.35	3.00	79.00	138.00	220.00	1121.00
ExpenseRatio	319363	1.22	0.42	0.11	0.98	1.18	1.43	4.43
TurnoverRatio	311399	80.24	75.81	1.00	33.00	61.00	102.00	806.00
FlowGrowth (%)	320507	0.14	7.75	-45.51	-1.50	-0.45	0.74	139.58
FlowVol	313894	2021.79	5069.23	4.56	154.39	532.64	1772.13	79658.24
Cash (%)	298489	3.06	4.68	-21.30	0.45	1.88	4.11	51.24

Table 2: **Summary Statistics by SPG:** Dependent Variables: the natural logarithm of funds age ($ln(Age)$) and TNA ($ln(TNA)$), winsorized expense ($Expenses$) and turnover ($Turnover$) ratios. Regressions are run separately for each SPG; the main independent variables is dummy with value of 1 for funds belonging to the SPG of interest. They include month and fund family fixed effects and demeaned control variables (omitted for brevity): log of fund age and TNA, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility (excluding the dependent variable from each regression). Standard errors are two-way clustered by fund and month.

	ln(TNA)	ln(Age)	Expenses	Turnover
Undervalued	0.0127	0.0631***	-0.00325	-1.191
	(0.35)	(2.71)	(-0.29)	(-0.61)
Sector	-0.0649	-0.0486	-0.0213*	7.382**
	(-1.27)	(-1.63)	(-1.76)	(2.48)
Dividends	0.117*	-0.00204	-0.0572***	-14.39***
	(1.85)	(-0.05)	(-3.68)	(-5.88)
Derivatives	-0.0209	-0.0358	-0.0366***	13.62***
	(-0.39)	(-1.04)	(-2.75)	(4.42)
CompAdv	0.0354	0.0227	0.0416***	-2.349
	(0.56)	(0.66)	(2.65)	(-0.85)
ProductsServices	0.0720	0.0705**	0.0716***	4.361
	(1.19)	(2.07)	(4.21)	(1.40)
Quantitative	-0.208***	-0.0915***	-0.0615***	22.69***
	(-3.33)	(-2.66)	(-3.64)	(5.29)
SmallCap	0.0345	-0.112***	0.0657***	-2.284
	(0.55)	(-3.48)	(4.17)	(-0.77)
Defensive	0.0663	0.0162	-0.0241*	-8.686**
	(1.18)	(0.46)	(-1.75)	(-2.38)
MidCap	-0.00108	-0.0270	0.0385**	4.844
	(-0.01)	(-0.77)	(2.14)	(1.63)
Foreign_ADR	-0.110*	-0.0219	-0.0450***	-3.580
	(-1.77)	(-0.58)	(-2.65)	(-1.26)
PE_Ratio	-0.0156	0.0760**	0.00876	-5.448
	(-0.21)	(2.02)	(0.51)	(-1.47)
IntrinsicValue	0.129	0.00908	0.0415***	-15.89***
	(1.57)	(0.17)	(2.81)	(-6.15)
FixedIncome	0.0376	0.0686	-0.0145	1.667
	(0.64)	(1.50)	(-0.85)	(0.49)
Tax	-0.0603	-0.0566	0.00194	-8.457**
	(-0.68)	(-1.31)	(0.10)	(-2.01)
Foreign_EM	-0.0382	0.0758*	0.00431	-5.762*
	(-0.41)	(1.68)	(0.21)	(-1.86)
Overall	5.542***	4.899***	1.226***	81.12***
	(324.74)	(472.28)	(278.77)	(99.07)
Obs	286,524	286,524	286,524	286,524

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 3: **Similarities in Returns and Holdings:** Comparison of differences in raw returns and in holdings dispersion computed relative to funds within the same SPG against dispersion computed relative to funds outside of the SPG. Return difference is computed as the absolute difference in raw returns between each funds and the average fund in its SPG or the average fund outside its SPG. Dispersion is computed at the fund level, as the natural logarithm of the sum of squared differences between the weights as a percentage of TNA ($ln(WDisp)$) of each fund and the average across all funds in the same SPG or across all other SPGs (see section 4.4.3). The differences are estimated controlling for fund age, log assets, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility, average stock market capitalization, average stock book-to-market ratio, average one-year past return, and month as well as fund family fixed effects. All independent variables are demeaned. Standard errors are two-way clustered by fund and month.

	(1) Within Mandate	(2) Outside Mandate	(3) Difference
mret_diff	1.394***	1.448***	-0.0546***
	(198.70)	(215.49)	(-19.13)
ln_disp	-4.751***	-4.334***	-0.417***
	(-437.29)	(-349.71)	(-51.19)
Controls	Yes	Yes	Yes
FE	Family+Month	Family+Month	Family+Month
Cluster	Fund+Month	Fund+Month	Fund+Month
R2	0.231	0.284	0.163
Obs	286524	286524	286524

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 4: **Risk-Adjusted Performance by SPG:** Dependent Variables: 6-factor alpha and value added. Regressions are run separately for each SPG; the main independent variables is dummy with value of 1 for funds belonging to the SPG of interest. They include month and fund family fixed effects and demeaned control variables (omitted for brevity): log of fund age and TNA, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility, average stock market capitalization, average stock book-to-market ratio, average one-year past return. Standard errors are two-way clustered by fund and month.

	Alpha	ValueAdded
Undervalued	0.00543	17.19
	(0.52)	(1.12)
Sector	-0.0187*	-3.860
	(-1.79)	(-0.29)
Dividends	-0.0529*	-76.99**
	(-1.88)	(-2.17)
Derivatives	0.00396	14.67
	(0.34)	(0.72)
CompAdv	0.0372*	-2.637
	(1.94)	(-0.12)
ProductsServices	0.0224	21.34
	(1.14)	(1.17)
Quantitative	-0.0177	-29.84**
	(-1.01)	(-2.23)
SmallCap	0.0265	11.74
	(1.38)	(0.75)
Defensive	-0.00459	17.95
	(-0.29)	(1.20)
MidCap	0.00725	11.21
	(0.27)	(0.45)
Foreign_ADR	0.00718	17.59
	(0.49)	(0.94)
PE_Ratio	-0.00966	2.348
	(-0.55)	(0.12)
IntrinsicValue	0.00844	4.960
	(0.44)	(0.20)
FixedIncome	-0.0149	-27.04
	(-0.89)	(-1.12)
Tax	0.00203	16.43
	(0.14)	(0.74)
Foreign_EM	-0.00296	8.705
	(-0.14)	(0.32)
Overall	-0.0938***	-70.71***
	(-74.00)	(-35.68)
Obs	267,294	267,294

t statistics in parentheses; * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 5: **Risk-Adjusted Performance by SPG cross-sectional:** Dependent Variables: cumulative 6-factor value added across each fund's months of existence. Regressions are run separately for each SPG; the main independent variables is dummy with value of 1 for funds belonging to the SPG of interest. They include fund family and month fixed effects. Standard errors are clustered by fund and month.

	Value Added FF6					
	3m		12m		24m	
Undervalued	50.82	48.46*	-81.57	-22.81	-91.77	-28.42
	(1.21)	(1.68)	(-0.98)	(-0.66)	(-1.15)	(-0.84)
Sector	7.450	-1.150	13.38	5.323	19.38	7.178
	(0.38)	(-0.03)	(0.60)	(0.14)	(0.83)	(0.19)
Dividends	-158.8***	-135.9***	-179.3***	-164.3***	-146.8***	-139.6***
	(-2.60)	(-3.40)	(-3.25)	(-3.45)	(-2.89)	(-2.98)
Derivatives	-46.09	-50.46	30.39	-41.00	38.82	-37.42
	(-0.70)	(-1.30)	(0.71)	(-0.89)	(0.76)	(-0.83)
CompAdv	27.46	35.61	64.57	69.95	43.47	60.01
	(0.75)	(0.87)	(1.33)	(1.43)	(1.12)	(1.25)
ProductsServices	26.80	79.71*	70.12	98.90**	49.81	67.97
	(0.97)	(1.91)	(1.29)	(1.99)	(1.16)	(1.39)
Quantitative	-8.344	8.384	-30.89	-4.168	-29.18	-6.065
	(-0.31)	(0.22)	(-1.05)	(-0.09)	(-0.99)	(-0.13)
SmallCap	-17.78	11.47	9.831	31.18	17.00	39.57
	(-0.80)	(0.28)	(0.47)	(0.63)	(0.79)	(0.82)
Defensive	-27.58	-4.872	-11.50	7.120	-6.110	18.41
	(-1.17)	(-0.10)	(-0.49)	(0.12)	(-0.29)	(0.32)
MidCap	114.4	111.1**	139.1*	154.1***	118.0	144.9***
	(1.48)	(2.36)	(1.88)	(2.75)	(1.56)	(2.63)
Foreign_ADR	12.79	-19.49	57.13	-18.11	63.23	-30.63
	(0.51)	(-0.44)	(1.40)	(-0.34)	(1.29)	(-0.59)
PE_Ratio	18.55	1.658	5.742	8.882	20.74	16.51
	(0.60)	(0.03)	(0.20)	(0.15)	(0.74)	(0.28)
IntrinsicValue	-94.22	-137.4***	-57.00	-84.52	-70.91	-90.87
	(-1.38)	(-2.79)	(-1.28)	(-1.44)	(-1.47)	(-1.58)
FixedIncome	-3.644	-13.80	-5.013	-10.68	-7.255	-0.599
	(-0.10)	(-0.24)	(-0.13)	(-0.16)	(-0.18)	(-0.01)
Tax	-12.86	-56.31	9.914	-71.95	10.67	-71.08
	(-0.25)	(-1.04)	(0.18)	(-1.12)	(0.19)	(-1.12)
Foreign_EM	103.5	73.26	126.6	109.3	144.6*	139.0**
	(1.29)	(1.31)	(1.34)	(1.64)	(1.68)	(2.12)
Overall	-73.26***	-66.47***	-66.55***	-59.66***	-75.50***	-70.03***
	(-3.72e+16)	(-6.55)	(-3.96e+16)	(-4.93)	(-2.09e+16)	(-5.89)
FE	Family	No	Family	No	Family	No
Cluster	Fund	Fund	Fund	Fund	Fund	Fund
Obs	2,740	2,995	2,740	2,740	2,995	2,740

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 6: **Risk factor exposures by SPG:** Dependent Variables: funds' loadings on the Fama-French 5 factors plus momentum, computed using the prior 12 months of daily returns. Regressions are run separately for each SPG and factor; the main independent variables is a dummy with value of 1 for funds belonging to the SPG of interest. Only the coefficients with respect to the dummy variable are displayed for each regression. Regressions include fund family and month fixed effects. Standard errors are clustered by fund and month.

	Factor loading (beta)					
	Market	SMB	HML	MOM	RMW	CMA
Undervalued	0.00286	-0.0446***	-0.00169	0.00124	0.000636	-0.0194***
	(0.62)	(-3.50)	(-0.21)	(0.34)	(0.10)	(-3.16)
Sector	0.00239	-0.00989	0.00480	0.00470	-0.00383	-0.00511
	(0.46)	(-0.61)	(0.51)	(1.04)	(-0.52)	(-0.69)
Dividends	-0.0230***	-0.167***	0.0986***	-0.0459***	0.113***	0.115***
	(-2.97)	(-10.03)	(9.72)	(-8.35)	(12.92)	(11.85)
Derivatives	-0.0216***	-0.0325**	-0.00121	-0.00557	-0.00728	0.000676
	(-2.89)	(-2.02)	(-0.13)	(-1.06)	(-0.84)	(0.08)
CompAdv	-0.00339	-0.0188	-0.107***	0.0202***	-0.0635***	-0.0689***
	(-0.49)	(-0.88)	(-9.52)	(3.39)	(-6.08)	(-6.86)
ProductsServices	0.0202***	0.0719***	-0.0771***	0.0276***	-0.0872***	-0.0347***
	(2.60)	(3.62)	(-6.93)	(4.42)	(-8.33)	(-3.93)
Quantitative	0.00979	-0.0352	0.0324***	0.0179***	0.0592***	0.0208**
	(1.57)	(-1.50)	(2.64)	(2.80)	(6.77)	(2.33)
SmallCap	0.0189***	0.392***	0.0360***	0.00579	-0.0140	-0.0164**
	(2.83)	(20.75)	(2.84)	(1.08)	(-1.35)	(-2.39)
Defensive	-0.0168**	-0.0151	0.00461	-0.0118**	0.00979	0.0114
	(-2.52)	(-1.02)	(0.46)	(-2.01)	(1.10)	(1.18)
MidCap	0.0119*	-0.00511	-0.0287**	0.0143**	-0.0175*	0.0170*
	(1.69)	(-0.33)	(-2.41)	(2.56)	(-1.71)	(1.79)
Foreign_ADR	0.00546	-0.0190	-0.0236*	0.0132**	-0.0245**	-0.0292**
	(0.72)	(-0.89)	(-1.79)	(2.47)	(-2.21)	(-2.25)
PE_Ratio	0.00645	-0.0212	0.0376***	-0.00517	0.0311***	0.0204*
	(0.92)	(-0.91)	(2.80)	(-0.67)	(2.62)	(1.91)
IntrinsicValue	-0.00806	-0.0470**	0.0708***	-0.0417***	0.0295**	0.0285**
	(-1.01)	(-2.13)	(5.59)	(-6.38)	(2.55)	(2.53)
FixedIncome	-0.0146*	-0.0250	0.00561	-0.0214***	-0.00164	0.0224**
	(-1.67)	(-1.49)	(0.55)	(-4.16)	(-0.15)	(2.23)
Tax	0.00433	-0.0513*	-0.0149	0.000586	0.0122	-0.00608
	(0.64)	(-1.93)	(-1.27)	(0.11)	(1.12)	(-0.67)
Foreign_EM	-0.00223	0.0370	-0.0288*	0.0120*	-0.0413***	-0.0397***
	(-0.27)	(1.38)	(-1.92)	(1.72)	(-3.63)	(-3.13)
Overall	0.980***	0.222***	-0.00706**	0.0308***	-0.0545***	-0.0322***
	(970.71)	(34.83)	(-2.04)	(21.38)	(-20.45)	(-13.33)
Obs	284,528	284,528	284,528	284,528	284,528	284,528

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 7: **Dispersion in Characteristics of Stocks Held:** Comparison of holdings dispersion measures computed relative to funds within the same SPG against dispersion computed relative to funds outside of the SPG. Dispersion is computed at the fund level, as the natural logarithm of the sum of squared differences between characteristics ($ln(CDisp)$) vectors of each fund and the average across all funds in the same SPG or across all other SPGs (see section 4.4.3). We also compute Characteristics Dispersion based on subsets of characteristics: priced ($ln(CDisp\_``Priced'')$), non-priced ($ln(CDisp\_``NotPriced'')$), items on the asset side of the balance sheet ($ln(CDisp\_Assets)$), items on the liabilities side of the balance sheet ($ln(CDisp\_Liabilities)$), items on the income statement ($ln(CDisp\_Income)$), security market variables ($ln(CDisp\_Market)$), information availability ($ln(CDisp\_Information)$), sentiment ($ln(CDisp\_Sentiment)$), Fama-French 48 industry weights ($ln(CDisp\_Industries)$), and characteristics of the funds' strategy ($ln(CDisp\_Strategy)$). The differences are estimated controlling for fund age, log assets, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility, and month as well as fund family fixed effects. All independent variables are demeaned. Standard errors are two-way clustered by fund and month.

	(1) Within Mandate	(2) Outside Mandate	(3) Difference
ln(Cdisp_"Priced")	-0.850***	-0.627***	-0.223***
	(-52.97)	(-39.74)	(-24.67)
ln(Cdisp_"NotPriced")	-1.116***	-1.040***	-0.0755***
	(-106.04)	(-101.04)	(-17.83)
ln(Cdisp_Strategy)	-0.481***	-0.423***	-0.0572***
	(-34.87)	(-31.50)	(-8.41)
ln(Cdisp_Information)	-1.399***	-1.141***	-0.259***
	(-70.32)	(-52.96)	(-18.97)
ln(Cdisp_Sentiment)	-2.808***	-2.734***	-0.0741***
	(-286.51)	(-289.82)	(-22.31)
ln(Cdisp_Industries)	-3.145***	-3.091***	-0.0540***
	(-313.80)	(-318.37)	(-20.44)
ln(Cdisp_Assets)	-1.179***	-1.089***	-0.0903***
	(-86.45)	(-82.05)	(-17.46)
ln(Cdisp_Liabilities)	-1.782***	-1.688***	-0.0941***
	(-190.20)	(-187.43)	(-21.47)
ln(Cdisp_Income)	-5.149***	-5.033***	-0.116***
	(-313.67)	(-410.12)	(-12.71)
ln(Cdisp_Markets)	-3.351***	-3.245***	-0.106***
	(-263.65)	(-251.28)	(-17.05)
Controls	Yes	Yes	Yes
FE	Family+Month	Family+Month	Family+Month
Cluster	Fund+Month	Fund+Month	Fund+Month
R2	0.361	0.382	0.0355
Obs	286,524	286,524	286,524

t statistics in parentheses; * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 8: **Industry Exposure by SPG:** Differences between the average industry exposure within a particular SPG and the average exposure across all other SPGs. Differences are estimated in separate regressions of asset-weighted average stock characteristics at the fund level on dummy variables for each SPG (see section 4.5), controlling for the natural logarithm of fund age and TNA, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility and average market cap, book-to-market ratio and one-year past return. Industries are the Fama-French 48 – a selected sample is displayed for brevity. Regressions include month and fund family fixed effects. Standard errors are two-way clustered by fund and month.

	Util	Telcm	Hshld	Oil	Banks	Tech	Drugs
Undervalued	-0.33***	-0.10	0.04	-0.44*	-0.05	0.00	0.02
	(-3.53)	(-0.69)	(0.95)	(-1.84)	(-0.16)	(0.89)	(0.08)
Sector	0.02	-0.21	-0.06	0.39	0.63*	0.00	-0.13
	(0.16)	(-1.48)	(-0.94)	(1.15)	(1.69)	(1.56)	(-0.53)
Dividends	1.93***	1.85***	0.16**	0.50	4.55***	-0.02***	-0.58*
	(10.13)	(8.45)	(2.41)	(1.39)	(8.62)	(-6.98)	(-1.86)
Derivatives	0.01	0.36**	0.03	-0.32	-0.07	0.00	0.25
	(0.06)	(2.14)	(0.43)	(-1.10)	(-0.19)	(1.45)	(0.89)
CompAdv	-0.97***	-0.54***	0.25***	-1.38***	-1.52***	0.01**	1.08***
	(-9.45)	(-3.48)	(3.09)	(-4.09)	(-3.92)	(2.57)	(2.91)
ProductsServices	-0.92***	-0.45**	-0.05	-0.39	-3.23***	0.01***	1.39***
	(-8.25)	(-1.99)	(-0.63)	(-1.01)	(-7.31)	(3.86)	(3.12)
Quantitative	0.33**	0.20	-0.08	0.40	0.71	0.00	-0.26
	(2.03)	(1.20)	(-1.50)	(1.14)	(1.61)	(1.45)	(-0.90)
SmallCap	-0.29**	-1.61***	-0.12*	2.69***	-1.99***	-0.01***	0.69
	(-2.51)	(-9.55)	(-1.97)	(4.79)	(-4.22)	(-2.83)	(1.21)
Defensive	0.06	0.07	0.01	-0.03	0.96*	-0.00	-0.38
	(0.43)	(0.32)	(0.17)	(-0.08)	(1.82)	(-1.21)	(-1.15)
MidCap	0.47***	-0.80***	-0.10	-1.68***	-3.22***	-0.00	-1.88***
	(2.62)	(-4.03)	(-1.39)	(-4.23)	(-7.99)	(-0.68)	(-6.34)
Foreign_ADR	-0.38*	-0.08	-0.10	0.47	0.28	0.00	0.58*
	(-1.70)	(-0.35)	(-1.24)	(1.02)	(0.60)	(0.37)	(1.87)
PE_Ratio	0.17	0.07	0.02	-0.03	1.60***	-0.00	-0.84**
	(1.06)	(0.36)	(0.25)	(-0.07)	(2.71)	(-1.47)	(-2.55)
IntrinsicValue	0.24	1.63***	-0.09	-1.42***	2.05***	-0.00	-0.77**
	(1.18)	(3.63)	(-0.95)	(-3.03)	(3.52)	(-0.37)	(-2.30)
FixedIncome	0.28**	0.28	0.01	0.28	0.48	-0.00	0.30
	(2.09)	(1.10)	(0.14)	(0.84)	(1.02)	(-0.42)	(0.64)
Tax	-0.37*	0.08	0.08	1.63***	0.11	0.00	0.24
	(-1.79)	(0.36)	(0.96)	(2.64)	(0.21)	(0.85)	(0.49)
Foreign_EM	-0.02	-0.10	-0.16*	-0.56	-0.80	0.00	-0.32
	(-0.06)	(-0.32)	(-1.70)	(-0.99)	(-1.45)	(1.19)	(-0.95)
Obs	286,524	286,524	286,524	286,524	286,524	286,524	286,524

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 9: **Priced Characteristics by SPG:** This table reports differences between priced stock characteristics within a particular mandate and the average across all other mandates. These differences are estimated in separate regressions of asset-weighted average stock characteristics at the fund level on dummy variables for each mandate (see section 4.5), controlling for fund age, log assets, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility, and month as well as fund family fixed effects. Standard errors are two-way clustered by fund and month.

	MktBeta	MktCap	BookToMkt	Momentum	Investment	Profitability
Undervalued	-0.297	9.598***	-0.0687	-1.624*	-0.347	-0.148
	(-0.63)	(3.43)	(-0.89)	(-1.96)	(-0.60)	(-0.18)
Sector	0.857	3.215	-0.0578	-0.483	-0.445	-0.424
	(1.46)	(0.97)	(-0.55)	(-0.53)	(-0.64)	(-0.46)
Dividends	-2.614***	35.98***	0.919***	-10.66***	-1.839***	-8.244***
	(-4.03)	(9.91)	(6.67)	(-6.80)	(-3.02)	(-9.13)
Derivatives	0.525	5.110	0.116	0.659	1.811*	-1.843
	(0.68)	(1.47)	(1.11)	(0.49)	(1.79)	(-1.42)
CompAdv	-0.150	7.352	-0.866***	3.154***	-0.139	8.753***
	(-0.23)	(1.62)	(-6.52)	(2.62)	(-0.19)	(7.69)
ProductsServices	0.273	-17.21***	-0.382***	8.224***	2.568***	5.485***
	(0.34)	(-4.09)	(-3.06)	(4.69)	(2.67)	(4.40)
Quantitative	0.146	8.047*	0.297*	1.152	-2.217***	-0.401
	(0.22)	(1.77)	(1.75)	(0.80)	(-4.58)	(-0.35)
SmallCap	2.576***	-72.05***	0.169	7.070***	5.542***	-2.700*
	(3.02)	(-19.26)	(0.85)	(4.27)	(4.93)	(-1.74)
Defensive	-1.014	1.290	-0.0336	-2.127	-0.390	0.208
	(-1.41)	(0.41)	(-0.38)	(-1.64)	(-0.52)	(0.20)
MidCap	-0.0107	-14.37***	-0.206	1.411	-1.206**	3.600***
	(-0.02)	(-4.27)	(-1.01)	(1.29)	(-2.45)	(3.17)
Foreign_ADR	1.467*	4.024	-0.118	3.076**	0.215	1.805
	(1.81)	(0.92)	(-0.66)	(2.45)	(0.35)	(1.36)
PE_Ratio	-0.831	4.653	-0.00850	-1.691	-1.606**	-1.923
	(-1.01)	(0.98)	(-0.06)	(-1.00)	(-2.30)	(-1.32)
IntrinsicValue	-1.606*	8.970*	0.318**	-7.638***	-2.983***	-3.096**
	(-1.66)	(1.96)	(2.33)	(-7.06)	(-4.05)	(-2.43)
FixedIncome	-0.754	5.287	0.0866	-3.142***	-0.370	-2.747**
	(-0.85)	(1.34)	(0.81)	(-2.83)	(-0.44)	(-1.99)
Tax	2.496***	15.74***	0.202	0.225	-0.686	-0.243
	(2.83)	(2.85)	(1.09)	(0.14)	(-0.65)	(-0.16)
Foreign_EM	-1.459*	-10.25*	-0.290**	2.355*	0.754	1.748
	(-1.68)	(-1.96)	(-2.11)	(1.76)	(1.47)	(1.19)
Obs	286,524	286,524	286,524	286,524	286,524	286,524

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 10: **Non-Priced Characteristics by SPG:** This table reports differences between various average non-priced stock characteristics within a particular mandate and the characteristics' average across all other mandates. These differences are estimated in separate regressions of asset-weighted average stock chracteristics at the fund level on dummy variables for each mandate (see section 4.5), controlling for fund age, log assets, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility, and month as well as fund family fixed effects. Standard errors are two-way clustered by fund and month.

	CurrAssets	Inventories	NonPerfAst	PPandE	Intangibles	AssetGrowth	Cash
Undervalued	-0.58	0.84	-0.43	-1.51	0.37	0.07	-0.59
	(-0.47)	(1.14)	(-1.41)	(-1.55)	(0.33)	(0.14)	(-0.50)
Sector	-1.88	-0.31	0.79*	0.66	0.46	0.04	-1.83
	(-1.35)	(-0.48)	(1.73)	(0.55)	(0.30)	(0.07)	(-1.36)
Dividends	-20.92***	-2.94***	1.27***	4.19***	-6.02***	-6.28***	-15.97***
	(-13.60)	(-3.65)	(3.13)	(3.20)	(-4.55)	(-8.21)	(-11.07)
Derivatives	-0.51	0.05	0.10	-0.61	-1.98	0.12	1.35
	(-0.31)	(0.06)	(0.22)	(-0.53)	(-1.27)	(0.19)	(0.73)
CompAdv	9.71***	-0.03	-2.05***	-6.25***	9.17***	2.05***	9.11***
	(5.30)	(-0.03)	(-4.04)	(-4.62)	(5.16)	(3.28)	(4.51)
ProductsServices	15.43***	0.04	-1.46***	-1.87	5.55***	4.45***	13.02***
	(8.13)	(0.04)	(-3.71)	(-1.17)	(3.44)	(5.22)	(6.20)
Quantitative	-5.30***	2.62***	0.46	0.18	-8.90***	-2.53***	-4.52***
	(-3.01)	(3.30)	(0.87)	(0.15)	(-5.14)	(-3.83)	(-2.96)
SmallCap	15.45***	0.65	1.92***	5.93***	-10.30***	4.08***	11.80***
	(7.06)	(0.62)	(3.28)	(3.50)	(-5.64)	(4.40)	(4.70)
Defensive	-1.00	0.85	0.53	-2.20	1.21	-0.59	-0.83
	(-0.57)	(0.70)	(0.69)	(-1.40)	(0.76)	(-0.78)	(-0.48)
MidCap	1.67	0.98	-2.17***	3.18**	12.38***	0.19	-5.12***
	(1.01)	(1.19)	(-4.59)	(2.04)	(6.21)	(0.31)	(-3.35)
Foreign_ADR	2.36	-0.56	-0.43	-0.97	1.30	1.83**	3.86**
	(1.25)	(-0.66)	(-0.75)	(-0.60)	(0.75)	(2.57)	(2.10)
PE_Ratio	-5.48**	3.39***	0.68	-1.37	-2.57	-0.92	-5.37***
	(-2.52)	(3.20)	(1.24)	(-0.83)	(-1.41)	(-1.09)	(-2.78)
IntrinsicValue	-8.28***	-1.88*	1.33*	-4.17**	5.22**	-3.55***	-6.06***
	(-4.17)	(-1.72)	(1.74)	(-2.11)	(2.12)	(-4.92)	(-2.86)
FixedIncome	-3.84**	-2.30**	1.33**	2.83**	-3.99**	-1.03	-3.00
	(-2.05)	(-2.43)	(2.26)	(1.98)	(-2.38)	(-1.37)	(-1.61)
Tax	-4.56**	-1.79	-0.85**	3.58	-5.31***	-0.68	-2.26
	(-2.15)	(-1.52)	(-1.97)	(1.65)	(-2.88)	(-0.87)	(-1.07)
Foreign_EM	5.23**	-1.62	-0.63	-0.98	5.15**	2.14***	5.19**
	(2.34)	(-1.34)	(-0.95)	(-0.53)	(2.20)	(2.82)	(2.54)
Obs	286,524	286,524	286,524	286,524	286,524	286,524	286,524

t statistics in parentheses; * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 11: **(Cont'd) Non-Priced Characteristics by SPG:** This table reports differences between various average non-priced stock characteristics within a particular mandate and the characteristics' average across all other mandates. These differences are estimated in separate regressions of asset-weighted average stock chracteristics at the fund level on dummy variables for each mandate (see section 4.5), controlling for fund age, log assets, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility, and month as well as fund family fixed effects. Standard errors are two-way clustered by fund and month.

	CurrLiab	Leverage	DeferTax	LTDebt	OperatAct	EarnGrowth	R_and_D
Undervalued	0.78	0.04	3.96**	-2.12**	-1.53	-0.30	0.63
	(1.02)	(0.12)	(2.41)	(-2.52)	(-1.10)	(-1.47)	(0.47)
Sector	-0.50	0.15	-0.20	0.01	1.39*	-0.13	-2.07*
	(-0.56)	(0.37)	(-0.11)	(0.01)	(1.97)	(-0.59)	(-1.76)
Dividends	-6.39***	2.64***	17.41***	3.25***	2.76***	0.19	-3.75***
	(-6.81)	(6.05)	(8.21)	(3.11)	(4.06)	(0.80)	(-3.73)
Derivatives	-0.05	-0.23	3.88*	-0.52	-4.23	0.06	0.64
	(-0.05)	(-0.59)	(1.84)	(-0.56)	(-1.28)	(0.26)	(0.43)
CompAdv	4.40***	-1.37***	0.98	-6.97***	0.41	0.57	0.81
	(4.78)	(-2.75)	(0.43)	(-5.96)	(0.26)	(1.64)	(0.56)
ProductsServices	4.93***	-2.90***	-9.29***	-2.94**	-2.44	-0.07	6.70**
	(4.58)	(-6.06)	(-4.56)	(-2.23)	(-1.54)	(-0.18)	(2.47)
Quantitative	0.53	1.70***	5.81***	-0.02	3.92***	-0.07	-3.20***
	(0.49)	(3.95)	(2.79)	(-0.02)	(3.69)	(-0.26)	(-3.60)
SmallCap	-2.11*	-2.47***	-20.93***	3.35**	-4.82**	-0.65	6.79***
	(-1.81)	(-4.91)	(-9.96)	(2.11)	(-2.43)	(-1.46)	(2.64)
Defensive	-0.43	0.51	-4.61*	-0.92	-0.71	-0.02	0.33
	(-0.38)	(1.13)	(-1.95)	(-0.90)	(-0.44)	(-0.06)	(0.15)
MidCap	3.40***	-0.32	-10.53***	5.68***	1.66**	0.35	-2.19**
	(3.10)	(-0.63)	(-5.09)	(4.83)	(2.34)	(0.87)	(-2.43)
Foreign_ADR	0.51	-0.59	2.13	-0.85	-0.00	0.23	0.80
	(0.39)	(-1.02)	(0.81)	(-0.67)	(-0.00)	(0.80)	(0.45)
PE_Ratio	-2.71**	1.46***	2.59	0.67	0.80	-0.12	-1.22
	(-2.02)	(2.63)	(0.95)	(0.56)	(1.12)	(-0.34)	(-0.70)
IntrinsicValue	-2.24*	2.37***	4.03	2.96*	5.44*	-0.39	-5.14***
	(-1.75)	(3.13)	(1.60)	(1.71)	(1.73)	(-0.64)	(-3.22)
FixedIncome	-1.33	0.39	-1.00	3.69**	0.66	-0.22	-0.53
	(-1.15)	(0.81)	(-0.44)	(2.18)	(0.83)	(-0.73)	(-0.34)
Tax	-1.43	0.26	7.87**	-1.76	2.20	0.32	-1.23
	(-1.05)	(0.48)	(2.50)	(-1.17)	(1.55)	(0.94)	(-0.46)
Foreign_EM	0.80	-0.65	-6.86**	1.15	-1.97*	0.95*	-0.08
	(0.58)	(-1.17)	(-2.32)	(0.82)	(-1.79)	(1.83)	(-0.08)
Obs	286,524	286,524	286,524	286,524	286,524	286,524	286,524

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 12: **(Cont'd) Non-Priced Characteristics by SPG:** This table reports differences between various average non-priced stock characteristics within a particular mandate and the characteristics' average across all other mandates. These differences are estimated in separate regressions of asset-weighted average stock chracteristics at the fund level on dummy variables for each mandate (see section 4.5), controlling for fund age, log assets, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility, and month as well as fund family fixed effects. Standard errors are two-way clustered by fund and month.

	DivYield	Issuance	Repurchase	Illiquidity	Age
Undervalued	-0.32***	0.14	2.69***	0.00	6.24**
	(-2.74)	(0.43)	(2.81)	(0.24)	(2.47)
Sector	-0.25*	-0.20	-0.31	-0.00	2.13
	(-1.73)	(-0.48)	(-0.23)	(-0.06)	(0.65)
Dividends	3.60***	3.04***	4.90***	0.01***	62.11***
	(10.45)	(6.80)	(3.80)	(2.70)	(15.68)
Derivatives	0.32	-0.31	0.91	-0.00	5.30
	(1.44)	(-0.65)	(0.65)	(-1.59)	(1.51)
CompAdv	-1.04***	-1.19**	-3.84**	-0.00	-14.15***
	(-5.71)	(-2.40)	(-2.45)	(-0.88)	(-3.52)
ProductsServices	-0.73***	-1.57***	-6.56***	-0.01***	-25.29***
	(-3.46)	(-3.78)	(-4.62)	(-3.39)	(-6.59)
Quantitative	0.27	2.21***	11.41***	0.01**	16.30***
	(1.55)	(3.59)	(6.84)	(2.07)	(3.88)
SmallCap	-1.19***	-2.74***	-20.43***	-0.01**	-57.35***
	(-7.20)	(-4.94)	(-12.59)	(-2.39)	(-17.66)
Defensive	0.27	0.77	1.90	0.00	-0.96
	(1.44)	(1.58)	(1.52)	(0.48)	(-0.32)
MidCap	-0.67***	-1.10*	0.89	-0.00	-20.65***
	(-2.89)	(-1.75)	(0.66)	(-1.04)	(-5.95)
Foreign_ADR	-0.33	0.17	1.28	0.00	-4.36
	(-1.51)	(0.29)	(0.79)	(0.80)	(-0.91)
PE_Ratio	-0.15	0.80	1.31	0.01	9.02*
	(-0.73)	(1.46)	(0.77)	(1.28)	(1.90)
IntrinsicValue	0.41*	-0.05	9.25***	-0.00	12.62**
	(1.72)	(-0.06)	(4.48)	(-0.71)	(2.56)
FixedIncome	0.60**	-0.04	0.65	0.01*	9.07**
	(2.30)	(-0.06)	(0.50)	(1.73)	(2.22)
Tax	-0.06	1.45**	4.14**	-0.00	13.77***
	(-0.28)	(2.30)	(2.12)	(-1.61)	(2.71)
Foreign_EM	-0.65**	-1.19*	-6.19***	-0.00	-21.58***
	(-2.54)	(-1.67)	(-3.08)	(-1.38)	(-4.23)
Obs	286,524	286,524	286,524	286,524	286,524

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 13: **(Cont'd) Non-Priced Characteristics by SPG:** This table reports differences between various average non-priced stock characteristics within a particular mandate and the characteristics' average across all other mandates. These differences are estimated in separate regressions of asset-weighted average stock chracteristics at the fund level on dummy variables for each mandate (see section 4.5), controlling for fund age, log assets, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility, and month as well as fund family fixed effects. Standard errors are two-way clustered by fund and month.

	ADR_perc	ForeignInc	NumStocks	Cash_perc	CommStk_perc
Undervalued	-0.03	0.10	-5.06**	1.52	6.25**
	(-0.54)	(0.61)	(-2.04)	(0.60)	(2.40)
Sector	-0.02	0.16	-6.25***	-1.33	5.94*
	(-0.25)	(0.62)	(-2.83)	(-0.49)	(1.96)
Dividends	0.89***	0.44	-6.79**	-3.42	-24.32***
	(6.14)	(1.64)	(-2.01)	(-1.11)	(-5.27)
Derivatives	0.02	0.11	3.95	1.26	-5.11
	(0.20)	(0.45)	(0.65)	(0.32)	(-1.11)
CompAdv	-0.01	0.20	-12.70***	2.96	4.03
	(-0.12)	(0.73)	(-5.20)	(0.78)	(0.91)
ProductsServices	-0.03	-0.03	-3.16	-2.92	8.84***
	(-0.26)	(-0.10)	(-1.34)	(-0.93)	(2.61)
Quantitative	-0.54***	-1.30***	19.75***	-14.86***	20.51***
	(-6.26)	(-5.24)	(5.20)	(-5.52)	(5.84)
SmallCap	-0.45***	-1.22***	43.78***	4.97*	-4.49
	(-5.35)	(-4.39)	(5.36)	(1.71)	(-1.42)
Defensive	0.11	0.30	-3.72*	4.95	-8.94**
	(1.06)	(1.12)	(-1.86)	(1.37)	(-2.21)
MidCap	-0.31**	0.27	-4.72	-3.70	8.99***
	(-2.14)	(0.85)	(-1.26)	(-1.43)	(2.79)
Foreign_ADR	0.12	0.63**	-6.07***	-5.13*	3.78
	(0.85)	(2.07)	(-2.84)	(-1.80)	(0.98)
PE_Ratio	-0.22**	-0.25	-4.87**	0.92	5.67
	(-2.06)	(-0.81)	(-2.00)	(0.21)	(1.32)
IntrinsicValue	0.11	0.46	-11.77***	1.03	-11.87**
	(0.81)	(1.61)	(-6.45)	(0.25)	(-2.03)
FixedIncome	-0.01	-0.05	-3.78*	20.21***	-29.74***
	(-0.08)	(-0.17)	(-1.77)	(4.45)	(-4.72)
Tax	0.09	-0.13	7.84	-8.29**	11.34***
	(0.74)	(-0.45)	(0.91)	(-2.18)	(2.62)
Foreign_EM	0.44***	0.70**	-9.87***	6.38	-7.03
	(3.29)	(2.13)	(-2.81)	(1.40)	(-0.95)
Obs	286,524	286,524	286,524	274,422	274,422

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 14: **(Cont'd) Non-Priced Characteristics by SPG:** This table reports differences between various average non-priced stock characteristics within a particular mandate and the characteristics' average across all other mandates. These differences are estimated in separate regressions of asset-weighted average stock chracteristics at the fund level on dummy variables for each mandate (see section 4.5), controlling for fund age, log assets, expense ratio, turnover ratio, monthly fund flows, monthly fund flow volatility, and month as well as fund family fixed effects. Standard errors are two-way clustered by fund and month.

	NumAnalysts	NumRecc	MeanRecc	StDevRecc	NumDow	MeanDow	StDevDow
Undervalued	-1.52*	-1.09**	-0.64**	0.15	-0.43	0.57	1.54***
	(-1.82)	(-2.26)	(-2.05)	(1.14)	(-1.08)	(1.45)	(3.64)
Sector	-0.81	-0.82	-0.35	-0.01	-0.38	0.68	0.49
	(-0.77)	(-1.44)	(-1.00)	(-0.05)	(-0.87)	(1.38)	(1.00)
Dividends	-5.06***	-2.37***	3.38***	0.34	-2.40***	0.71	3.16***
	(-5.65)	(-4.12)	(6.59)	(1.39)	(-6.48)	(1.16)	(5.29)
Derivatives	-0.63	-0.35	0.28	0.11	0.49	-0.11	0.98*
	(-0.47)	(-0.51)	(0.68)	(0.79)	(0.56)	(-0.20)	(1.87)
CompAdv	-2.07*	-0.63	-2.74***	-0.06	-1.00	2.42***	1.09
	(-1.74)	(-0.83)	(-5.80)	(-0.32)	(-1.60)	(3.84)	(1.64)
ProductsServices	2.76**	1.74***	-1.59***	-0.31	1.34*	-0.21	-1.55**
	(2.48)	(2.72)	(-3.44)	(-1.59)	(1.75)	(-0.34)	(-2.25)
Quantitative	-0.61	-0.32	1.27***	-0.08	0.32	0.54	1.06
	(-0.42)	(-0.40)	(2.71)	(-0.49)	(0.46)	(0.92)	(1.59)
SmallCap	15.11***	8.04***	0.05	-0.50	5.53***	-6.39***	-9.27***
	(7.15)	(6.98)	(0.08)	(-1.61)	(6.38)	(-8.55)	(-11.34)
Defensive	-0.80	-0.58	-0.23	-0.14	0.35	0.39	-0.55
	(-0.79)	(-0.90)	(-0.56)	(-0.78)	(0.34)	(0.69)	(-0.95)
MidCap	-1.98	-0.73	1.37***	0.08	-1.42***	-0.14	-1.60***
	(-1.54)	(-0.85)	(2.65)	(0.39)	(-2.77)	(-0.24)	(-2.68)
Foreign_ADR	1.04	0.74	-0.70	0.26	-0.10	1.08	-0.44
	(0.56)	(0.74)	(-1.49)	(1.60)	(-0.13)	(1.55)	(-0.66)
PE_Ratio	-2.98***	-1.88***	-0.43	-0.29*	-0.53	0.00	0.40
	(-2.69)	(-3.15)	(-0.84)	(-1.77)	(-0.74)	(0.00)	(0.50)
IntrinsicValue	-3.38***	-1.91***	1.95***	0.34**	-0.54	0.22	1.27*
	(-2.97)	(-2.96)	(3.68)	(2.13)	(-1.12)	(0.29)	(1.67)
FixedIncome	-1.33	-0.53	0.52	0.06	-0.75	0.29	0.63
	(-1.41)	(-0.76)	(1.06)	(0.38)	(-1.49)	(0.57)	(0.99)
Tax	2.34	0.91	-0.14	0.04	-1.08	-0.36	3.11***
	(1.02)	(0.73)	(-0.28)	(0.16)	(-1.18)	(-0.50)	(4.16)
Foreign_EM	0.06	0.12	-1.29**	-0.04	0.34	-0.05	-1.08
	(0.05)	(0.13)	(-2.38)	(-0.17)	(0.54)	(-0.07)	(-1.35)
Obs	286,524	286,524	286,524	286,524	286,524	286,524	286,524

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 15: **Flows by SPG and Investor Type:**

	FlowGrowth	FlowGrowth	FlowGrowth
FlowGrowth_R * Dividends	-0.201**	-0.214***	-0.216***
	(-2.52)	(-2.74)	(-2.76)
FlowGrowth_I * Dividends	0.0569	0.0647	0.0596
	(1.07)	(1.23)	(1.15)
FlowGrowth_Rt * Dividends	-0.0709	-0.0650	-0.0720
	(-1.07)	(-0.98)	(-1.07)
FlowGrowth_R * Derivatives	-0.0805	-0.0672	-0.0585
	(-0.94)	(-0.80)	(-0.71)
FlowGrowth_I * Derivatives	0.0162	0.000722	0.00159
	(0.33)	(0.01)	(0.03)
FlowGrowth_Rt * Derivatives	0.134**	0.126*	0.128**
	(2.12)	(1.96)	(2.04)
FlowGrowth_R * CompAdv	0.199**	0.211***	0.192**
	(2.51)	(2.67)	(2.55)
FlowGrowth_I * CompAdv	0.0974*	0.0888*	0.102**
	(1.89)	(1.75)	(2.10)
FlowGrowth_Rt * CompAdv	-0.0671	-0.0750	-0.0724
	(-0.90)	(-1.00)	(-1.02)
FlowGrowth_R * Quantitative	-0.179**	-0.172**	-0.155*
	(-2.17)	(-2.09)	(-1.89)
FlowGrowth_I * Quantitative	-0.156**	-0.159**	-0.167**
	(-1.99)	(-2.04)	(-2.21)
FlowGrowth_Rt * Quantitative	0.187***	0.179***	0.170***
	(2.99)	(2.90)	(2.86)
FlowGrowth_R * Foreign_ADR	-0.189**	-0.174**	-0.206***
	(-2.51)	(-2.35)	(-2.74)
FlowGrowth_I * Foreign_ADR	0.0814	0.0752	0.0968
	(1.29)	(1.19)	(1.52)
FlowGrowth_Rt * Foreign_ADR	0.0197	0.0134	0.0478
	(0.28)	(0.19)	(0.68)
FlowGrowth_R * IntrinsicValue	0.0648	0.0836	0.117
	(0.61)	(0.79)	(1.15)
FlowGrowth_I * IntrinsicValue	0.0620	0.0494	0.0442
	(1.03)	(0.85)	(0.78)
FlowGrowth_Rt * IntrinsicValue	-0.163*	-0.171*	-0.141*
	(-1.82)	(-1.91)	(-1.67)
FlowGrowth_R * FixedIncome	0.148	0.128	0.143
	(1.38)	(1.21)	(1.32)
FlowGrowth_I * FixedIncome	-0.0624	-0.0580	-0.0466
	(-0.98)	(-0.95)	(-0.77)
FlowGrowth_Rt * FixedIncome	-0.141*	-0.137	-0.156*
	(-1.65)	(-1.62)	(-1.74)
FlowGrowth_R * Foreign_EM	0.0590	0.0781	0.0627
	(0.63)	(0.83)	(0.68)
FlowGrowth_I * Foreign_EM	0.121*	0.109	0.102
	(1.77)	(1.60)	(1.52)
FlowGrowth_Rt * Foreign_EM	0.195***	0.180***	0.161**
	(3.00)	(2.76)	(2.55)
FE	No	Month	Family+Month
Cluster	Fund+Month	Fund+Month	Fund+Month
R2	0.0330	0.0370	0.0637
Obs	253302	253302	253301

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 16: **Offering Likelihood by SPG:**

	D_retail	D_institutional	D_retirement
Undervalued	-0.0108	-0.00690	0.000330
	(-1.10)	(-0.58)	(0.03)
Sector	0.00496	0.0226	0.00334
	(0.34)	(1.39)	(0.30)
Dividends	0.0106	-0.0444**	-0.0236**
	(0.80)	(-2.41)	(-1.97)
Derivatives	0.0221*	-0.0141	-0.0237
	(1.72)	(-0.78)	(-1.56)
CompAdv	-0.0254	0.0101	0.0114
	(-1.61)	(0.53)	(0.79)
ProductsServices	-0.0200	-0.00414	-0.00810
	(-1.12)	(-0.20)	(-0.59)
Quantitative	0.0176	-0.0130	0.0178
	(0.91)	(-0.61)	(1.19)
SmallCap	-0.0419**	0.0225	0.0245*
	(-2.44)	(1.19)	(1.81)
Defensive	0.0276**	0.0227	0.0164
	(2.07)	(1.14)	(1.43)
MidCap	0.000852	0.0162	-0.00135
	(0.06)	(0.85)	(-0.10)
Foreign_ADR	0.00195	0.0793***	0.00485
	(0.12)	(4.00)	(0.25)
PE_Ratio	-0.00142	-0.00245	0.0428***
	(-0.08)	(-0.13)	(2.63)
IntrinsicValue	0.0399**	-0.0760***	-0.0120
	(2.38)	(-3.19)	(-0.73)
FixedIncome	0.0290**	0.0115	-0.00437
	(2.08)	(0.57)	(-0.31)
Tax	0.0106	-0.0208	-0.0706***
	(0.68)	(-0.74)	(-4.88)
Foreign_EM	-0.0335*	-0.00465	0.0141
	(-1.96)	(-0.21)	(0.62)
Obs	286,576	286,576	286,576

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 17: **TNA Shares by Investor Type:**

	% retail	% institutional	% retirement
Undervalued	-0.0164**	0.0188*	0.00245
	(-2.01)	(1.90)	(0.22)
Sector	-0.00998	-0.00960	0.0122
	(-0.96)	(-0.80)	(1.00)
Dividends	0.0426***	-0.0480***	0.00692
	(3.78)	(-3.13)	(0.37)
Derivatives	0.00614	0.00201	-0.0256*
	(0.50)	(0.16)	(-1.82)
CompAdv	0.0115	0.0125	0.00945
	(0.88)	(0.79)	(0.69)
ProductsServices	0.00778	0.0284	-0.0304
	(0.72)	(1.60)	(-1.36)
Quantitative	0.0103	-0.0505***	0.0174
	(0.63)	(-3.10)	(0.69)
SmallCap	-0.0358***	0.0601***	-0.0161
	(-2.88)	(4.04)	(-0.60)
Defensive	-0.0136	-0.0205	-0.0440**
	(-1.08)	(-1.57)	(-2.26)
MidCap	0.0132	-0.00676	-0.0105
	(1.02)	(-0.48)	(-0.65)
Foreign_ADR	-0.0279*	-0.0179	0.00515
	(-1.85)	(-1.24)	(0.36)
PE_Ratio	0.00497	-0.0176	0.0529**
	(0.39)	(-1.00)	(2.48)
IntrinsicValue	0.0253*	0.00762	-0.00114
	(1.66)	(0.41)	(-0.05)
FixedIncome	-0.0260**	-0.00342	-0.0319*
	(-2.00)	(-0.23)	(-1.90)
Tax	0.0152	-0.00224	-0.0138
	(1.07)	(-0.11)	(-0.23)
Foreign_EM	0.00310	0.0439**	0.0242
	(0.19)	(2.25)	(1.40)
Obs	240,902	170,608	37,522

t statistics in parentheses; * $p<0.10$, ** $p<0.05$, *** $p<0.01$