Title

The Value of Being First: Evidence from Mutual Fund Flows

Author

Zhou, Youqing

Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

The Value of Being First: Evidence from Mutual Fund Flows

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Management

by

Youqing zhou

Dissertation Committee:
Associate Professor Christopher Schwarz, Chair
Associate Professor Zheng Sun
Assistant Professor Chong Huang

2019

DEDICATION

To

my family and friends

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express the deepest gratitude to my committee chair, Professor Christopher Schwarz, for his unfaltering support. He has continually conveyed a spirit of excellence in regard to research and teaching. Without his guidance and help this dissertation would not have been possible.

I would like to thank my committee members, Professor Zheng Sun and Professor Chong Huang, whose comments and suggestions have helped me think through many difficult problems. In addition I would like to thank Professor Lu Zheng, for her guidance through each stage of the process.

Special thanks go to my fellow PhD and friend Jie Gao. She has been offering unconditional support ever since we came into this program. I also thank Noel Negrete for her administrative work and the seminar participants at University of California Irvine for their inputs.

My financial support was provided by the University of California, Irvine.

CURRICULUM VITAE

Youqing Zhou

2004-08	B.A. in Economics, B.A. in Mathematics, Wuhan University
2008-10	M.Phil in Economis, Lingnan University
2013-19	PhD in Management, University of California, Irvine

FIELD OF STUDY

Empirical Asset Pricing, Institutional Investors, Behavioral Finance

ABSTRACT OF THE DISSERTATION

The Value of Being First: Evidence from Mutual Fund Flows

By

Youqing Zhou

Doctor of Philosophy in Management

University of California, Irvine, 2019

Professor Christopher Schwarz, Chair

Much of consumer choice is now driven by online shopping and search results. An open question is how much the value of being the “first” result is. In this paper, I use mutual funds to investigate the value. I find that fund with names that start with letters in the first half of the alphabetic ranking attract 0.85% more net flow every quarter comparing with funds in the second half. I find that this effect is driving by excess inflows rather than outflows, consistent with search results driving the findings. This bias toward early alphabetic funds is driven by retail investors, not institutional investors. Although funds with alphabetical advantage gain more investment flows, they do not necessarily perform worse than those with less flows. However, I also find no evidence that fund companies are taking advantage of this bias. Overall, I find significant economic value from being early in search results.

1. INTRODUCTION

In the new digital age, a considerable amount of shopping activities has shifted from brick-and-mortar stores to online marketplace, for example purchasing goods and services on e-commerce websites, and picking investment products on vendors' platforms. Hence much of consumer choice is now driven by online shopping experience. Merchants choose different formats of online stores in terms of advertisement, web design, payment methods, and so on. Are there any common factor or technology that all businesses adopt? Most shopping sites provide a search service that allows customers to quickly narrow down their target range. So yes, the way that search results are presented may affect consumers' decision. More specifically, the order that candidate items are displayed plays a role in forming consumption patterns. An open question is how much the value of being the "first" result is. In this paper, I use mutual funds to investigate this value.

What determines who show up first? Is there a default algorithm to rank search results? The online investment marketplace provides a suitable setup to illustrate the problem. Figure 1 is a snapshot of the mutual fund screener on Wall Street Journal[1]. Even with a few filters applied, there are more than 200 mutual funds for an investor to consider. This large number of candidate funds appear mostly homogenous. Few investors would have the expertise or patience to inspect every single one.

A closer look at the output from the fund screener indicates that these funds are ordered alphabetically by their names. In fact, other popular online mutual fund marketplace adopt this alphabetical sorting algorithm by default as well. Together they could form a common factor – the alphabetical ranking of mutual fund names – that drives investment

[1] http://online.wsj.com/public/quotes/mutualfund_screener.html

flows into and out of mutual funds. In this paper, I examine whether or not the returned search order, represented by mutual fund names, impacts investor flows.

An important question to ask is whether the alphabetical ranking of mutual fund names is correlated with mutual funds' skill. Are those A funds more capable of earning investment returns than Z funds? If so then investors are well justified in choosing A funds over Z funds. In this case it is less interesting to investigate this issue. But no. Intuitively it is difficult to associate a mutual fund's alphabetical position with its skill. Most mutual funds are named after their founders, or have an auspicious meaning for the purpose of building a brand name. It suffices to say that a mutual fund's name is not related to its investment skill.

The mutual fund literature has plenty of studies on what affects mutual fund investor's decisions. According to rational theory, investors would pursue maximum return and minimum risk at the same time. A few performance measures are commonly used by investors to predict future returns, for example funds' simple return, CAPM alpha, and other forms of alpha. Regardless of whether they can truly represent a fund's capability of earning future returns, investors would naturally look at them as reference. Limited attention is also frequently discussed in the literature. It may affect investors' choices under certain circumstances. Sirri & Tufano (1998) contends that selecting funds is not an everyday task for most households. In general, investors tend to purchase funds that are easier or less costly for them to identify. The salience of a fund could be enhanced in many ways, such as stellar performance, advertising, and/or media coverage.

For example, Jain & Wu (2000) discovers that advertised mutual funds render themselves easier for investors to identify and thus attract more money. The cost of advertising and marketing go into mutual fund operating expenses. Hence Barber *et al.*

(2005) finds that the ongoing 12B-1 fees appear to attract investment flows. Window dressing by mutual funds is also in line with the goal of garnering investors' attention. Solomon *et al.* (2014) points out that media coverage of mutual fund holdings is the channel that makes winner funds receive substantial extra flows. Kaniel & Parham (2017) analyzes a prominent ranking list on Wall Street Journal that recommends mutual funds. They conclude that, due to being mentioned in the ranking list, those funds experience remarkable extra flows.

All these elements and efforts reduce the search cost of mutual funds. There surely exist other factors that can serve this purpose. Based on the fact that most fund screeners typically return results in alphabetical order, I argue that the alphabetical order of the name adds to the salience of mutual funds and investors display tendency towards those funds with early alphabets. After controlling other characteristics such as past performance, size, and age, I find that funds in the first half of the alphabetic ranks attracts 0.85% more flow every quarter comparing with the funds in the bottom half. The Growth category turns out to have the strongest name ranking effect – a growth fund that's alphabetically disadvantaged receives 1.82% less investor flows.

In line with this idea, several studies have examined the ranking issue in different context. Feenberg *et al.* (2017) investigates NBER paper downloading via the weekly issuance of an email to subscribers. The email lists new papers which are listed based on the time they are submitted. They find that, although the order or ranking of papers is irrelevant of paper quality and authors' reputation, subscribers download more of the papers that show up earlier in the list. The paper ascribes the bias to cognitive fatigue and time constraint. More interestingly, in the economics discipline, most research papers list authors by last names, not by contribution. Based on this fact, van Praag & van Praag (2008) finds that authors whose name begin with letters placed high in the alphabet garner more attention and develop faster productivity rate.

A further test could show how search order is impacting flows. In addition to net flows, I manage to separate inflows and outflows. The search results order should only impact fund inflows as that is when investors are searching for new funds and have limited attention. Since investors only own a small number of funds, the alphabetical order should not impact outflows. After hand collecting data from NSAR filings, I rerun my flow models and find that indeed inflows, not outflows, are significantly affected by the alphabetic ranking of fund names.

One factor that may impact this bias is the type of investor. Specially, mutual fund investors can be partitioned into groups based on how sophisticated they are (Dhar & Zhu (2006), Keswani & Stolin (2008)). Institutional investors are considered to be more aware of cognitive biases than retail investors. When I split funds into these two types, I find that the impact is driven by retail investors. There is no impact from search order when looking at institutional flows. Thus, search order is only important for unsophisticated investors where search costs are likely higher.

Finally I investigate how this bias has impacted the industry. First, I look whether or not mutual fund companies "game" the system by focusing their fund names at the beginning of the alphabet. However, I find that fund names are similar to that expected based on family names and words in the dictionary. Second, I look at whether or not the bias leads to costs for investors. While investors could save on search costs, investors could be choosing suboptimal funds. However, I find that is not the case. Performance is not impacted by the name of the fund. This is somewhat expected given that mutual fund performance is not persistent and therefore unpredictable (e.g. Carhart (1997)).

Overall, this paper makes several contributions to the literature. First, I show the importance of being first in search results as well as quantify the value. Given more financial services, and more generally consumer product purchases occurs through online search, this paper gives a better understanding of search result impact. Second, I should that investors suffer from search costs when looking for new products, even after using screening tools. Finally, this paper again demonstrates that other fund characteristics besides performance and fees impact fund flows.

The remaining of this paper is organized as follows. Section 2 describes the data. Section 3 presents the results that mutual fund flows are associated with the alphabetic ranking of the names. I also discuss which investor groups are driving the order bias and how inflows and outflows show different pattern of this bias. Section 4 discusses whether mutual funds are utilizing this bias and whether investors are making smart decisions. Section 5 concludes.

2. Data

As is standard in the mutual fund literature, data on mutual fund returns, total net assets, and other fund information come from Center for Research in Security Prices (CRSP) Survivor-Bias-Free US Mutual Fund Database. I filter domestic equity funds by their style codes and other relevant information. Unfortunately, no one set of style codes was available over my entire sample people; thus, I used a combination of style classifications to create my sample. I select funds with the following Lipper Objective Codes or Class Names: MLCE, CA, DL, DSB, LSE, ELCC, SESE, EI, EIEI, GI, G, MLGE, MLVE, MR, SG, SCCE, SCGE, SCVE, MC, MCGE, MCVE, MCCE, LCCE, LCGE, and LCVE. If a fund doesn't have any of the above Lipper Objective Codes or Class Names, I select funds with the following Strategic Insights Objective Codes: AGG, GMC, GRI, GRO, ING, and SCG. Then I use Wiesenberger Objective Code to include G, G-I, AGG, GCI, GRI, GRO, LTG, MCG, and SCG. For funds missing all the style codes, I compute the average ratio of assets invested in common stocks and require a minimum of 80% for a fund to be included in the sample.

After selecting equity funds, I categorize them base on their investment styles. Since my set of style codes is not consistent over the sample, I combine categories into the six main classifications: Small Cap, Mid Cap, Large Cap, Growth, Value, and Others to ensure consistent classifications across time. As is typically in the literature, I exclude index funds which I identify identified by the index fund flag. Institutional funds are identified with a list of words that would indicate the fund is institutional as well as by the institutional indicator in CRSP.

The CRSP "fundno" identifier represents each fund share class as an individual series. I merge multiple share classes of a fund by the identifier "wficn" produced by Wharton Research Data Services. For instance, to derive fund level Total Net Assets, I sum up TNAs over all share classes that have the same wficn. To compute fund returns, I average share class returns weighted by corresponding share class Total Net Assets. Besides raw returns, I estimate fund alphas using the observations of 24 previous months. For non-quantity variables, I take the value of the largest share class when collapsing to fund level except for the institutional indicator.

Finally, as is common, I remove funds with less than five million dollars of assets under management are screened out to remove any undue impact from small funds. My final sample consists of 3,646 domestic equity mutual funds and 148,283 fund-quarter observations between 1999 and 2015. Each year funds come in and out of the industry. There are 2023 funds in the first year 1999, growing to a peak of 2,666 funds in 2,888. The industry then shrinks to a low of 1,827 funds in the year 2015. Panel A of Table 1 lists how many mutual funds there are in each category as defined by the investment object. Each category has been through the same expansion and contraction periods as the entire fund industry has.

Panel B reports the summary statistics of fund Total Net Assets, flows, performance, riskiness, expenses, turnover, and loads information. The average fund manages about $1 billion dollars of assets while the median size is 211 million dollars. This means the market is dominated by a few extraordinarily large funds. There is also considerable cross-sectional variation in fund flows, with an average net outflow of 7 million per quarter. Note that the average fund flow in percentage is a value weighted measure. It's moderately positive, which indicates that large funds see relatively more net inflows than small funds. I summarize several measures of fund performance. Fund raw returns are slightly positive while CAPM alpha, Fama-French-three-factor alpha, and Carhart alpha are nearly zero, if not negative.

This is consistent with the well-documented aggregate underperformance of mutual funds. Riskiness is defined as the standard deviation of past 12 month returns. About 70% of all these funds have either a front-end load or back-end load, or both. The quartile statistics imply that mutual funds charge similar load fees.

The two main variables I construct for my sample is investor flows and my ranking of names. Mutual fund flows are assumed to occur at the end of a month, thus it is calculated as the difference between the value-added lag TNA and the TNA of this period. Specifically, $flow_t = TNA_t - (TNA_{t-1} + merged\ assets_{t-1}) \times (1 + ret_t)$. I also compute fund flows as if it occurs at the beginning of a month: $flow_t = \frac{TNA_t}{1+ret_t} - TNA_{t-1} - merged\ assets_{t-1}$. The analysis is done in the same procedure. Quarterly flows are the sum of 3 consecutive monthly flows. Flows are scaled by net assets in regression analysis.

In addition to using net flows inferred from the CRSP database, I also gather data from NSAR filings on inflows and outflows following Schwarz and Potter (2016). I match fund names by hand each semiannual period to link the CRSP fund with the appropriate fund. To verify I have the correct match, I compare the assets on the NSAR filing and that listed in CRSP. NSAR filings list inflow and outflows by month on the form directly. I can therefore simiply compute the quarterly information by summing the inflow and outflows over time.

Finally, the key object of study is the alphabetic ranking of mutual funds. This is determined by fund names. The first character of a fund name is under one of these three major categories: (1) letters such as, not exclusively, “a” or “A”; (2) a numeric digit such as “2”, “5”, or “8”; (3) symbols such as “#”. I remove type 2 and 3 funds since there are very few of them. Then I assign a score of -1 to letter “a” or “A”, a score of -2 to letter “b” or “B”, …, and a score of -26 to letter “z” or “Z”. I process the second character of fund names in a similar way, but with the values on a lower scale. For example, if the second letter of a fund is “a”, it gets a score of -0.1, adding to whatever score the first letter gets. In this way, I can preserve the lexicographic order of fund names and have a finer partition of ranking. Panel C of Table 1 reports the distribution of mutual fund name initials of 1999, 2005, 2010, and 2015. The next section elaborates on this key measure.

3. The order bias in mutual fund investing

3.1 The effect of alphabetic ranking of fund names

This section examines whether the alphabetic ranking of fund names predicts fund flows. The key measure is a fund's position as sorted alphabetically by its name. I assign scores to fund names according in the following way. Firstly, I take the first letter of a fund's name. If it is 'A' or 'a', the score is -1. If it is 'B' or 'b', the score is -2. And so on. The last alphabet is 'Z' or 'z' and the score is -26. Thus, higher scores represent funds in the upper top of the alphabet list. Secondly, I refine the ranking measure to one decimal place by using the second letter from fund names. If the second letter is 'A', the fund gets a score of -0.1, adding to its first letter. If the second letter is 'B', the fund gets an additional score of -0.2. The goal is to measure fund's position more precisely and meanwhile, to preserve the lexicographic order.

My conjecture is that funds with a higher position down in the alphabetic queue attracts more attention from investors. That said, the fund name is indeed a categorical variable. I do not want to over interpret the quantitative meaning of the lexicographic ranking. Therefore, in the benchmark model, I adopt a dummy measure of the name ranking. The dummy variable is set to 1 if the fund's name ranks in the top half, and turns to 0 if the fund's name ranks in the second half. For example, the dummy variable for an 'A' fund is in the top half and has a value of 1. Meanwhile, a 'Z' fund is in the bottom of the list and thus the dummy should be 0. In the baseline tests I also include the linear measure as defined above, but the dummy measure will be the key independent variable throughout.

A strand of the fund literature has focused on the non-linear relation between fund flows and performance. To be inclusive, I use the piecewise measure of fund performance as prescribed in Sirri & Tufano (1998). First, mutual funds are sorted by performance within its category from low to high (1-100) and then partitioned into five linear measures. Specifically, for any fund, there are five pieces of the measure. The bottom, or the 5th quintile $PERF_5$ is defined as $\min(RANK_{t-1}, 20)$. The 4th quintile $PERF_4$ is defined as $\min(RANK_{t-1}, 20)$. The 3rd quintile $PERF_3$ is $\min(20, RANK_{t-1} - PERF_4 - PERF_5)$. The 2nd quintile $PERF_2$ is $\min(20, RANK_{t-1} - PERF_3 - PERF_4 - PERF_5)$. Lastly, the top, or the best quintile $PERF_1 = \min(20, RANK_{t-1} - PERF_2 - PERF_3 - PERF_4 - PERF_5)$. This set of variables would capture the heterogeneous flow-performance relation across the performance spectrum.

The mutual fund literature has discovered many other fund characteristics to be associated with fund flows. To follow the literature, I use the natural logarithm of TNA as fund size, and measure fund riskiness as the standard deviation of fund returns in the past 12 months. In the baseline model, I sum up fund flows into the same investment objective to control for the style effect. Other fund characteristics include fund age in years, and fund expense. Sirri & Tufano (1998) emphasizes the role of costly search as in mutual fund flows. I follow their model specification which computes an enhanced version of fund expense. This alternative expense measure turns out to be not crucial to our hypothesis testing. Later I will trim down to use the simple expense variable from CRSP.

Panel A of Table 2 reports the regression of fund flows on the name ranking variable and other characteristics. These regressions are run quarter by quarter. Standard errors and t-statistics are calculated from the vector of quarter results, as in Fama and MacBeth (1973). P-value are given in italics below the coefficient estimates. The first four columns use the dummy measure of mutual fund name ranking while the last two columns use the linear measure. The t-statistics indicate that the name ranking plays an interesting role in explaining fund flows. For a fund that falls into the first half of the alphabetic rank, it attracts 0.85% more flow every quarter comparing with the funds in the second half[2]. For an average fund with approximately 1 billion dollar of assets under management, this translates into 8.5 million dollars more fund flows per quarter. To interpret the results for the linear measure, if a fund moves up the alphabets by one tick, for instance from 'B' to 'A', the fund will gain 0.05% flows per quarter on average. If we roughly calibrate the effect of having an earlier name by using this coefficient estimate, a fund that changes its name from 'Z' to 'A' will see an increase of 1.25% in fund flow per quarter. Given a median Total Net Assets of 1 billion, this gain is more than 10 million dollars.

All the other fund characteristics behave as expected. The piecewise design of fund performance measure shows a clear pattern of convex flow-performance relation. The worst performing funds have a significant coefficient of 12 basis points while the top funds have a coefficient of 26 basis points. Other funds in between have either insignificant or quantitatively small estimates. This convexity shows up in every regression model regardless of whether fund performance is measured by raw return or alternative asset pricing models. Fund size and fund age both negatively affect fund flows in a significant way. The riskiness of fund returns also has a negative coefficient estimate, which indicates that investors tend to avoid risk. Investment flow into the entire category of a fund in question naturally has a positive coefficient. Expense ratio negatively affects fund flows. 1% increase in expense ratio brings about 1.23% decrease in fund flows. As Sirri & Tufano (1998) shows,

[2] In untabulated results, I add name rank of the fund family as additional control. Both name rank variables are negative.

marketing expenditure attenuates the loss in fund flows due to high expense. It doesn't affect other coefficients so that I suspend it in other analysis.

In addition to raw returns, I also measure fund performance by excess return, CAPM alpha, Fama-French alpha, and four factor alpha separately. Panel B continues to explore those possibilities. To focus on the key hypothesis, I report the coefficient estimates on the name variable only. Similar results go through for each model specification. The estimate stays in the range of 0.74 – 0.85. For example, when I measure fund performance by the alpha from the four factor model, an early fund on average receives an increase of 0.75% fund flows associated just with its name.

I categorize mutual funds by their investment objective as provided by CRSP and other sources. Mutual funds' flows are affected by trendy investment fads and fashion, which oftentimes spread to a whole style category. Table 2 controls the style effect by including the total investment flow into that category. Another question regarding style investing is that the investor base may be heterogeneous across style categories. Table 3 runs the same tests for each style category separately. The Growth category turns out to have the strongest name ranking effect. In Panel A, a growth fund that ranks high above the alphabets receives 1.82% more investment flows. Large cap funds and mid cap funds also show a significant negative correlation 0f 0.58 (or 0.83) between fund flows and a fund's name ranking. Mutual funds in other categories have negative coefficient on the name ranking variable as well. But they are not statistically significant. The magnitude of the estimates go down as I use more sophisticated measure of fund performance while remains economically significant. In summary, all the baseline tests confirm that a fund's name, its alphabetic ranking in particular, plays a significant role in attracting investment flows.

3.2 Who is driving the bias: Institutional or retail investors

One would expect certain investor groups are more cautious in selecting mutual funds. In this section I examine whether there is heterogeneity in investors regarding choosing mutual funds. As discussed in many papers, institutional investors in general are more aware of cognitive traps than individual investors. Hence at share class level, I identify institutional (or retail) shares by the institutional (or retail) indicator provided by CRSP. I also run text analysis for fund names to partition them into those two groups. Then the same tests are run for both groups. Table 4 reports the summary statistics and regression results.

The regression analysis in Panel B and C shows a striking difference between institutional investors and retail investors. For our interest, institutional investors show insignificant reaction to mutual fund name's alphabetic ranking, while retail shares have a significant estimate of 0.72. That is almost identical to the estimate on the whole sample. The regression using four factor alpha as performance measure in Panel C shows similar pattern. Institutional investors don't simply choose mutual funds because they are ranked higher by their names while retail investors show some inattention to the default order mutual funds are presented. The magnitude also matches the estimate for the whole sample, i.e. 0.61 versus 0.74. Therefore, the effect of mutual fund name's alphabetic ranking is mostly driven by retail shares.

The institutional and retail investors are different in many other ways as well. After breaking down the performance measure, it is clear that institutional investors react more strongly to loosing funds. At the other end of the spectrum, they do not show much interest in winner funds. On the contrary, retail investors avoid losing funds only moderately while chasing winner funds with some enthusiasm. This situation changes somehow if one uses more sophisticated model to measure fund performance. As Panel C shows, when mutual funds are evaluated by four factor alphas, institutional clients are almost nonchalant to this performance metric. On the other hand, the retail shares have approximately a convex relation between fund flows and performance. The coefficient estimate on top quintile is more than three times bigger than its correspondent on the bottom quintile.

As for other fund characteristics, institutional and retail investors also have different attitudes. Most significantly, institutional clients don't care about fund riskiness, which is simply measured by the variation of past returns. Retail investors are somehow intimidated by this metric. The expense variable has a much bigger negative impact on flows into institutional shares than on flows into retail shares. Many of these phenomena worth further investigation.

3.3 Who is driving the bias: New or existing investors

When investors consider buying mutual funds, they select funds from a long list of candidates. The bigger the candidate pool is, the more likely investors get tired fumbling through the list. Imagine when a search engine presents many pages of search results, how many would go to the second or third page? If there is just one page of results, people may be interested in reading each one of them more carefully. Similarly, an ordinary investor owns only a few mutual funds, they can examine each one with much greater care no matter how they are listed in the repository. Hence the selling decisions are not subject to the alphabetic ranking effect of mutual fund names. This section I investigate whether new investors or existing investors are causing the unequal fund flows due to their name ranking.

A simple way is to assume inflows going into a fund are contributed by new investors. Outflows are for sure the result of existing investors withdrawing money. Inflow and outflow data are obtained from form N-SAR filings with the Securities and Exchanges Commission. Registered investment management companies have been obliged to disclose their financial information in form N-SAR. Among many things, I focus on Item 28. This item includes four sub items of inflow and outflow for each month of the half year reporting period. I define *Shares Sold: New Sales* as fund inflow, *Shares Redeemed and Repurchased* as fund outflow. The difference between those two is the net flow. I use this information to compare and match with CRSP data.

Matching the N-SAR filings with CRSP takes a lot of tedious labor. In addition, SEC filings and CRSP contain different mutual fund portfolios (Schwarz & Potter, 2016). For those matched by name and assets, I impose the condition that the net flow as calculated from N-SAR inflow and outflow shall have less than 2% difference than the CRSP flow data. This requirement further matches more than 80% of the data.

The final sample comprises of 2363 equity mutual funds during the period of March 1999 through December 2009. Table 5 Panel A reports the summary statistics for the matched funds and compare them to the whole CRSP sample. The average size of the NSAR matched funds are those large funds in the CRSP sample. The size and flows are more spread out than the whole sample. The returns are a bit lower but other fund characteristics match well. For our interest, the average net flow as calculated from NSAR inflow and outflow is -2.22 million, which is a bit higher than the average -6.45 million flow reported by CRSP. This discrepancy may not yet be justified by the standard deviation. It remains to be seen whether the matched sample are consistently different than the whole sample.

I proceed to analyze the sample that has NSAR flows. The same regression model, presented in the first four columns of Panel B, shows similar results as with the whole CRSP sample. An average fund that ranks in the first half by name gains 0.32% fund flows compared with a fund that ranks in the second half by name. The magnitude is less than that of the estimate with the whole sample (0.85). But they are qualitatively the same. The last three columns use NSAR flow data. The Net flow column adopts the same model specification as used in the tests of CRSP data. The coefficient on the name ranking variable is estimated to be 0.31, the same as the regression that uses CRSP flow data, which means that the data is matched in a decent manner.

Inflow and outflow are distinct from the net difference between them. The *Inflow* column uses NSAR inflow as the dependent variable and includes its counterpart—the outflow as an independent variable, i.e. $inflow = f(name\ ranking, outflow, others)$. The *Outflow* column uses NSAR outflow as the dependent variable and includes NSAR inflow as a right-hand side variable, i.e. $outflow = f(name\ ranking, inflow, others)$. In both columns, the counterparts turn out to be highly correlated with each other. This perhaps points to how active this fund is being traded.

It is yet more interesting to look at the estimates on the name ranking variable, marked in bold. Inflows are significantly affected by the alphabetic ranking of fund names. A fund that ranks in the first half by name gains 0.38% fund inflows compared with a fund that has an alphabetically disadvantaged initial. This confirms that new investors pay more attention to those funds with other things being equal. Outflows, on the other hand, have an insignificant estimate. This is consistent with the hypothesis that since investors typically have limited number of holdings in their portfolios, they can give ample time and consideration to every single candidate when deciding which ones to sell. Hence investors are not subject to the behavioral bias caused by the alphabetic ranking of fund names. It can be concluded that new investors are the culprit in regard of ignoring mutual funds just because of their names.

4. More questions

4.1 Do funds play the name game

Following the fact that investment flows are associated with a fund's name, a natural question to ask is that whether mutual funds attempt to gain advantage by giving itself an earlier name on the alphabet list. To examine this issue, I obtain some commonly used statistics of letter distribution to compare with mutual fund names. The first statistics is the relative frequency of the first letter of English words obtained from Wikipedia[3]. I graph it in Panel A of Figure 1 along with the distribution of mutual fund names. If mutual funds are giving themselves an earlier name to gain advantage, its distribution should show a pattern similar to first-order stochastic dominance compared with the natural distribution of English letters. In the figure, the fund name distribution sits to the left of the English letter distribution almost everywhere. That is a subtle sign of gaming the names to some extent.

The other distribution of English words is my own work from scratch. I count the pages that each letter occupies in a Merriam-Webster Dictionary (Copyright 2004). Although the number of words on each page is different, I assume it averages out evenly for each letter. Hence I produce a frequency statistics from the count of pages. It is graphed in Panel B along with the distribution of mutual fund names. This time the two distributions line up side to side nicely, suggesting that the mutual fund names are not quite different than the vocabulary in the dictionary.

A last source I would like to make use of is American last names. Many American corporations are named after their founders. Mutual funds are no exception. I pull out a

[3] https://en.wikipedia.org/wiki/Letter_frequency

statistics of American last names by the United Census Bureau[4] and compare them with mutual fund names. Again the two distributions match each other almost identically. We cannot conclude that mutual funds are giving themselves earlier name to gain more investment flows.

Another way to look at this problem is to examine the trend of mutual fund names. Do they move ahead on the alphabet list? For the new funds could get a better name when they join the market. In Table 6, I summarize the median name scores for each style category at different period.

4.2 Does the name ranking also affect performance

What will happen after those alphabetically early funds take in more investment flows? Do they perform better than those ranked further down the list? This section I examine the performance of mutual funds that are distinguished by their names and thus investment flows. I compute fund returns by assuming proceeds are reinvested every quarter, rolling out to 5 years maximum. The early funds are those rank top 50% by their names. The late funds are those rank in the second half by names. Panel A of Figure 2 presents the whole sample that contains funds from all the style categories. In this graph, the early funds and late funds don't have visible difference in their out of sample performance. Jain & Wu (2000) finds that advertised mutual funds, which attract more investment flows, do not perform better in the future. My results lean towards the other direction.

The next step is to look at each category individually, since we have confirmed that three categories – Large Cap, Mid Cap, and Growth funds – have shown significant name

[4] https://www.census.gov/topics/population/genealogy/data/2000_surnames.html

ranking effect. Figure 2.B, C, and D present the Large Cap category funds, the Mid Cap, and the Growth category funds respectively. The Large Cap category shows almost no difference in performance between the early funds and late funds. This result is understandable because large funds are usually well arbitraged. Mispricing is less common among large funds. For the Mid Cap category, the early funds perform slightly better than late funds. The difference in returns accumulate to be 4% after 5 years. It is similar for the Growth funds. However, the difference is negligible. For the first year or two, the funds with early alphabets are not visibly different than those with late alphabets regarding performance. Since mutual funds gain extra investment flows due to their alphabetically early names, and if we assume fund names are not associated with manager skills, this is evidence that mutual funds do not have decreasing return to scales.

5. Conclusions

This paper investigates a behavioral channel that might affects mutual fund flows. When investors are searching through a menu of mutual funds, their attention lapse as they scroll down the candidate list. Some may give up considering funds that are at the bottom just because they have a name initials like X, or Y, or Z. Hence, top alphabetical names give mutual funds an advantage of gaining investor flows. Specifically, I test whether the lexicographical rank of mutual funds is associated with fund flows. I find that for a fund that falls into the first half of the alphabetic rank, it attracts 0.85% more flow every quarter compared with the funds in the second half. The Growth category turns out to have the strongest name ranking effect – a growth fund that's alphabetically disadvantaged receives 1.82% less investor flows.

Consistent with previous studies that show that professional investors are more sophisticated than individuals, I find institutional shares of mutual funds immune from the name order bias. It's the retail investors that display the tendency towards early alphabets. With additional NSAR data, I tear apart new investor from existing investors. Fund inflows, which could come from new investors, are significantly affected by the alphabetic order of listing. Whereas fund share redemptions do not reveal this order bias.

The results indicate that choice is sensitive to the order of presentation. It has roots in both the theory of costly search and limited attention. I argue that the alphabetical order of the name adds to the salience of mutual funds and the search costs are effectively reduced for those funds with early alphabets. This order bias will persist as long as the pool of candidate funds is large enough to diverse investors' attention. Finally, although alphabetically early funds gain more investment flows, their performance are not necessarily worse than those with less flows due to their names. If we assume fund names are not associated with manager skills, this may shed light on whether mutual funds have decreasing return to scales.

References

Agarwal, V., Gay, G. D., & Ling, L. (2014). Window Dressing in Mutual Funds. *The Review of Financial Studies, 27*, 3133-3170.

As easy as ZYX. (2001, August 30). *The Economist*.

Bailey, W., Kumar, A., & Ng, D. (2011). Behavioral biases of mutual fund investors. *Journal of Financial Economics, 102*, 1-27.

Barber, B. M., Huang, X., & Odean, T. (2016). Which factors matter to investors? Evidence from mutual fund flows. *The Review of Financial Studies, 29*(10), 2600-2642.

Barber, B. M., Odean, T., & Zheng, L. (2005). Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows. *Journal of Business, 78*(6), 2095-2119.

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

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

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *LII*(1), 57-82.

Chevalier, J., & Ellison, G. (1997). Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Polirical Economy, 105*(6), 1167-1199.

Del Guercio, D., & Reuter, J. (2014, August). Mutual Fund Performance and the Incentive to Generate Alpha. *The Journal of Finance, LXIX*(4), 1673-1704.

Del Guercio, D., & Tkac, P. A. (2002). The Determinants of the Flow of Funds of Managed Portfolios: Mutual Funds vs. Pension Funds. *The Journal of Financial and Quantitative Analysis, 37*, 523-557.

Dhar, R., & Zhu, N. (2006, May). Up Close and Personal: Investor Sophistication and the Disposition Effect. *Management Science, 52*(5), 726-740.

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

Huang, J., Wei, K. D., & Yan, H. (2007). Participation Costs and the Sensitivity of Fund Flows to Past Performance. *The Journal of Finance, LXII*(3), 1273-1311.

Ivkovic, Z., & Weisbenner, S. (2009). Individual Investor Mutual Fund Flows. *Journal of Financial Economics, 92*, 223-237.

Jain, P. C., & Wu, J. (2000). Truth in Mutual Fund Advertising: Evidence on Future Performance and Fund Flows. *The Journal of Finance, LV*(2), 937-958.

Jurajda, S., & Munich, D. (2010). Admission to selective schools, alphabetically. *Economics of Education Review, 29*, 1100-1109.

Kaniel, R., & Parham, R. (2017). WSJ Category Kings - The impact of media attention on consumer and mutual investment decisions. *Journal of Financial Economics, 123*, 337-356.

Keswani, A., & Stolin, D. (2008). Which Money Is Smart? Mutual Fund Buys and Sells of Individual and Institutional Investors. *The Journal of Finance, LXIII*(1), 85-118.

Schwarz, C. G., & Potter, M. E. (2016). Revisiting Mutual Fund Portfolio Disclosure. *The Review of Financial Studies, 29*(12), 3519-3544.

Sensoy, B. A. (2009). Performance Evaluation and Self-designated Benchmark Indexes in the Mutual Fund Industry. *Journal of Financial Economics, 92*, 25-39.

Sirri, E. R., & Tufano, P. (1998). Costly Search and Mutual Fund Flows. *The Journal of Finance, LIII*(5), 1589-1622.

Solomon, D. H., Soltes, E., & Sosyura, D. (2014). Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics, 113*, 53-72.

van Praag, C., & van Praag, B. M. (2008). The Benefits of Being Economics Professor A (rather than Z). *Economica, 75*, 782-796.

Figure 1: Search results from Wall Street Journal mutual fund screener

This is a screen shot of the mutual fund screener on The Wall Street Journal. Filter applied are (1) Lipper Category = Large-Cap Growth Funds; (2) Total return = HIGHEST; (3) Time Period = 3 Years; (4) Expense Ratio = 0-2.5%.

Source: http://online.wsj.com/public/quotes/mutualfund_screener.html

MUTUAL FUND SCREENER: Results

Page 1 (Equities 1-25 of 129) Next >

Fund Name	Ticker	NAV as of 8/28/2017	Total Net Assets	Load Adjusted Returns			
				1 Yr Return	5 Yr Return	10 Yr Return	Since Inception
AB Lg Cap Gr;A	APGAX	45.23	$1,954,500,000	13.31%	16.38%	10.45%	N/A
AB Lg Cap Gr;Adv	APGYX	48.93	$1,965,900,000	18.63%	17.68%	11.19%	5.74%
AB Lg Cap Gr;B	APGBX	34.33	$21,500,000	14.52%	16.44%	10.00%	N/A
AB Lg Cap Gr;C	APGCX	34.76	$312,900,000	17.47%	16.50%	10.06%	N/A
AB Lg Cap Gr;I	ALLIX	48.68	$236,900,000	18.64%	17.77%	11.37%	10.58%
AB Lg Cap Gr;K	ALCKX	46.01	$90,600,000	18.26%	17.35%	10.98%	10.17%
AB Lg Cap Gr;R	ABPRX	43.61	$53,800,000	17.88%	17.00%	10.66%	7.82%
ActivePassive LC Gr;A	APLGX	11.09	$33,800,000	19.52%	N/A	N/A	0.56%
Alger Inst:Cap App F;A	ALAFX	29.62	$22,900,000	15.42%	N/A	N/A	14.81%
Alger Inst:Cap App F;Z	ALZFX	30.12	$57,300,000	22.13%	N/A	N/A	16.62%
AllianzGI:Foc Gr;B	PGFBX	N/A	N/A	5.68%	14.92%	9.36%	N/A
Amer Cent:Select;R6	ASDEX	69.12	$800,000	19.33%	14.88%	8.92%	18.67%
Amer Cent:Ultra;I	TWUIX	43.35	$238,000,000	20.71%	16.01%	9.49%	4.90%
Amer Cent:Ultra;Inv	TWCUX	42	$9,093,000,000	20.46%	15.78%	9.27%	N/A
Amer Cent:Ultra;R6	AULDX	43.37	$173,700,000	20.89%	N/A	N/A	13.82%
American Funds Gro;R6	RGAGX	48.09	$22,383,000,000	20.67%	16.51%	8.06%	6.00%
Baron Fifth Ave Gro;Inst	BFTIX	23.85	$91,900,000	26.37%	16.72%	8.25%	7.16%
BlackRock:Foc Growth;A	MDFOX	3.67	$63,100,000	15.56%	14.48%	8.67%	-15.00%
BlackRock:Foc Growth;I	MAFOX	3.94	$35,400,000	22.32%	16.14%	9.69%	-14.00%
CB Large Cap Gr;A	SBLGX	38.25	$1,626,100,000	8.42%	15.68%	8.61%	7.37%
CB Large Cap Gr;I	SBLYX	42.23	$5,002,700,000	15.37%	17.47%	9.66%	8.22%
CB Large Cap Gr;IS	LSITX	42.27	$988,600,000	15.46%	N/A	N/A	15.51%
CB Large Cap Gr;R	LMPLX	36.88	$72,700,000	14.73%	16.69%	8.96%	3.75%
Columbia:Disc Gro;R5	CQURX	9.92	$5,000,000	18.83%	16.05%	8.73%	7.82%
Columbia:Disc Gro;Z	CLQZX	9.64	$109,800,000	18.76%	15.89%	8.67%	7.77%

Table 1: Summary Statistics

This table uses data from January 1999 through December 2015. There are 3646 domestic equity mutual funds and 148283 fund-quarter observations during the sample period. Total net assets are assets under management reported in million dollars. Fund flows are calculated as the difference between two consecutive TNAs that's not attributable to asset growth or merge and acquisition. Quarterly flows are the sum of three consecutive monthly fund flows. CAPM alpha is the intercept from the regression of $R_t - R_{f,t} = \alpha + \beta(R_m - R_{f,t})$. Fama French alpha is the intercept from the three-factor model $R_t - R_{f,t} = \alpha + \beta(R_m - R_{f,t}) + h \cdot HML_t + s \cdot SMB_t$. Riskiness is the standard deviation of 12 past returns. Statistics on front/rear loads are summarized from the sample of funds that have either front-end or rear-end loads respectively.

Panel A

Year	Large Cap	Mid Cap	Small Cap	Growth	Value	Other
1999	646	271	426	213	131	336
2003	765	351	504	181	162	390
2007	678	378	560	254	175	505
2011	609	322	466	217	138	469
2015	472	254	404	182	115	400

Panel B

	Mean	Standard deviation	25% percentile	Median	75% percentile
Total net assets, million $	1023.9	2481.4	53.4	211.1	794.9
Quarterly flows, million $	-7.2	95.9	-14.3	-1.4	3.9
Quarterly flows, % of TNA	1.87%	16.31%	-4.60%	-1.20%	3.69%
Quarterly returns	1.82%	4.28%	-0.77%	1.66%	4.29%
CAPM alpha	0.04%	0.82%	-0.34%	-0.02%	0.35%
Fama French alpha	-0.02%	1.18%	-0.30%	-0.06%	0.20%
Carhart alpha	-0.04%	1.23%	-0.31%	-0.07%	0.17%
Riskiness	4.77%	1.41%	3.89%	4.53%	5.46%
Annual expense	1.27%	0.87%	0.97%	1.21%	1.49%
Turnover ratio	89.07%	122.57%	33.68%	63.23%	108.26%
Maximum front loads	2.49%	1.71%	0.97%	2.43%	3.84%
Maximum rear loads	1.02%	0.89%	0.24%	0.75%	1.72%

Panel C

Year	ABCD	EFG	HIJK	LMN	OPQ	RST	UVW	XYZ
1999	21.64%	16.18%	9.88%	15.34%	11.54%	15.68%	9.50%	0.26%
2005	24.07%	15.62%	10.56%	13.28%	10.43%	16.80%	9.14%	0.10%
2010	23.56%	14.64%	11.72%	14.62%	10.35%	15.30%	9.68%	0.13%
2015	25.88%	13.55%	11.16%	14.72%	9.77%	14.37%	10.51%	0.04%

Table 2: Effect of name ranking

This table examines whether the alphabetic ranking of fund names predicts fund flows. The dependent variable is mutual fund's flow as a percentage of its Total Net Assets. The first four columns use a dummy measure of the ranking. For funds that rank in the top half by name, the dummy variable is set to 1. For funds in the second half, the dummy is set to 0. I also rank fund name from last quarter, labeled as 'lag dummy'. This set of tests also utilize an alternative measure of fund expense as suggested by Sirri & Tufano (1998). The last two columns use a linear measure of the ranking, which is based on the lexicographic order of fund names. Specifically, fund names begin with 'A' has a score of -1, fund names begin with 'B' has a score of -2, and so on. I refine the ranking measure to one decimal place by using the second letter from fund names. If the second letter is 'A', the fund gets a score of -0.1 in addition to its first letter. If the second letter is 'B', the fund gets an additional score of -0.2. Funds are ranked by performance with other funds in its investment objective from low to high (1-100) and then partitioned into five linear measures. For example, the bottom or 5th quintile ($PERF_5$) is defined as $\min(RANK_{t-1}, 20)$, the 4th quintile is defined as min(20, $RANK_{t-1} - PERF_5$), and so forth, up to the highest performance quintile. Panel A uses fund raw return as the performance measure. Panel B reports the dummy regression results with other performance measures. These regressions are run quarter by quarter. Standard errors are calculated from the vector of quarter results, as in Fama and MacBeth (1973). T statistics are given in italics below the coefficient estimates.

Panel A

	Dummy measure				Linear measure	
First letter dummy	0.85		0.85			
	3.59		*3.57*			
First letter lag dummy		0.86		0.85		
		3.78		*3.76*		
Name Rank					0.05	
					3.02	
Name Rank lag						0.05
						3.17
Breakdown of performance						
Bottom quintile	0.12	0.12	0.12	0.12	0.12	0.12
	2.59	*2.59*	*2.52*	*2.52*	*2.60*	*2.59*
4th quintile	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	-0.72	*-0.71*	*-0.76*	*-0.74*	*-0.72*	*-0.71*
3rd quintile	0.04	0.04	0.04	0.04	0.04	0.04
	3.06	*3.03*	*3.08*	*3.05*	*3.08*	*3.05*

2nd quintile	0.04	0.04	0.04	0.04	0.04	0.04
	2.38	*2.38*	*2.46*	*2.46*	*2.37*	*2.37*
Top quintile	0.26	0.26	0.27	0.27	0.27	0.27
	8.38	*8.40*	*8.50*	*8.52*	*8.42*	*8.44*
Size	-0.67	-0.67	-0.72	-0.72	-0.67	-0.67
	-7.04	*-7.05*	*-7.64*	*-7.63*	*-7.03*	*-7.04*
Total style flow	1.27	1.27	1.27	1.27	1.27	1.26
	7.83	*7.80*	*7.70*	*7.67*	*7.81*	*7.80*
Riskiness	-0.59	-0.59	-0.54	-0.54	-0.58	-0.58
	-4.67	*-4.66*	*-4.23*	*-4.23*	*-4.59*	*-4.59*
Age (years)	-0.13	-0.13	-0.14	-0.14	-0.13	-0.13
	-12.73	*-12.76*	*-12.93*	*-12.95*	*-12.69*	*-12.68*
Expense sum	-0.72	-0.71			-0.72	-0.72
	-4.01	*-4.00*			*-4.07*	*-4.07*
Expense ratio			-1.23	-1.23		
			-5.27	*-5.27*		

Panel B

	Excess return		CAPM alpha		FF 3 factor alpha		Four factor alpha	
First letter dummy	0.82		0.74		0.77		0.74	
	3.83		*3.09*		*3.37*		*3.23*	
First letter lag dummy		0.83		0.76		0.78		0.75
		4.09		*3.19*		*3.43*		*3.30*
Breakdown of performance								
Bottom quintile	-0.09	-0.09	0.14	0.14	0.13	0.13	0.11	0.11
	-2.72	*-2.73*	*5.37*	*5.38*	*5.84*	*5.83*	*4.35*	*4.34*
4th quintile	0.00	0.00	0.09	0.09	0.09	0.09	0.09	0.09
	-0.02	*-0.05*	*6.32*	*6.36*	*7.80*	*7.78*	*5.66*	*5.62*
3rd quintile	0.03	0.03	0.12	0.12	0.08	0.08	0.09	0.09
	2.10	*2.12*	*5.42*	*5.43*	*5.76*	*5.74*	*5.26*	*5.28*
2nd quintile	0.03	0.03	0.12	0.12	0.15	0.15	0.13	0.13
	2.35	*2.31*	*5.80*	*5.86*	*3.75*	*3.75*	*5.00*	*5.00*
Top quintile	0.13	0.13	0.42	0.42	0.33	0.33	0.35	0.35
	2.23	*2.23*	*7.69*	*7.68*	*4.88*	*4.88*	*5.78*	*5.78*
Size	-0.62	-0.62	-0.92	-0.92	-0.85	-0.85	-0.83	-0.83
	-6.73	*-6.73*	*-9.54*	*-9.53*	*-8.88*	*-8.87*	*-8.66*	*-8.65*

Total style flow	1.21	1.21	1.35	1.35	1.30	1.30	1.28	1.28
	7.85	*7.83*	*7.77*	*7.74*	*7.52*	*7.50*	*7.33*	*7.30*
Riskiness	-0.56	-0.56	-0.44	-0.44	-0.45	-0.45	-0.43	-0.43
	-4.10	*-4.10*	*-2.92*	*-2.92*	*-2.83*	*-2.83*	*-2.62*	*-2.63*
Age in years	-0.13	-0.13	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11
	-12.7	*-12.8*	*-12.2*	*-12.2*	*-12.1*	*-12.1*	*-12.5*	*-12.5*
Expense sum	-0.69	-0.68	-0.39	-0.39	-0.38	-0.38	-0.37	-0.37
	-3.67	*-3.66*	*-2.22*	*-2.21*	*-2.16*	*-2.14*	*-2.11*	*-2.10*

Table 3: Effect of name ranking within each style

This table presents the previous benchmark regressions with each individual style. Fund styles are defined by the CRSP style codes. The table reports only the coefficient on the name ranking variable. The first two rows use the dummy variable of name ranking. The following two rows use the linear measure of name ranking. Panel A measure fund performance by raw return. Panel B, C, D, and E use excess return, CAPM alpha, Fama French three factor alpha, and four factor alpha, respectively. These regressions are run quarter by quarter. Standard errors are calculated from the vector of quarter results, as in Fama and MacBeth (1973). T statistics are given in italics below the coefficient estimates.

Panel A: Raw return as performance measure

	Large Cap	Mid Cap	Small Cap	Growth	Value	Other
Dummy						
First letter dummy	0.58	0.83	0.18	1.82	0.32	1.77
	2.92	*2.73*	*0.67*	*3.75*	*0.57*	*1.64*
First letter lag dummy	0.59	1.13	0.15	1.76	0.41	1.48
	2.86	*3.71*	*0.54*	*3.80*	*0.74*	*1.37*
Linear						
Name Rank	0.02	0.05	0.00	0.14	0.07	0.12
	1.55	*2.58*	*0.10*	*3.68*	*1.75*	*1.50*
Name Rank lag	0.03	0.06	0.00	0.13	0.05	0.11
	2.05	*3.24*	*0.01*	*3.62*	*1.68*	*1.44*
Number of funds	658	329	482	214	146	436

Panel B: Excess return as performance measure

	Large Cap	Mid Cap	Small Cap	Growth	Value	Other
Dummy						
First letter dummy	0.56	0.81	0.11	2.02	0.37	1.71
	2.75	*2.64*	*0.41*	*3.79*	*0.74*	*1.70*
First letter lag dummy	0.58	1.14	0.09	2.01	0.49	1.39
	2.80	*3.98*	*0.32*	*3.86*	*0.96*	*1.40*
Linear						

Name Rank	0.02	0.05	0.00	0.14	0.07	0.12
	1.55	*2.58*	*0.10*	*3.68*	*1.75*	*1.50*
Name Rank lag	0.03	0.06	0.00	0.13	0.05	0.11
	2.05	*3.24*	*0.01*	*3.62*	*1.68*	*1.44*

Panel C: CAPM alpha as performance measure

	Large Cap	Mid Cap	Small Cap	Growth	Value	Other
Dummy						
First letter dummy	0.41	0.45	0.31	1.56	-0.34	1.89
	2.00	*1.51*	*1.23*	*3.10*	*-0.67*	*1.75*
First letter lag dummy	0.41	0.73	0.29	1.50	-0.17	1.56
	1.92	*2.98*	*1.12*	*3.09*	*-0.35*	*1.42*
Linear						
Name Rank	0.02	0.02	0.02	0.12	0.02	0.14
	1.44	*1.51*	*1.17*	*3.38*	*0.47*	*1.55*
Name Rank lag	0.03	0.04	0.02	0.11	0.00	0.13
	1.92	*2.25*	*1.02*	*3.31*	*0.08*	*1.46*

Panel D Fama French three factor alpha as performance measure

	Large Cap	Mid Cap	Small Cap	Growth	Value	Other
Dummy						
First letter dummy	0.39	0.65	0.37	1.71	-0.42	1.84
	1.87	*2.50*	*1.39*	*3.36*	*-0.85*	*1.69*
First letter lag dummy	0.38	0.91	0.37	1.63	-0.23	1.51
	1.80	*3.20*	*1.33*	*3.28*	*-0.52*	*1.37*
Linear						
Name Rank	0.02	0.04	0.02	0.12	0.01	0.14
	1.27	*2.57*	*1.34*	*3.46*	*0.25*	*1.46*
Name Rank lag	0.02	0.05	0.02	0.11	-0.01	0.13
	1.77	*2.63*	*1.24*	*3.30*	*-0.32*	*1.36*

Panel E Four factor alpha as performance measure

	Large Cap	Mid Cap	Small Cap	Growth	Value	Other
Dummy						
First letter dummy	0.36	0.62	0.36	1.71	-0.36	1.82
	1.78	*2.29*	*1.29*	*3.48*	*-0.77*	*1.67*
First letter lag dummy	0.36	0.88	0.35	1.65	-0.18	1.46
	1.69	*3.24*	*1.22*	*3.39*	*-0.42*	*1.32*
Linear						
Name Rank	0.02	0.04	0.02	0.13	0.01	0.13
	1.21	*2.61*	*1.26*	*3.65*	*0.28*	*1.44*
Name Rank lag	0.02	0.05	0.02	0.12	-0.01	0.13
	1.67	*2.77*	*1.20*	*3.39*	*-0.24*	*1.34*

Table 4: Institutional investors versus retail investors

This table reports the analysis with institutional shares and retail shares separately. The subject at study is mutual fund flows and the key independent variable is a fund name's alphabetic ranking. Institutional shares and retail shares are identified by both the indicators provided by CRSP, and a fund's name. Panel A reports the summary statistics. Panel B reports the baseline regression results done for each group separately, measuring fund performance using raw returns. Panel C presents the same tests, using four factor alpha as fund performance measure. These regressions are run quarter by quarter. Standard errors are calculated from the vector of quarter results, as in Fama and MacBeth (1973). T statistics are given in italics below the coefficient estimates.

Panel A

	Institutional shares			Retail shares		
	Mean	Standard deviation	Median	Mean	Standard deviation	Median
Total net assets, million $	293.40	450.86	87.06	1072.76	4174.73	141.97
Quarterly flows, million $	1.74	24.59	-0.09	-11.21	149.47	-1.20
Quarterly flows, %	3.49%	16.13%	-0.10%	4.26%	91.91%	-1.51%
Quarterly returns	1.88%	3.51%	1.70%	1.78%	4.61%	1.63%
CAPM alpha	0.08%	0.93%	0.01%	0.03%	0.83%	-0.03%
Fama French alpha	0.02%	1.38%	-0.04%	-0.03%	1.21%	-0.07%
Riskiness	4.71%	1.11%	4.51%	4.97%	7.40%	4.54%
Annual expense	0.98%	0.34%	0.97%	1.40%	0.90%	1.35%
Turnover ratio	84.20%	78.87%	66.77%	90.13%	127.53%	62.99%
Maximum front loads	1.86%	1.47%	1.82%	3.52%	1.60%	3.79%
Maximum rear loads	1.66%	1.12%	1.60%	1.16%	0.88%	0.96%

Panel B

	Institutional shares		Retail shares	
	Dummy	Linear	Dummy	Linear
First letter dummy	3.14		0.72	
	1.20		*3.33*	

First letter lag dummy		3.18				0.79		
		1.22				*3.91*		
Name Rank			0.03				0.04	
			0.23				*2.25*	
Name Rank lag				0.08				0.04
				0.88				*2.47*
Breakdown of performance								
Bottom quintile	0.32	0.32	0.31	0.30	0.09	0.09	0.09	0.09
	2.05	*2.04*	*1.91*	*1.91*	*3.37*	*3.35*	*3.33*	*3.30*
4th quintile	0.22	0.22	0.22	0.22	-0.01	-0.01	-0.01	-0.01
	0.53	*0.54*	*0.54*	*0.53*	*-0.23*	*-0.22*	*-0.22*	*-0.21*
3rd quintile	-0.12	-0.12	-0.13	-0.12	0.06	0.06	0.06	0.06
	-0.42	*-0.43*	*-0.45*	*-0.43*	*2.37*	*2.37*	*2.36*	*2.35*
2nd quintile	0.05	0.05	0.06	0.05	0.02	0.02	0.02	0.02
	0.54	*0.53*	*0.56*	*0.55*	*1.15*	*1.12*	*1.13*	*1.13*
Top quintile	-0.15	-0.15	-0.16	-0.16	0.29	0.29	0.29	0.29
	-0.53	*-0.53*	*-0.55*	*-0.56*	*8.53*	*8.55*	*8.57*	*8.59*
Size	-2.10	-2.10	-2.10	-2.11	-0.79	-0.79	-0.79	-0.79
	-1.86	*-1.86*	*-1.87*	*-1.89*	*-8.66*	*-8.66*	*-8.65*	*-8.68*
Total category flow	2.56	2.56	2.60	2.53	1.44	1.44	1.44	1.44
	3.69	*3.69*	*3.62*	*3.60*	*7.38*	*7.37*	*7.30*	*7.31*
Riskiness	0.41	0.41	0.39	0.41	-0.36	-0.36	-0.37	-0.37
	0.48	*0.47*	*0.47*	*0.49*	*-2.89*	*-2.89*	*-2.91*	*-2.91*
Age in years	-0.73	-0.74	-0.72	-0.72	-0.13	-0.13	-0.13	-0.13
	-2.66	*-2.67*	*-2.70*	*-2.69*	*-13.05*	*-13.01*	*-13.06*	*-13.05*
Expense sum	-6.14	-6.12	-6.02	-6.06	-1.32	-1.32	-1.33	-1.34
	-2.38	*-2.37*	*-2.38*	*-2.38*	*-6.74*	*-6.71*	*-6.84*	*-6.83*

Panel C

	Institutional shares		Retail shares	
	Dummy	Linear	Dummy	Linear
First letter dummy	2.76		0.61	
	1.07		*2.85*	
First letter lag dummy	2.81		0.69	
	1.09		*3.38*	

Name Rank			0.01				0.04	
			0.06				*2.30*	
Name Rank lag				0.05				0.04
				0.61				*2.57*
Breakdown of performance								
Bottom quintile	0.59	0.59	0.58	0.59	0.10	0.10	0.10	0.10
	1.02	*1.02*	*1.03*	*1.03*	*3.57*	*3.56*	*3.56*	*3.55*
4th quintile	-0.41	-0.41	-0.40	-0.40	0.09	0.09	0.09	0.09
	-0.79	*-0.79*	*-0.78*	*-0.79*	*4.67*	*4.67*	*4.68*	*4.68*
3rd quintile	0.40	0.40	0.39	0.39	0.06	0.07	0.07	0.07
	1.45	*1.45*	*1.45*	*1.45*	*2.78*	*2.79*	*2.80*	*2.82*
2nd quintile	0.16	0.16	0.15	0.14	0.16	0.16	0.16	0.16
	1.89	*1.89*	*1.76*	*1.67*	*6.23*	*6.24*	*6.25*	*6.24*
Top quintile	0.22	0.22	0.22	0.23	0.33	0.33	0.33	0.33
	1.60	*1.56*	*1.50*	*1.61*	*6.23*	*6.23*	*6.24*	*6.24*
Size	-2.35	-2.34	-2.34	-2.35	-0.96	-0.96	-0.96	-0.95
	-2.06	*-2.06*	*-2.07*	*-2.09*	*-10.51*	*-10.50*	*-10.52*	*-10.55*
Total category flow	2.63	2.63	2.67	2.60	1.49	1.49	1.49	1.49
	3.58	3.58	3.56	3.51	7.55	7.55	7.48	7.49
Riskiness	0.89	0.88	0.88	0.90	-0.28	-0.28	-0.28	-0.28
	0.95	*0.94*	*0.96*	*0.97*	*-1.99*	*-2.00*	*-2.01*	*-2.01*
Age in years	-0.69	-0.70	-0.68	-0.68	-0.10	-0.10	-0.10	-0.10
	-2.56	-2.57	-2.60	-2.59	-12.76	-12.72	-12.77	-12.77
Expense sum	-6.58	-6.56	-6.44	-6.48	-0.94	-0.94	-0.96	-0.96
	-2.18	*-2.18*	*-2.18*	*-2.19*	*-4.82*	*-4.79*	*-4.90*	*-4.90*

Table 5: New investors versus existing investors

Inflow and outflow data are obtained from form N-SAR filings with the Securities and Exchanges Commission. From Item 28 of the filing, I define *Shares Sold: New Sales* as fund inflow, and *Shares Redeemed and Repurchased* as fund outflow. The difference between those two is the net flow. I require that this net flow shall have less than 2% difference than the CRSP flow data. The final sample comprises of 2363 matched funds. Panel A reports the descriptive statistics and compare them with those of the entire CRSP sample. The first four columns in Panel B are the baseline regression results for the matched sample. The last three columns present regression analysis using NSAR flows. The Net flow column adopts the same model specification as used in the tests of CRSP data. The Inflow column uses NSAR inflow as the dependent variable and includes NSAR outflow as an independent variable. The Outflow column uses NSAR outflow as the dependent variable and includes NSAR inflow as an right hand side variable. Other fund characteristics are the same as the baseline regression. Panel C uses four factor alpha as the performance measure, other things equal. These regressions are run quarter by quarter. Standard errors are calculated from the vector of quarter results, as in Fama and MacBeth (1973). T statistics are given in italics below the coefficient estimates.

Panel A

	NSAR matched data			CRSP data		
	Mean	Standard deviation	Median	Mean	Standard deviation	Median
Total net assets, million $	1657.29	5969.12	267.43	1023.9	2481.4	211.1
Quarterly flows, million $	-6.45	193.26	-2.10	-7.2	95.9	-1.4
NSAR quarterly net flows, million $	-2.22	196.60	-1.72			
NSAR quarterly inflows, million $	108.75	366.44	16.59			
NSAR quarterly outflows, million $	110.97	329.85	21.54			
Quarterly flows, %	-0.14%	11.24%	-1.46%	1.87%	16.31%	-1.20%
Quarterly returns	1.08%	4.50%	0.89%	1.82%	4.28%	1.66%
CAPM alpha	0.04%	0.60%	0.01%	0.04%	0.82%	-0.02%
Fama French alpha	-0.03%	0.47%	-0.04%	-0.02%	1.18%	-0.06%
Riskiness	5.04%	1.49%	4.75%	4.77%	1.41%	4.53%
Annual expense	1.27%	0.42%	1.23%	1.27%	0.87%	1.21%
Turnover ratio	88.06%	89.02%	66.7%	89.1%	122.6%	63.23%

Maximum front loads	2.58%	1.77%	2.62%	2.49%	1.71%	2.43%
Maximum rear loads	1.14%	0.92%	0.96%	1.02%	0.89%	0.75%

Panel B

	CRSP net flows				NSAR flows		
					Net flow	Inflow	Outflow
First letter	0.32		0.31				
	2.36		*2.26*				
First letter lag		0.31		0.30	0.31	**0.38**	**0.11**
		2.20		*2.11*	*2.20*	***2.64***	***1.10***
NSAR outflow						0.53	
						10.89	
NSAR inflow							0.28
							8.06
Breakdown of performance							
Bottom performance	0.11	0.11	0.11	0.11	0.11	0.07	-0.08
	5.81	*5.72*	*5.74*	*5.64*	*5.78*	*4.93*	*-5.74*
4th performance	0.03	0.03	0.03	0.03	0.03	0.02	-0.03
	2.28	*2.29*	*2.33*	*2.33*	*2.33*	*1.55*	*-2.43*
3rd performance	0.06	0.06	0.06	0.06	0.06	0.05	-0.03
	5.72	*5.68*	*5.73*	*5.69*	*5.59*	*4.85*	*-4.43*
2nd performance	0.05	0.05	0.05	0.05	0.05	0.04	-0.02
	5.01	*4.98*	*5.02*	*5.00*	*5.01*	*4.35*	*-3.07*
Top performance	0.23	0.23	0.23	0.23	0.23	0.24	-0.05
	9.26	*9.25*	*9.29*	*9.27*	*9.20*	*9.43*	*-3.69*
Size	-0.01	-0.01	-0.05	-0.05	-0.01	0.05	0.08
	-0.21	*-0.22*	*-1.10*	*-1.12*	*-0.25*	*1.30*	*2.53*
Total category flow	1.08	1.08	1.08	1.08	1.09	0.96	-0.40
	11.07	*11.06*	*10.58*	*10.58*	*10.94*	*11.36*	*-6.82*
Riskiness	-0.45	-0.45	-0.42	-0.41	-0.45	-0.21	0.54
	-3.27	*-3.25*	*-2.99*	*-2.97*	*-3.27*	*-1.76*	*7.15*
Age in years	-0.10	-0.10	-0.11	-0.11	-0.11	-0.14	-0.02
	-13.7	*-13.6*	*-14.4*	*-14.3*	*-13.71*	*-18.05*	*-3.76*
Expense sum	-0.64	-0.64			-0.52	-0.19	0.59
	-3.96	*-3.97*			*-3.26*	*-1.35*	*5.02*
Expense ratio			-1.20	-1.20			
			-4.55	*-4.55*			

Panel C

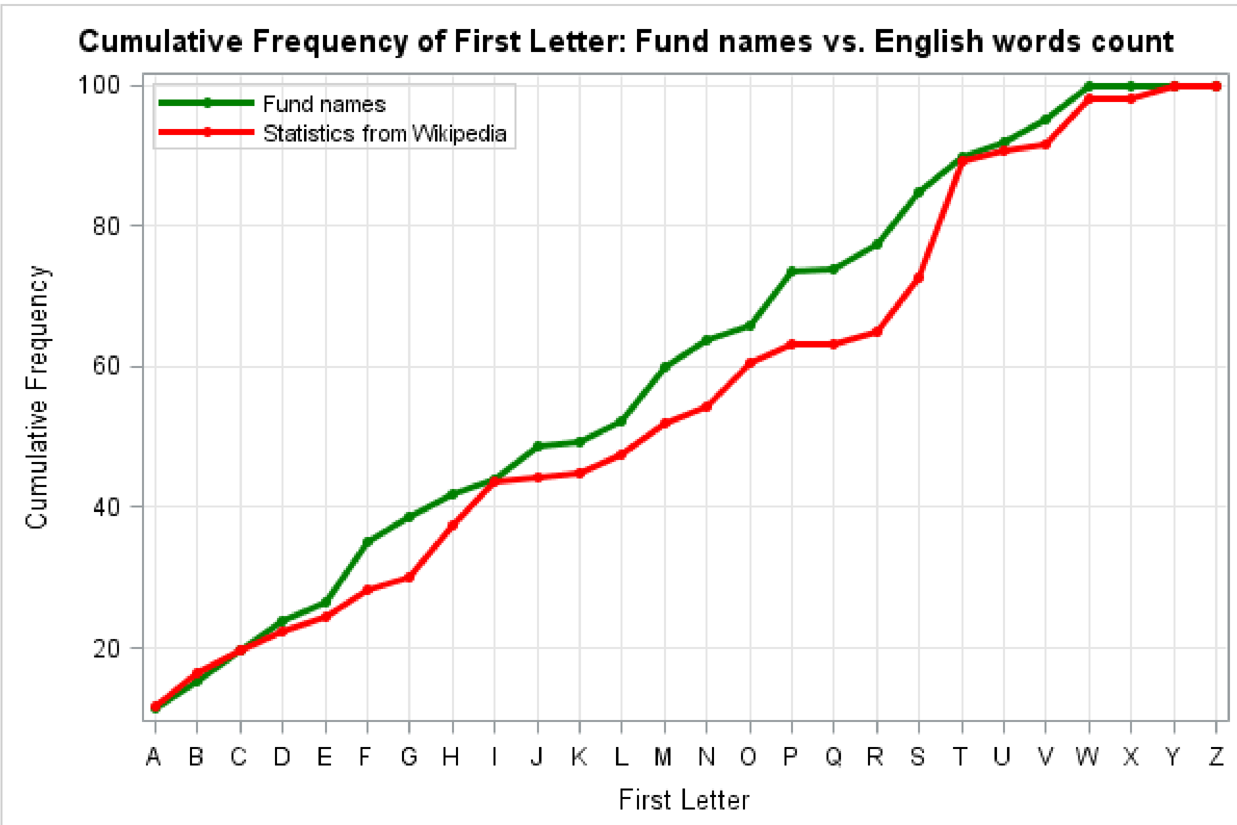
	CRSP net flows				NSAR flows		
					Net flow	Inflow	outflow

First letter	0.30		0.29				
	2.27		2.22				
First letter lag		0.28		0.27	0.28	**0.34**	**0.08**
		2.06		*2.00*	*2.06*	***2.51***	***0.93***
NSAR outflow						0.55	
						11.64	
NSAR inflow							0.29
							8.50
Breakdown of performance							
Bottom performance	0.15	0.15	0.15	0.15	0.15	0.11	-0.10
	10.74	*10.71*	*10.67*	*10.64*	*10.69*	*10.86*	*-8.05*
4th performance	0.05	0.05	0.05	0.05	0.05	0.04	-0.03
	5.96	*6.02*	*5.86*	*5.92*	*5.93*	*5.73*	*-3.92*
3rd performance	0.09	0.09	0.09	0.09	0.09	0.08	-0.04
	7.95	*7.98*	*7.96*	*7.99*	*8.03*	*8.03*	*-5.39*
2nd performance	0.07	0.07	0.07	0.07	0.07	0.06	-0.03
	5.48	*5.48*	*5.58*	*5.59*	*5.45*	*5.61*	*-3.64*
Top performance	0.26	0.26	0.26	0.26	0.27	0.27	-0.06
	10.35	*10.37*	*10.37*	*10.39*	*10.39*	*11.09*	*-3.87*
Size	-0.15	-0.15	-0.18	-0.18	-0.15	-0.08	0.13
	-3.42	*-3.42*	*-3.61*	*-3.62*	*-3.44*	*-1.74*	*4.00*
Total category flow	0.99	0.99	0.99	0.99	1.00	0.90	-0.38
	12.15	*12.15*	*11.64*	*11.65*	*12.02*	*11.68*	*-6.97*
Riskiness	-0.36	-0.35	-0.33	-0.33	-0.36	-0.12	0.49
	-2.77	*-2.76*	*-2.60*	*-2.59*	*-2.76*	*-1.03*	*5.98*
Age in years	-0.09	-0.09	-0.09	-0.09	-0.09	-0.12	-0.03
	-12.5	*-12.5*	*-12.7*	*-12.7*	*-12.52*	*-15.98*	*-5.13*
Expense sum	-0.28	-0.28			-0.16	0.12	0.41
	-1.75	*-1.76*			*-1.03*	*0.86*	*3.38*
Expense ratio			-0.63	-0.63			
			-2.47	*-2.48*			

Figure 2: Do mutual funds play the name game

This set of figures compare the distribution of mutual fund name initials with that of English words or common American last names. The relative frequency of the first letter of English words comes from two sources. One is a statistics by Wikipedia[5]. The other is my own statistics from a Merriam-Webster dictionary. I count the pages that each initial letter occupies and transform them into relative frequency.

Figure 2.A: Relative frequency of the first letter of English words. Source: Wikipedia.



[5] https://en.wikipedia.org/wiki/Letter_frequency.

Figure 2.B: Relative frequency as the first letter of an English word. Source: Merriam-Webster dictionary.

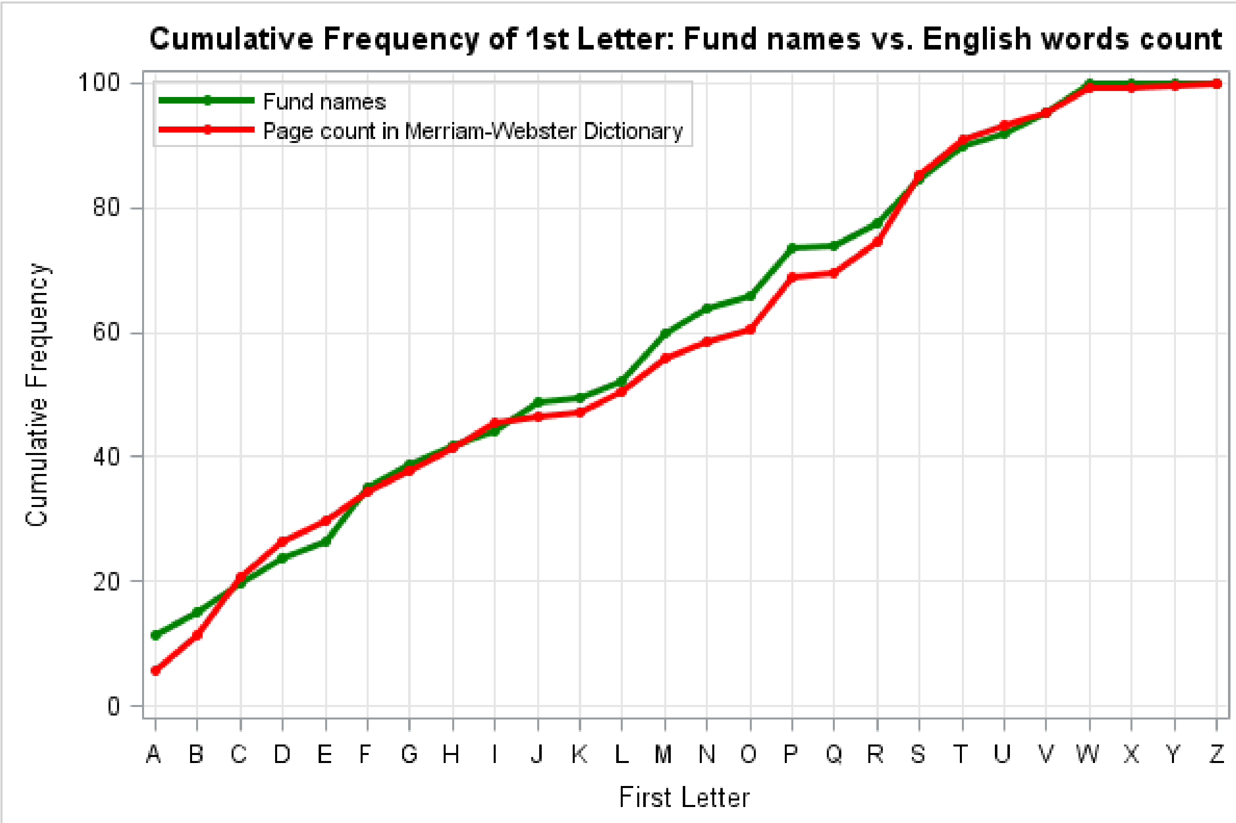


Figure 1.C Count of top 1000 American last names by first letter. Source: United Census Bureau.

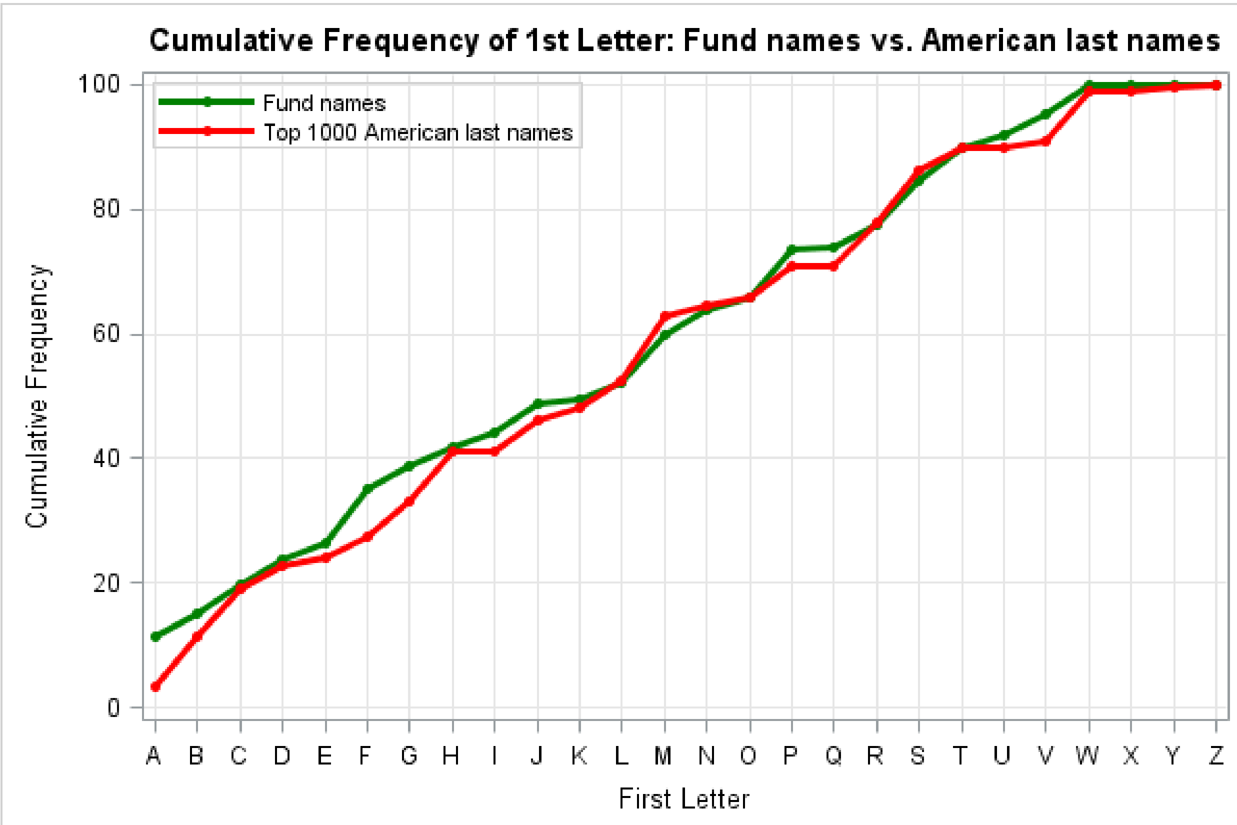


Table 6: Time trend of mutual fund names

This table reports for each style category the median name scores. Specifically, fund names begin with 'A' has a score of -1, fund names begin with 'B' has a score of -2, and so on. I refine the ranking measure to one decimal place by using the second letter from fund names. If the second letter is 'A', the fund gets a score of -0.1 in addition to its first letter. If the second letter is 'B', the fund gets an additional score of -0.2.

DATE	Large Cap	Mid Cap	Small Cap	Growth	Value	Other
31-Dec-01	-12.37	-10.53	-12.47	-12.41	-10.47	-8.33
31-Dec-04	-10.92	-12.37	-12.47	-10.33	-10.42	-11.44
31-Dec-07	-12.37	-12.37	-13.13	-10.33	-10.47	-10.47
31-Dec-10	-10.47	-12.37	-13.13	-10.16	-8.47	-10.47
31-Dec-13	-10.47	-12.47	-13.33	-8.895	-9.14	-12.11

Figure 3: Performance of mutual funds sorted by name

I compute fund returns by assuming proceeds are reinvested every quarter. The early funds are those rank top 50% by their names. The later funds are those rank in the second half by names. Figure 2.A puts all mutual funds together. Figure 2.B through D present the Large Cap category funds, the Mid Cap category, and the Growth category funds respectively.

Figure 3.A

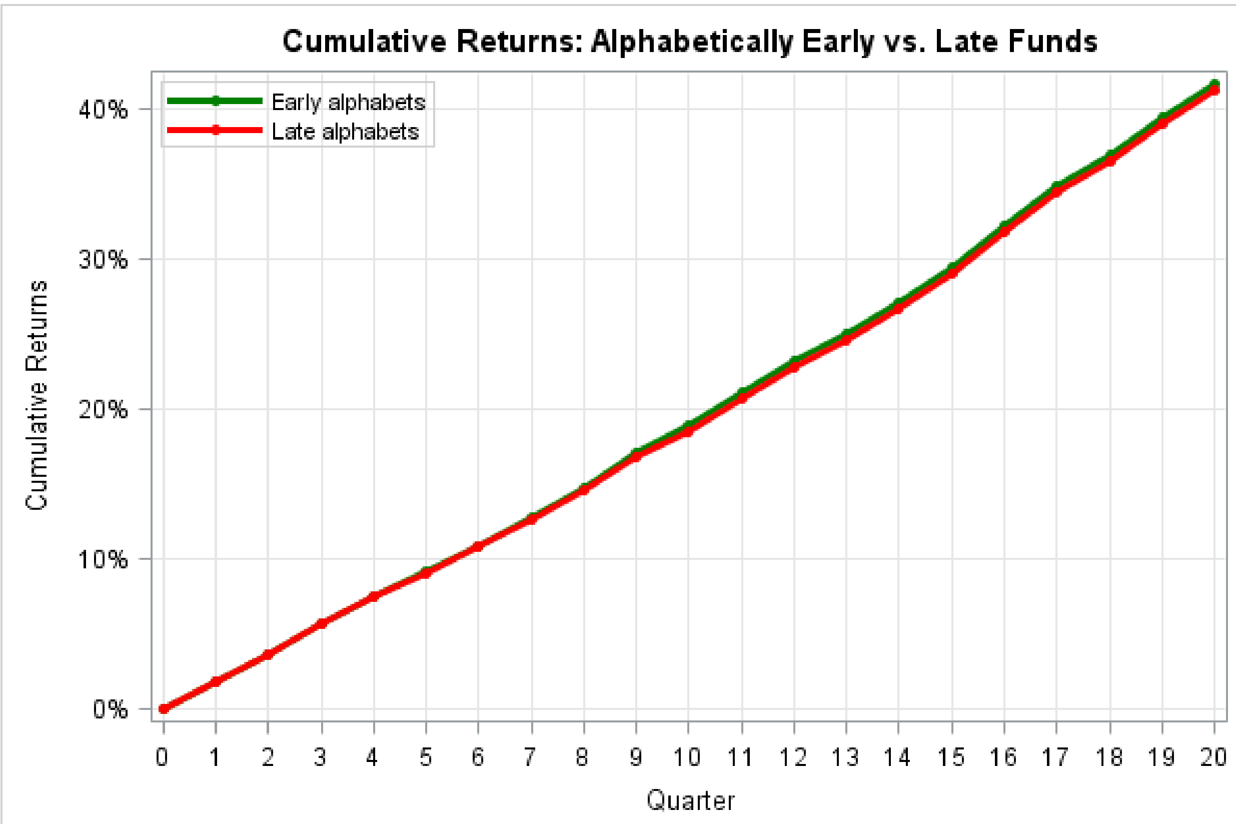


Figure 3.B

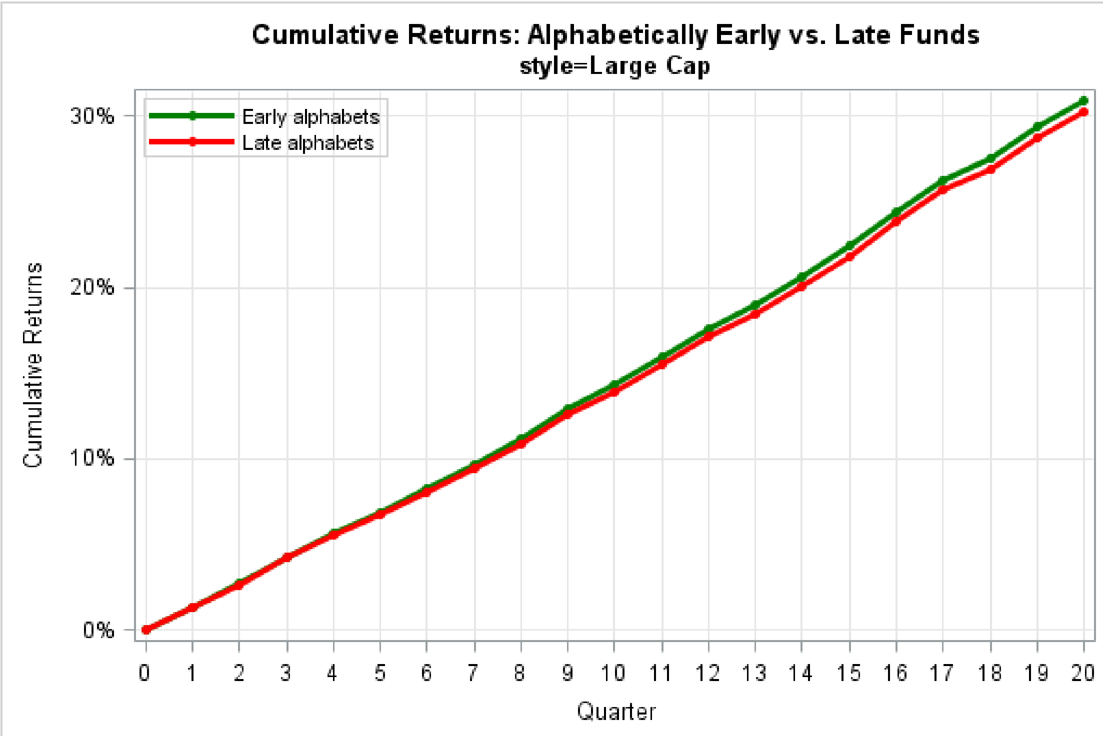


Figure 3.C

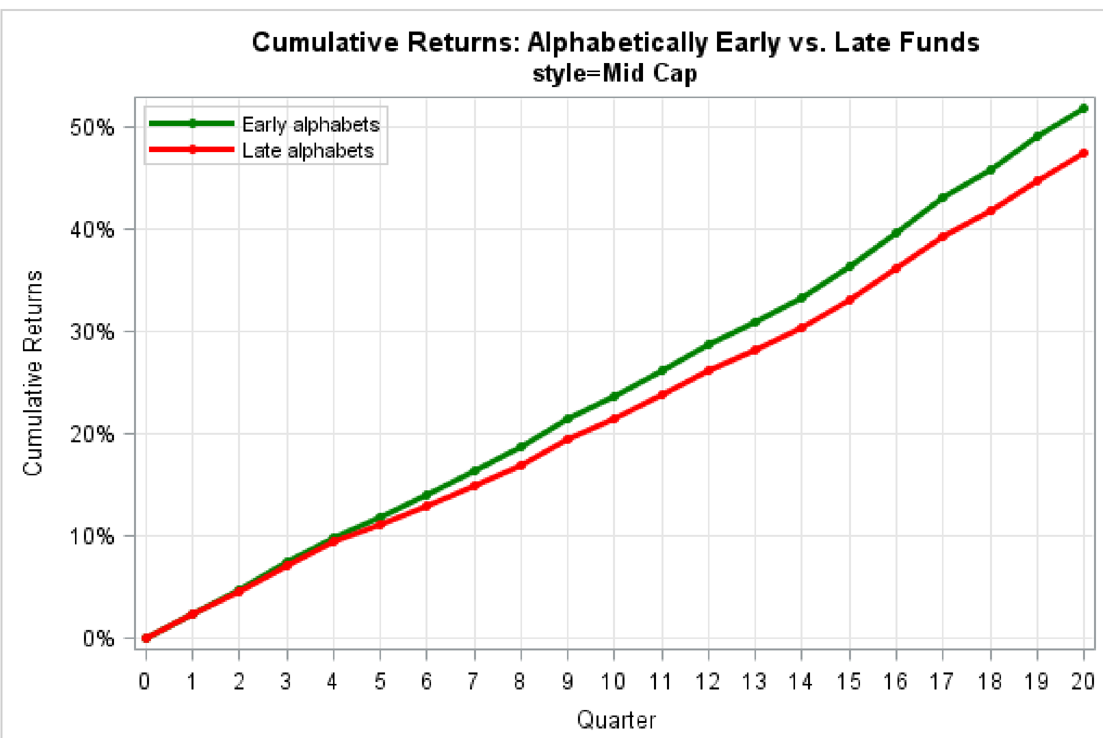
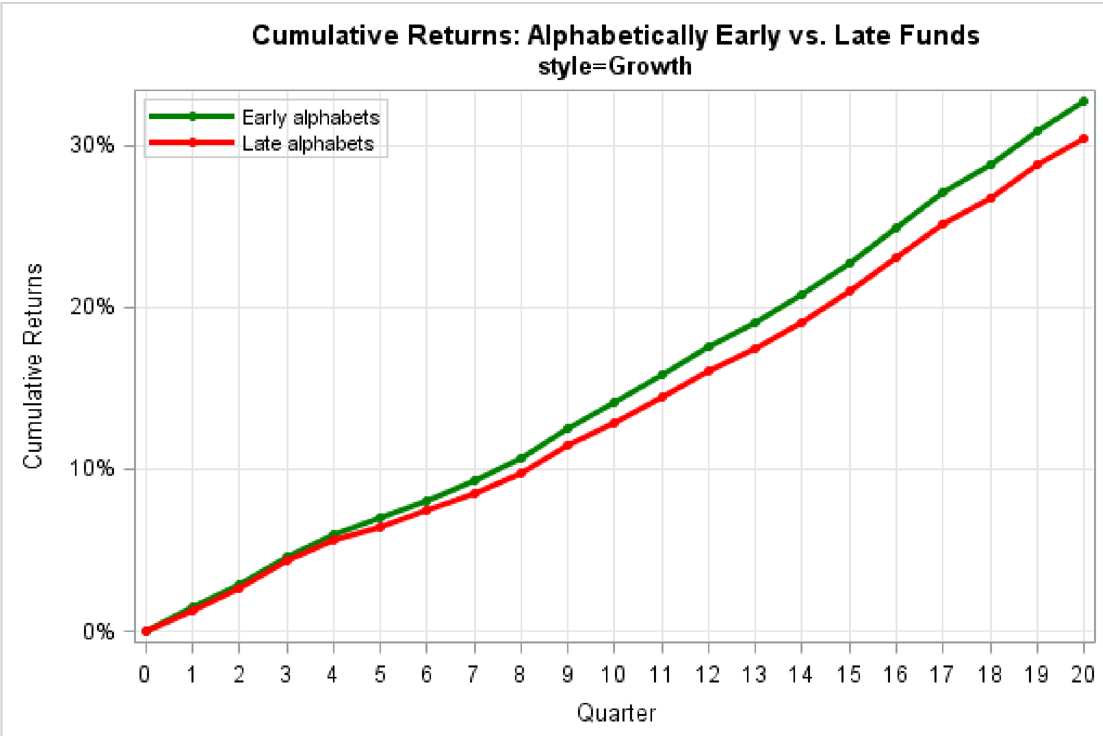


Figure 3.D



Cumulative Returns: Alphabetically Early vs. Late Funds
style=Growth
Early alphabets
Late alphabets
Cumulative Returns
30%
20%
10%
0%
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Quarter