The power of primacy:

Alphabetic bias, investor recognition, and market outcomes

Heiko Jacobs and Alexander Hillert*

First draft: January 2013

This draft: October 2014

Abstract

Extensive research has revealed that the convention of alphabetical name ordering tends to provide an advantage to those positioned in the beginning of the alphabet. This paper is the first to explore implications of this alphabetic bias in the natural setting of financial markets. Most notably, we find that stocks with names early in alphabet have about 5% to 15% higher trading activity and liquidity. These findings are related to firm visibility as well as investor sophistication. International evidence, mutual fund flows and other settings further support the idea that ordering effects are strong enough to affect economic aggregates.

Keywords: Trading behavior, behavioral finance, rank effects, name effects, limited attention

JEL Classification Codes: G02, G12, G14

*Corresponding author: Heiko Jacobs, Finance Department, University of Mannheim, L 5, 2, 68131 Mannheim. E-Mail: jacobs@bank.bwl.uni-mannheim.de. Phone: +49 621 181 3453. Fax:+49 621 181 1534. Alexander Hillert, Finance Department, University of Mannheim, L9, 1-2, 68131 Mannheim. E-Mail: hillert@bwl.uni-mannheim.de. Phone: +49 621 181 1462. Fax:+49 621 181 1664. Parts of this research project have been conducted while the first author visited the School of Banking & Finance at the University of New South Wales. We would like to thank (in alphabetical order...) Tim Adam, Russell Jame, Markku Kaustia, Alok Kumar, Ron Masulis, Ernst Maug, Ralitsa Petkova, Erik Theissen, Martin Weber as well as seminar participants at the 41st Annual Meeting of the European Finance Association 2014, at the Third Helsinki Finance Summit on Investor Behavior 2013, at the 20th Annual Meeting of the German Finance Association (DGF) 2013, at the 76th Annual Meeting of the German Academic Association for Business Research (VHB), at the University of New South Wales, at the Humboldt University of Berlin/ESMT, and at the University of Mannheim.

1. Introduction

Being on or near the top of a list is in many ways important in academia. For example, Brogaard et al. (2014) uncover that articles placed first, second, and third in an issue of top finance and economic journals enjoy all else equal about 50%, 26%, and 17% more future citations than articles published near the back of an issue. The authors conclude that it is not completely clear whether this placement effect reflects the editors' private information about article quality or the readers' limited attention. Other studies, however, point more directly to the latter explanation. Often these phenomena are strongly related to the widespread convention of alphabetization, for example with regard to credits on coauthored publications in economics. Einav and Yariv (2006) and van Praag and van Praag (2008) demonstrate that, due to the increased visibility of first-named authorship, researchers with initials closer to the beginning of the alphabet benefit by various measures of academic success. These even seem to include better chances of receiving tenure at top ten economics departments or prestigious awards such as the Nobel Prize.

There are more examples for this form of "alphabetic bias" in academia. Richardson (2010) reveals that in the case of a well-established medical imaging journal, reviewers whose last name started with an A received almost twice as many review invitations as their colleagues towards the end of the alphabet. The journal relied on software which automatically presented a typically alphabetically sorted list of potential reviewers, just as several thousands of other publications have reportedly done. Editors seemed to "almost certainly" (p. 215) exhibit the natural tendency to invite the first panel of appropriate reviewers. A closely related mechanism applies to positions as visiting academics (Merritt (1999)). Moreover, Meer and Rosen (2011) document a comparatively low probability of alumni with names located at the end of the alphabet donating to their university. This appears to be driven by volunteers, who personally call potential donors in alphabetical order, and often simply run out of time before they reach the end. As the convention of alphabetization is a fact of everyday life, "alphabetic bias" is clearly far from being limited to academia.[1] Collectively, this leads "The Economist" to arrive at the (hyperbolical) conclusion: "Over the past century, all kinds of unfairness and discrimination have been denounced or made illegal. But one insidious form continues to thrive: alphabetism. This [...] refers to discrimination against those whose surnames begin with a letter in the lower half of the alphabet." (Economist (2001), p.13).

This research project is to the best of our knowledge the first to study the role of alphabetization in the specific context of financial markets. We thereby particularly concentrate on the stock market, which offers a natural and promising setting for several reasons.

First, investment decisions in stock markets are economically important. Second, the setting offers a rich set of data on market participants' behavior, control variables, and aggregated outcomes, enabling us to distinguish more precisely among alternative hypotheses than some existing empirical studies in other settings do. Third, due to the thousands of stocks available, investors face a considerable search problem (e.g. Barber and Odean (2008), Merton (1987)), making ordering effects likely to occur. Fourth, there is a widespread convention of listing company names alphabetically in newspapers, indices, data bases, broker statements, watch lists and most other sources of information. Figure 1 provides motivating examples from popular websites.

Please insert figure 1

Bloomberg.com allows to browse through stocks by the first letter of the firm name (on top of the page) or by sectors. Any subsample selected by one of these methods yields an alphabetically sorted list. Russell.com presents the official list of Russell 3000 members on 37 pages with alphabetical ordering of firm names. Similarly, finance.yahoo.com presents

[1] For instance, it drives the chance of getting access to over-subscribed public services (Jurajda and Münich (2010)) or the likelihood of receiving votes on alphabetically ordered ballots (e.g. Bakker and Lijphart (1980), Ho and Imai (2008)).

members of the Nasdaq Composite on 49 pages. For any market segment, nytimes.com offers firm variables like size or proximity to the 52 week high, but again firms with names early in the alphabet are displayed on the first pages.

Against this background, we test the hypothesis that firms towards the beginning of such lists enjoy privileged treatment. Put simply, we comprehensively explore stock-market implications of the intuitive conjecture of "The Economist" (2001, p. 13), which is also backed up by eye movement data (e.g. Lohse (1997)): "It has long been known that a taxi firm called AAAA cars has a big advantage over Zodiac cars when customers thumb through their phone directories." Our findings, whose essence is captured in figure 2, support this claim.

Please insert figure 2

The graph compares trading activity (as measured by monthly turnover) and costs of trading (as measured by the monthly Amihud (2002) illiquidity ratio) for NYSE/AMEX firms grouped by sorting their firm name in alphabetical order. The benchmark in figure 2 is firms whose name is above the 75th percentile in a given month. Findings reveal that firms positioned earlier in the alphabet appear to be strikingly more recognized by investors. For instance, firms in the top 5% have about 13% higher turnover and 11% lower costs of trading than firms in the lowest 25% of the alphabet.

Importantly, the findings illustrated in figure 2 are estimates from regressions controlling for 35 firm characteristics over a time period ranging from January 1995 to December 2009. Controls include, for instance, market capitalization, industry membership, analyst coverage, earnings surprises, national and local media coverage, different types of firm news extracted from 8-K filings, or company name fluency.

Stock characteristics as well as actual trading data from fund managers and a large discount brokerage suggest that alphabetic bias is related to firm visibility as well as investor sophistication. For instance, differences-in-differences tests show that findings tend to be most pronounced among stocks with low analyst coverage, low institutional ownership, low industry market share, low advertising expenditures, or with a high level of retail investor trading activity. These insights lend further support to the idea that "alphabetic bias" is strong and pervasive enough to affect equilibrium outcomes.

We find similar results across firm universes, sample periods, model specifications, and econometric approaches. We also find some (albeit clearly weaker) evidence for the impact of a firm's alphabetical rank on breadth of ownership, on firm valuation as well as on coverage by analysts and the local media. Further evidence for the importance of rank effects comes from the alphabetic distribution of firm names, from the impact of name changes, from an experimental analysis of stock selection, and from international stock markets. With respect to the latter, we find alphabetic bias in large Western countries, but not in Japan, where firms are sorted based on a numerical stock code unrelated to alphabetic ordering. As hypothesized, it is instead the relative position within the numerical code system that predicts trading activity in Japan.

Finally, we turn to the mutual fund market, which offers another promising setting. Again, investors face multiple investment alternatives with many sources of information presenting lists of funds in alphabetical order. Thus, funds with names early in the alphabet may be comparatively more likely to be part of the choice set of investors facing time and capacity constraints. Moreover, the mutual fund business is a trillion dollar market, making the cross-sectional distribution of fund flows economically important. Finally, as the literature review below shows, recent work has uncovered the impact of name effects on mutual fund flows, rendering the ordering of names another important issue. Our empirical analysis from 1992 to 2012 supports this conjecture. The effect of alphabetic fund name ordering is concentrated in small funds for which search costs tend to be higher.

In this subgroup, funds within the top of the alphabet enjoy about 3% to 6% more net inflows p.a., holding all other factors fixed.

As browsing from the top of lists is natural human behavior, it seems likely that the underlying mechanisms of our findings are similar to the ones proposed for the academic settings sketched above. Comparable to the case of multiple authors, being named early on a stock list might boost visibility. Comparable to the case of review invitations and visiting positions, top-ranked stocks might also have a high chance of already satisfying the needs of market participants in search of investment alternatives. As in the case of alumni donations and the other settings, investors might also run out of time, capital, interest, or face other constraints before they reach the bottom of a stock list.

At least for some investors, several other more subtle effects may be at work. For instance, being among the first items on a list may (potentially unconsciously) often be associated with superior quality, such as in the example of finance papers placed first, second, and third in an issue. Confirming evidence is brought forward by Ang et al. (2010) who uncover that investors exhibit affect for class A shares relative to class B shares. Moreover, and similar to the impact of first-named authorship, a stock's permanent visibility may eventually translate into a higher degree of familiarity among investors, which in turn has been shown to be an important determinant of financial decision making (e.g. Huberman (2001), Grinblatt and Keloharju (2001)). Finally, there is a large amount of literature on the primacy effect, which Carney and Banaji (2012) characterize in their literature review as follows (p.1): "What is experienced first is remembered better (...), it drives attachment more strongly (...), creates stronger association with the self (...), influences impressions more decisively (...), and persuades more effectively (...)." Thus, even if market participants worked through a number of potential investments from the top to the bottom of e.g. a typical stock index, they may well eventually choose one of the stocks listed first.

Our study contributes to several strands of literature. First, it adds to the emerging research on name effects in financial markets. This work shows that factors unrelated to fundamentals can influence investor behavior and potentially also market outcomes. In a clinical study, Rashes (2001) uncovers excess comovement between two stocks with very similar ticker symbols and attributes this to confusion among predominantly small investors. Vivid examples for name effects are presented in Cooper et al. (2001) and Cooper et al. (2005b): During (after) the internet bubble, firms that simply changed their name to (away from) "dot.com names" enjoyed positive abnormal post-announcement returns.

Similarly, cosmetic name changes among mutual funds yield positive abnormal inflows without delivering better performance (Cooper et al. (2005a)). Green and Jame (2013) show that funds with more fluent names attract higher inflows. Two recent contribution uncover name effects also at the fund manager-level (Niessen-Ruenzi and Ruenzi (2013), Kumar et al. (2014)): stereotypes associated with female or foreign-sounding fund manager names lead, all else equal, to substantially lower inflows. Our contribution is to show that even the ordering of names, and not the meaning of the names themselves, influences economic aggregates, both in the stock market and the mutual fund market.

Second, our findings add to the understanding of the drivers of stock-level trading activity. For instance, turnover should be the same for all stocks in a CAPM world. However, there are large cross-sectional differences in real markets. In part, these effects are likely to be attributable to differences of opinion (e.g. Harris and Raviv (1993)) or investor overconfidence (e.g. Daniel et al. (1998), Odean (1998), Odean (1999)), yet many other factors are still not well understood (e.g. Chordia et al. (2007), Lo and Wang (2000), or Statman et al. (2006)). Identifying the mechanism behind these puzzling cross-sectional differences can help to enhance our understanding of what makes investors trade.

Several papers have progressed on this front by uncovering that specific firm-level variables deemed to be related to investor recognition and familiarity are positively correlated with stock-level trading activity. For instance, Loughran and Schultz (2004) and Jacobs and Weber (2012) explore the role of firm location. Grullon et al. (2004) focus on advertising expenditures, and Chen et al. (2004) concentrate on index addition effects.

The study most closely related to ours in this respect is Green and Jame (2013). They argue that firms with short, easy to pronounce, rememberable names attract, all else equal, more investors, which translates into higher stock-level turnover, lower transaction price impact, and higher firm value. We later control for the drivers of trading volume studied in all papers mentioned above. Our contribution is to show that a previously neglected and seemingly minor detail, essentially the first letter of a company name, has a surprisingly strong impact. Thus, seemingly innocuous choices by firms can effect how they are traded in financial markets.

Third, our study adds to the vibrant literature on limited investor attention. Due to the complexity of financial markets on the one hand and cognitive constraints on the other hand, market participants have to be selective with regard to information processing. As a consequence, investors cannot take all alternatives into account simultaneously (e.g. Kahneman (1973), Hirshleifer et al. (2009)). Both theory and evidence suggests that investors instead may resort to category thinking or other forms of complexity reducing heuristics, and thereby tend to neglect potentially value-relevant information.[2] Our contribution is to present a novel channel through which attention allocation of different groups of market participants can be assessed.

As the convention on alphabetical ordering is widespread, our findings may have economic implications for several actors in financial markets. For instance, they allow so-

[2] A non-exhaustive list of reference papers (in alphabetical order...) includes Barber and Odean (2008), Barberis and Shleifer (2003), Barberis et al. (2005), Boyer (2011), Cohen and Frazzini (2008), DellaVigna and Pollet (2009), Greenwood (2008), Green and Hwang (2009), Menzly and Ozbas (2010), Mullainathan (2002), Peng and Xiong (2006), and Peress (2008).

phisticated investors to identify stocks which, all else equal, are less costly to trade. Our results might be of relevance to exchange-listed firms in search of comparatively cheap methods to increase liquidity. With regard to the presentation and marketing of their product universe, results might be of interest to e.g. index providers, brokers, banks, or mutual fund companies.

2. Empirical Approach and Baseline Findings

2.1 QUANTIFYING THE DISTRIBUTION OF FIRM NAMES

Our goal is to test for the existence of an "alphabetic bias", which might be strong enough to affect economic aggregates in the context of investment decisions. Thus, referring back to the example sketched above, explanatory variables which quantify the difference between "AAAA cars" and "Zodiac cars" are of primary interest.[3]

We build on two measures, which are both based on the time-series of (historical and current) company names as provided by CRSP.[4] The first measure simply indicates the relative position of a firm's name in the firm universe under consideration (e.g. NYSE/AMEX firms or Nasdaq firms). In each month, we sort all eligible firms on their official company name in alphabetically ascending order. The variable, in the following referred to as *position continuous* and used in the baseline analysis, is then computed as the rank of the

[3] As in this example, we only consider firms with names starting with a letter. This is intended to assure a clean, conservative analysis. About 0.15% of monthly observations in our baseline sample represent firms for which the name starts with a number according to CRSP. Examples include 1st Source Corporation or 3Com. However, findings are very robust with respect to the inclusion of these firms.

[4] In close to 10% of firm months, CRSP reports names whose first letter is followed by a blank space. Examples include "M T S SYSTEMS CORP", "K TRON INTL INC" or "S & K FAMOUS BRANDS INC". In an attempt to letting the data "speak for themselves" and in order to provide conservative estimates, we do not modify these names unless noted otherwise. Eliminating the blank spaces or excluding these firm months from the analysis yields slightly stronger baseline results.

firm divided by the number of all firms. Thus, *position continuous* is uniformly distributed over the interval (0,1]. In general, and as we show in the next section, *position continuous* is essentially uncorrelated with respect to a large set of firm characteristics.

Our second measure takes potential non-linearities into account. The impact of alphabetization might be disproportionately strong for firms at the beginning of a list, and then decay gradually. To capture this line of reasoning mathematically, we introduce several disjunct dummies, in the following referred to as *position dummies*. *Position dummy* 5 takes on a value of one if the company name is among the first 5% in a given month and zero otherwise. *Position dummy* 25 is 1 if the firm name is between percentiles 5 and 25. Finally, *position dummy* 50 (*position dummy* 75) is 1 if the name is between percentiles 25 and 50 (50 and 75). Consequently, findings are benchmarked against the firms near the end of the alphabet (i.e., percentile 75 and above).

The upper half of figure 3 shows the pooled distribution of firm names by first letter. Findings are benchmarked against the distribution of words in general, which we construct from Google's "Ngram corpus" (see Michel et al. (2011) for details). The corpus we rely on is based on the content of hundreds of thousands of digitalized books written in American English and published between 1983 and 2009.

Differences between both word lists are most pronounced at the very beginning of the alphabet. 9.66% of firm names in our baseline sample (CRSP stocks between 1983-2011, see model 1 described in the following section) start with A. This is about 40% higher than the respective relative frequency for words in general. These differences do not vanish quickly once one considers subsequent letters. For instance, 48.7% of firm names start with "I" or earlier letters, while only 44.6% of words do so. The cumulative distribution functions in the lower half of the figure cross each other between "R" and "Q", which closely corresponds to percentile 75 used as a cut-off point for the *position dummies*.

Please insert figure 3

These insights are in line with extensive anecdotal evidence which suggests that at least a non-trivial fraction of firms appears to believe in increased visibility due to rank effects. Newspaper articles have pointed to the possibility of "alphabetical advantages" for a long time (e.g. Milwaukee Journal (1978), Chicago Tribune (1992)). Consistent with this view, several authors argue that the name "Amiga" was chosen in the 1980ies in part to appear before its competitors Apple and Atari (e.g. DeMaria and Wilson (2002)), and that the name "Amazon" was chosen in the 1990ies in part to profit from alphabetically ordered web directories (e.g. Byers (2006)). Beginning firm names with letters early in the alphabet is also sometimes recommended by management consultants (e.g. Kawasaki (2006)). This pattern is indeed clearly visible in the alphabetical distribution of firm names listed in Yellow Pages, which leads Einav and Yariv (2006) to conclude (p. 186): "In fact, this sort of influence on attention appears to be heavily exploited in the realm of advertising."

2.2 BASELINE RESULTS

The dependent variables in our baseline analysis are stock-level measures of trading activity and trading costs. With respect to the first, we focus on firm turnover as it "yields the sharpest empirical implications and is the most natural measure" (Lo and Wang (2000), p. 12). We logarithmize turnover as it is naturally skewed. As our measure of the costs of trading, we rely on the logarithmized Amihud (2002) illiquidity ratio. The ratio is computed as the monthly average of daily illiquidity measures, defined as the absolute daily stock return divided by daily trading volume in millions of dollars.[5]

Due to the differences in market structure and the well-documented double-counting

[5] Winsorizing turnover and the illiquidity ratio at the 99% level does not change any insights.

issue for stocks trading on Nasdaq (e.g. Atkins and Dyl (1997)), we run our analysis separately for NYSE/AMEX and Nasdaq stocks. Similar patterns across these disjunct subsamples would moreover suggest that our findings represent a more generalized phenomenon. To additionally estimate an overall effect across exchanges, we follow the method suggested in e.g. Loughran and Schultz (2005) by doubling trading volume for NYSE and AMEX stocks, pooling them with Nasdaq stocks, and adding a Nasdaq dummy to the regression.

We start with the framework proposed in Chordia et al. (2007), which aims at comprehensively identifying cross-sectional determinants of expected trading activity. The econometric approach involves running predictive regressions of a measure of trading activity for stock i in month t+1 ($Trading_{i,t+1}$) on up to n sets of lagged explanatory firm characteristics ($\sum_{k=1}^{n} Set_{k,i,t}$).

$$Trading_{i,t+1} = \gamma_{0,t} + \sum_{k=1}^{n} \gamma_{k,t} Set_{k,i,t} + \epsilon_{i,t+1} \tag{1}$$

$\gamma_{k,t}$ refers to the vector of coefficients obtained for the control variables of set k in month t. We adopt and extend the two sets of control variables considered in Chordia et al. (2007). In total, we consider two dependent variables (logarithmized turnover, logarithmized illiquidity ratio), three samples (NYSE/AMEX, Nasdaq, all) and three different models (with different controls and sample starts ranging from 1963 to 1995).

In order to provide conservative estimates, statistical inference in these 18 specifications is based on predictive panel regressions with year dummies and standard errors which are double-clustered by firm and month. Petersen (2009) argues that double-clustering produces correctly sized standard errors, regardless of whether a potential firm effect is permanent or temporary.

To comprehensively control for other variables, the empirical analysis combines data from several sources. The CRSP/Compustat Merged Database provides common stock as well as balance sheet information. Among others, we augment this data set with analyst data obtained from I/B/E/S, extensive local and national firm-level media coverage gathered from LexisNexis, and all 8-K filings compiled from EDGAR.

In each test, we only consider stock months for which all control variables are available. Untabulated findings show that this leads to our sample firms being slightly larger than the standard CRSP common stock universe.

Table 1 provides details regarding the motivation, the measurement, and references for all control variables which we briefly mention in the following.

Please insert table 1

The first set of lagged explanatory variables in Chordia et al. (2007) has the advantage of long data availability so that the beginning of the sample period can be set to July 1963 for AMEX and NYSE stocks. In comparison, the sample period starts in January 1983 for Nasdaq stocks, when trading volume data for these stocks becomes comprehensively available in CRSP. We refer to this setting as model 1. Controls consists of two variables for past signed returns, the book-to-market ratio, the firm's market capitalization, its age, its nominal share price, its leverage, and its beta. Panel A of table 2 presents main findings with regard to name ordering-induced cross-sectional differences in turnover (illiquidity).

Please insert table 2

We find that *position continuous* indeed matters for the cross-section of both trading activity and the costs of trading. For NYSE/AMEX (Nasdaq) stocks since 1963 (1983), the implied turnover increase for firms at the beginning of the alphabet relative to their counterparts at the end is all else equal about 10% (7%). Across exchanges, it is 9%. For the Amihud (2002) illiquidity ratio, findings are similar. Findings suggest that firms listed early in the alphabet have about 8% to 10% lower trading costs. These values are not only highly statistically significant, but also economically meaningful. We now test whether the effect remains persistent when including more comprehensive control variables.

For model specification 2, we add the number of analysts following a stock, analyst forecast dispersion, earnings surprises, and earnings volatility. While this specification is equivalent to the full model specification in Chordia et al. (2007), we consider additional variables shown to be related to trading activity. Given our unusually large number of controls, the following tests might be considered to be particularly conservative.

More specifically, we employ the return twelve months ago, a 52-week high dummy, squared market capitalization, dummies for S&P 500 membership, Dow Jones 30 membership (for NYSE/AMEX stocks), Nasdaq 100 membership (for Nasdaq stocks), idiosyncratic volatility, alpha, advertising expenditures, research and development expenditures, dummies for firms in urban and rural regions, the number of employees, and a set of dummies to control for the 49 Fama and French (1997) industries.

Model 2 relies partly on lagged monthly analyst (and other) data, which is not reliably available before 1980 (e.g. Hong et al. (2000)). Therefore, Model 2 starts in February 1980 for the NYSE/AMEX sample and again in January 1983 for Nasdaq stocks. Main findings for turnover (illiquidity) are displayed in panel B of table 2. The impact of *position continuous* remains very stable, thus inferences do not change. The impact of alphabetic name ranking on turnover (illiquidity) is estimated to range roughly from 9% to 10% (7% to 10.5%).

The major goal of model 3 is to control more comprehensively for firm-specific news and investor recognition, both of which are likely to be positively related to stock-level trading activity and liquidity. We include controls for in total about 1.15 million firm-level news articles in four national and 41 local newspapers, and moreover add state dummies. In addition, we construct in total ten variables which quantify different types of important firm news by relying on a parsing algorithm and 421,000 8-K filings gathered from EDGAR. 8-K filings are available from 1995 onwards, and coverage for many local newspapers also does not start before the middle of the 90ties. Therefore, the sample period for model 3 starts in February 1995.

Finally, we use data provided by Green and Jame (2013) to also control for the recently discovered impact of company name fluency. Fluency and alphabetic rank clearly represent distinct phenomena. The correlation between *position continuous* and the fluency score ranges from roughly -0.01 to about 0.03, depending on the time period and firm universe. Also from a conceptual point of view, both measures are very different. We are essentially testing for rank effects, which in the U.S. stock market are likely to manifest themselves in an "alphabetic bias". However, in the presence of a different ordering mechanism (such as in Japanese stock market) rank effects will be completely unrelated to firm names. Nevertheless, section 3.3 uncovers that the type of rank bias we document is particularly strong among the subsample of U.S. firms with high name fluency, which is consistent with the idea that investor sophistication is an underlying driver of both effects. Lagged company name fluency is available until December 2009, which consequently marks the end of the sample period for model 3.

Main findings for turnover (illiquidity) in this full model specification are displayed in panel C of table 2. It turns out that the explanatory variable of interest, *position continuous*, quantifies rank effects which are even slightly more pronounced than our estimates based on models 1 and 2. For instance, in the combined NYSE/AMEX/Nasdaq specification, the implied increase in turnover (decrease in trading costs) for firms positioned at the beginning of the alphabet is about 10.8% (10.7%), holding the 35 firm characteristics in model 3 constant. With t-statistics exceeding 4 (3), findings are also statistically highly significant.

As the second column of table 3 shows exemplarily, our other findings are broadly in line with insights from previous work (e.g. Chordia et al. (2007)).

Please insert table 3

Moreover, and as the first column of table 3 shows in more detail, the relative position of a firm's name in the alphabet appears to be reasonably random. For instance, the R^2 from a regression of *position continuous* on all firm characteristics (including industry dummies) of model 3 is just 0.025. The online appendix shows similar insights from univariate comparisons.[6]

3. Further Insights

3.1 ROBUSTNESS CHECKS

Alternative measure of firm identification We have replicated the analysis with ticker symbols instead of firm names. Findings turn out to be very similar. This is not surprising as in roughly 95% of all firm months, the first letter of the company name equals the first letter of the ticker symbol. A regression framework aimed at exploiting the cases where names and ticker symbols differ suggests that both are equally relevant from a statistical point of view.

Alternative measure of ordering Table 4 reports findings when we rely on disjunct *position dummies*, as described in section 2.1, to measure the relative position of a firm's name. Again, we run predictive panel regressions and report standard errors which are double-clustered by firm and month. In line with the idea of investors browsing from the top of lists of alphabetically sorted firm names, firms at the very beginning of the alphabet seem to be disproportionately more recognized by investors than all other firms.

[6] From a methodological point of view, alphabetic ordering thus may be useful as an instrument for addressing endogeneity in applications where stock-level trading activity or liquidity is a key explanatory variable.

Please insert table 4

For instance, estimates from the full model specification across exchanges indicate that the first 5% of firms have, all else equal, 11% higher trading activity and 12% lower costs of trading than their counterparts with names in percentiles 75 and above. The effect becomes smaller for firms later in the alphabet. For instance, while firms with names in percentiles 5 to 25 still enjoy more than 8% higher turnover and more than 6% lower costs of trading, there is hardly any effect for firms with names in percentiles 50 to 75. In sum, findings strongly support the insights from the baseline analysis.

Alternative firm universe In their stock selection process, market participants might not consider the whole stock universe, but apply preceding filter rules. Such screens may be related to stock exchanges, as in our baseline analysis. Arguably another popular way of restricting the firm universe is to focus on specific industries. To explore the idea that alphabetization-induced ordering effects should also be identifiable in these subgroups, we rerun the analysis for seven broad sectors (with at least 40 eligible firms in each month) as defined by SIC divisions. Dependent variables are again either turnover or illiquidity. Moreover, we again consider models 1 to 3 as well as both turnover and illiquidity. The online appendix reports the main findings from predictive panel regressions with double-clustered standard errors.

In 95% of the in total 42 regressions, the coefficient of *position continuous* goes in the predicted direction. Findings are also economically meaningful: about 90% of the regressions estimate the impact of *position continuous* to be at least 5%. Moreover, results are fairly robust across model specifications. In sum, findings strongly support the insights from the baseline analysis.

Alternative econometric approach We change the econometric approach by running predictive monthly Fama and MacBeth (1973)-type regressions, which corresponds to the method used in Chordia et al. (2007). As insights are similar across models and in order to conserve space, panels A and B of table 5 only report the results for the full regression specification (model 3).[7] In any setting and for both turnover (panel A) and liquidity (panel B), coefficients are statistically significant at the one percent level. Also the economic magnitude is very similar to the estimates of our baseline estimates. All else equal, estimates suggest a turnover (liquidity) difference of 7% to 12% (9% to 12%) between firms with names early in the alphabet and firms with names late in the alphabet. In the overall picture, inferences from the baseline analysis thus remain unchanged.

Please insert table 5

Alternative approach: name changes As they allow us to focus on within-firm variation, company name changes offer a conceptually different setting to examine the role of alphabetic ordering. In an attempt to balance the trade-off between a sufficiently large sample size and a comprehensive set of control variables, we rely on model 2 for the NYSE/AMEX/Nasdaq universe. We only consider firm name changes which go along with an absolute change in *position continuous* of at least 0.25. The average name change in this sample of in total 774 events goes along with an increase or decrease in the relative ranking of about 50%.

As panel C of table 5 shows, to some extent there seem to be patterns in the direction of the name changes which complement the analysis of the alphabetic distribution of names presented in figure 3. (Particular extreme) name changes tend to lead to a position earlier (later) in the alphabetical ranking, even though most of the estimates are not statistically significant.[8]

[7] The third column of table 3 exemplarily also tabulates the control variables.

[8] In order to be conservative, we eliminate blank spaces in firm names reported in CRSP for this analysis, as not doing so leads to stronger results (see also footnote 4).

Besides the material changes, in which we are primarily interested, within-firm variation in *position continuous* has other sources as well. From month to month, there are often relatively small changes induced by variation in the eligible firm universe, name changes of other firms, or name changes which change *position continuous* by less than 25%. To separate these effects, we first condition on those events for which we have at least 24 non-missing monthly observations in the 60 month periods both before and after a name change of at least 25%. We average *position continuous* both before and after each event in order to capture the impact of material name changes. Small month to month changes in *position continuous* are captured by *small change* which is one (minus one) in the case of a position earlier (later) in the alphabetical ranking and zero otherwise.

We then run panel regressions as before, with all control variables of model 2 (including month-fixed effects) and standard errors clustered at the firm level. However, we exclude the month of the name change, and introduce firm-fixed effects as well as a dummy which quantifies the period after the name changes.

Our central finding, displayed in panels D and E of table 5, is that the inferences from the between-firm baseline analysis also hold for the within-firm analysis. With turnover (the Amihud (2002) measure) as the dependent variable, the coefficient of *position continuous* is -0.142 (0.150) and statistically significant. The coefficients for *small change* have the predicted direction, but are (as expected) statistically not significant and economically small.

Finally, excluding all firm months from the name change analysis does not change any of the insights obtained from our baseline cross-sectional tests.

3.2 BREADTH OF OWNERSHIP, FIRM VALUATION, AND EXPECTED RETURNS

Firms with names at the beginning of the alphabet might also have higher breadth of ownership, holding all else equal. We thus run panel regressions similar to the ones in table 2, but we now use the logarithmized total number of shareholders as our dependent variable. Due to limited data availability, we run yearly regressions for all models. Panel A of table 6 shows the findings for the NYSE/AMEX/Nasdaq universe.

Please insert table 6

Regardless of whether we use *position continuous* or *position dummies*, firms listed early in the alphabet appear to have indeed ceteris paribus about 7% to 9% more shareholders, even when the Amihud illiqudity ratio is included as additional control.

Finally, previous work (e.g. Merton (1987)) indicates a positive relation between breadth of ownership and firm valuation. We use the logarithmized market-to-book ratio as a proxy for a firm's valuation and run regressions as specified above. Panel B of table 6 provides some supportive results. Firms with names at the very beginning of the alphabet seem to trade at about 5% higher valuations although this effect is not persistently significant. Moreover, the impact of alphabetic ordering on firm valuation goes beyond of what seems attributable to name-induced differences in the Amihud illiqudity ratio and the number of shareholders.

In sum, results for breadth of ownership and firm valuation are somewhat weaker than our baseline findings for turnover and liquidity. This is also reflected in (untabulated) name change tests for which we do find statistically significant relationships.

The slightly higher valuations for firms early in the alphabet raise the question to what extent "alphabetic bias" matters for expected returns. However, our insights in this respect are not different from what for instance Hong et al. (2008) or Green and Jame (2013) find for their analysis of the stock-level impact of local bias and name fluency bias, respectively. Both papers provide some illustrative calculations whose results Hong et al. (2008) summarize as follows: The difference in expected returns implied by local bias-driven valuation differences (similar in magnitude as ours) is "simply too small an effect to pick up with a standard return-forecasting exercise" (p. 37). Similarly, Green and Jame (2013) find "no significant relation between fluency and returns" (p. 823).

These judgements are confirmed by our (untabulated) Fama/MacBeth regressions with the full set of control variables (including or excluding liquidity and breadth of ownership), as they are unable to reliably detect an alphabetization-induced return difference. One interpretation of these findings is that firms early in the alphabet enjoy privileged treatment for instance in that their stocks can be traded at substantially lower transaction costs impact without measurable negative consequences for expected returns. Our insights also suggest that the striking cross-sectional patterns we uncover in this study are unlikely to be eliminated in the near future, as they do not offer easily exploitable arbitrage opportunities.

3.3 CROSS-SECTIONAL DETERMINANTS: FIRM VISIBILITY AND INVESTOR SOPHISTICATION

Rank effects may be strongest among otherwise less visible firms. To empirically test this hypothesis, we use our full model (i.e., NYSE/AMEX/Nasdaq universe, specification 3) although inferences do not depend on this choice. We then construct the following four measures of firm visibility, add them individually to the regressions with turnover or illiquidity as dependent variables, and interact them with *position continuous*: sales (e.g. Hong et al. (2008)), analyst coverage (e.g. Baker et al. (1999)), industry market share (e.g. Hou (2007)), and advertising expenditures (e.g. Grullon et al. (2004)). The average pairwise correlation between the four measures is only 0.23, suggesting that they capture different facets of firm visibility. Table 7 shows the interaction effects and gives an estimation of their economic significance.

Please insert table 7

All interaction effects go in the prediction direction. Most coefficients are statistically significant and moreover suggest economic effects of substantial magnitude. For instance, at the 75th percentile of the distribution of industry market share, the turnover (illiquidity) difference between firms with names at the beginning and at the end of the alphabet is estimated to be only 7% (6%). Moving to the 25th percentile increases these numbers to 15% (16%).

Another plausible conjecture is that our findings might be most pronounced in stocks with a high fraction of individual investors. This subgroup is widely considered to exhibit more biases, to face more cognitive constraints and to have less resources and knowledge than professional investors (see e.g. Barber and Odean (2013) or Grinblatt and Keloharju (2001)).

We explore the impact of cross-sectional differences in the investor base in two ways. First, we focus on subgroups of stocks which are known as a natural habitat of less sophisticated investors. Note that this implies that the effect can actually be stronger for relatively visible stocks provided that their investor base disproportionately consists of biased investors. Second, we directly study individual investors' trading decisions and portfolio holdings.

The trading of individual investors has been shown to be concentrated in small stocks (e.g. Kumar and Lee (2006), Dorn et al. (2008)), in stocks with lottery-type features (Kumar (2009)), in stocks covered by the media (e.g. Barber and Odean (2008), Engelberg et al. (2012)), in stocks with fluent names (Green and Jame (2013)), and in stocks with low institutional ownership (e.g. Lee et al. (1991), Lemmon and Portniaguina (2006)). We thus partition our sample along these dimensions.

Again, we rely on the most comprehensive model specification. We divide the sample into small and large stocks based on the median market capitalization in the previous month. As a second approach, we distinguish between lottery and non-lottery stocks, for which we follow the approach developed by Kumar (2009). Lottery firms (non-lottery firms) are firms with above (below) median idiosyncratic volatility, above (below) median idiosyncratic skewness, and below (above) median nominal share price. A third approach relies on (national) press coverage. As media coverage is highly positively correlated with firm size (e.g. Fang and Peress (2009)), we first orthogonalize the monthly number of articles with respect to lagged market capitalization before we perform a median split. We also distinguish between firms with fluent names (fluency score three and above) and non-fluent names (fluency score two and below), which roughly results in a 40/60 split of firm months. Finally, the fifth approach distinguishes firms with above and below median lagged institutional ownership. We gather quarterly ownership information for institutional managers with at least 100 million USD in assets under management, as reported on Form 13F filed with the SEC obtained via the Thomson-Reuters Institutional Holdings (13F) Database. For these four subsamples, we then run panel regressions as in table 5. Table 8 shows the main findings.

Please insert table 8

The results strongly confirm our predictions. Most notably, small firms (firms with low institutional ownership, firms with fluent names) positioned early in the alphabet are estimated to have all else equal about 14.5% (17%, 17.5%) higher turnover than corresponding firms late in the alphabet, while the corresponding estimate for large firms

(firms with high institutional ownership, firms with nonfluent names) is about 7% (3%, 7%). The findings for lottery stocks and national media also go in the predicted direction, but are less in magnitude.

In the overall picture, findings are even stronger in the case of illiquidity. Firms which have a name within the top of the alphabet and are at the same time likely to be a natural investment habitat of individual investors are about 14% to 20.5% more liquid than corresponding firms at the end of the alphabet. For large firms, non-lottery firms, firms neglected by the media, or firms with a high fraction of institutional investors, the effect is about 1.5% to 9% and in most cases statistically not distinguishable from zero.

We next turn to study the behavior of individual investors directly, and contrast their trading patterns with the behavior of fund managers as well as with the overall market. To this end, we compute the retail breadth of ownership and retail turnover by using data from a large discount broker from 1991 to 1996 (see Barber and Odean (2000), Barber and Odean (2001), and Kumar (2009) for more information). Due to the early sample period, only model 1 or 2 come into question for this analysis. We compute analogous measures for mutual fund managers based on the Thomson-Reuters Mutual Fund Holdings database. Panel A of table 9 shows the findings if we focus on retail investors.

Please insert table 9

The coefficients on *position continuous* suggest that firms early in the alphabet enjoy roughly 14% to 16% more retail trading activity. In contrast, the corresponding estimates for unconditional trading activity over the same time period are about 8% to 9% for the overall market, and 6% to 8% for mutual fund managers. Similarly, findings indicate that firms early in alphabet have about a 6% to nearly 8% higher breadth of retail ownership, whereas the corresponding estimates for market-wide or mutual fund breadth of ownership are 4% to nearly 6% during that time. In sum, individual investors appear to suffer more from “alphabetic bias” than other market participants.

Panel B of table 9 shows the findings if we benchmark mutual fund managers against the overall market. We can base this comparison on a sample period from 1983 to 2011 (for model 1 and model 2) or 1995 to 2009 (for model 3). While mutual fund managers also show an “alphabetic bias” in their holding and trading decisions, the effects are smaller than for the average market participant (as proxied by the estimates for the overall market). Under the assumption that mutual fund managers (retail investors) represent relatively (un)skilled investors, our tests suggest that “alphabetic bias” is all else equal negatively related to investor sophistication.

3.4 OTHER MARKET PARTICIPANTS: ANALYSTS AND THE MEDIA

It is worth noting that our regressions so far included controls for analyst coverage, as well as for national and local media coverage. We may now ask ourselves whether analysts and the media themselves might exhibit a tendency to cover firms which are, due to their early position in the alphabet, attention-grabbing. To explore this hypothesis, we run panel regressions with double-clustered standard errors, thereby concentrating on the full model and the whole stock universe.

As panel C of table 9 shows, the data partly support this hypothesis. In the case of the number of analysts as the dependent variable, the statistically significant results indicate that firms named early in the alphabet receive ceteris paribus about 3.6% more coverage than firms at the bottom of the alphabet. About one third of this effect is attributable to the higher liquidity and higher breadth of ownership of firms early in the alphabet. There is no effect for the national media. However, firms at the beginning of the alphabet enjoy about 4% more local media coverage, even after controlling for visibility-induced differences in Amihud illiquidity and breadth of ownership. Given our insights on investor sophistication as outlined above, one way to interpret these findings is to conjecture that the local media might be more influenced by non-fundamental factors than national media such as the Wall Street Journal (see e.g. Gurun and Butler (2012)).

3.5 INTERNATIONAL EVIDENCE ON ORDERING EFFECTS

Our interpretation of the findings so far rests on the assumption that investors facing a number of investment alternatives concentrate more on the alternatives offered first. To additionally validate this conjecture, we turn to international stock markets. More specifically, we contrast the pooled trading behavior in the five largest Western markets (Canada, UK, France, Germany, Australia) with the trading behavior in Japan.

In contrast to the Western countries, the presentation format in Japan is typically determined by numerical ticker symbols, and not by the Roman alphabet. Stock codes are issued by Japan's national numbering agency and can generally be understood to be assigned based on the order when the stock was listed. Moreover, the ordering contains a strong industry component. We control for these effects by assessing the relative rank of a stock based on the numerical stock code for each industry separately and by including control variables such as age, market capitalization, and the number of analysts. The Japanese version of *position continuous* turns out to be essentially uncorrelated with *position continuous* as implied by alphabetically ordered firm names. By construction, it is also uncorrelated with industry classification which we measure using the Datastream level 4 industry classification.

We then run regressions as before. The selection of the in total 17 controls is motivated by model 2 for the U.S. market, but restricted to those variables which are broadly available for international stocks. We rely on data from Datastream, Worldscope, and I/B/E/S. The controls are listed in the description of table 10, and details about the sample construction are reported in the online appendix. In contrast to the U.S. market, reliable and comprehensive stock-level time-series of historical names do not seem to be available for international stock markets. We thus focus on the most recent ten year period (January 2004 to December 2013), and base our analysis on the current names, which runs counter finding significant rank effects. Table 10 shows the main findings.

Please insert table 10

Both with respect to turnover (panel A) and illiquidity (panel B) and both with respect to *position continuous* and *position dummies*, the estimates for the pooled Western countries are very similar to the ones obtained for the U.S. market. Differences between firms early and late in the alphabet are in the area of 10% and highly statistically significant. This out-of-sample evidence thus further strengthens the insights obtained from the U.S. market.

In contrast, and as hypothesized, we do not find evidence for alphabetic bias in Japan. However, and again as predicted, the implied turnover (liquidity) difference between firms at the beginning and the end of the numerical stock code list is about 20% (15%), holding all other factors fixed. As in the U.S. stock market and the other Western countries, further tests with *position dummies* confirm that the firms at the very beginning of a typical stock list profit the most. Collectively, the analysis of international stock markets provides strong support for the idea of rank effects in stock selection.

3.6 EXPERIMENTAL EVIDENCE ON ORDERING EFFECTS

We run an experiment designed to mirror the presentation format in popular stock information sources (see e.g. the examples in figure 1). The experimental setting also allows us to isolate the impact of pure ordering effects, holding firm and investor characteristics fixed. More specifically, we present subjects a four-page list of unnamed stocks in randomized order, together with actual firm characteristics known to influence trading behavior. We ask subjects to indicate which stocks they would like to trade. Each stock in the first half of the list has a virtually identical twin in the second half, which however is not known to the subjects. In the absence of ordering effects, the likelihood of being selected should be very similar for both constituents of a given pair. However, and as the online appendix shows in more detail, the twin presented first receives far more attention than the twin towards the end of the list. In general, findings reveal strong ordering effects in stock selection.

4. Another Setting: Mutual Fund Flows

The mutual fund industry offers another promising setting in which to study the determinants of investor behavior. As in the stock market, investors have to choose from a large set of alternatives. This holds true even if one narrows down the decision to well-defined and relatively homogeneous market segments, such as open-end U.S. equity mutual funds with a domestic investment focus, as we do in the analysis below. Averaged across all months in our sample from January 1992 to December 2012, there are close to 1,500 funds which satisfy all data requirements. Thus, search costs are likely to play a role in the investment process. Indeed, Sirri and Tuffano (1998) test and verify the hypothesis that "consumers would purchase those funds that are easier or less costly for them to identify" (p. 1607). A substantial body of empirical work further provides support for this idea of cognitive overload in the mutual fund investment decision.[9]

Moreover, the mutual fund market is very large, making capital allocations economically important. In the average month of our sample, the combined assets under management total $1.69 trillion. Thus, if "alphabetic bias" does play a role in this large and

[9] A non-exhaustive list includes Barber et al. (2005), Cooper et al. (2005a), Haslem (2012), Hortacsu and Syverson (2004), Jain and Wu (2000), Koehler and Mercer (2009), and Solomon et al. (2014).

liquid market, then this type of ordering effect is likely to matter in general for financial decision making.

For the empirical analysis, we rely on the CRSP survivorship bias free mutual fund database. We use the third and fourth character of the CRSP style code to assign each fund month to one of in total 19 styles in our sample.[10] We use parsing algorithms (e.g. Gil-Bazo and Ruiz-Verdu (2009)) and manual screens to exclude index funds. We moreover drop very small funds that never have net total assets of over $5 million during the sample period. Finally, we exclude observations if any of the control variables (see below) cannot be computed. To avoid multiple counting, we aggregate all share classes of the same fund using MFLinks (e.g. Daniel et al. (1997), Kacperczyk et al. (2008)). The sample period begins in January 1992, when all (partly lagged) variables relied on are available in the desired frequency.

The dependent variable is the net inflow or outflow for fund i in month t. Following the consensus in the literature, a fund flow is defined as the percentage change in total net assets ($TNA_{i,t-1}$) which is not driven by the fund's return net of fees ($ret_{i,t}$):[11]

$$Fund\ flow_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1}}{TNA_{i,t-1}} - ret_{i,t} \tag{2}$$

The independent variables of interest are again measures which indicate the position of a given fund name in the universe of eligible funds. Consequently, we rely on *position continuous* as well as on *position dummies* both of which are defined as for the

[10]The CRSP style codes combine the information contained in Wiesenberger Objective Codes (1962-1993), Strategic Insight Objective codes (1993-1998), and Lipper Objective codes (from 1998 on), all of which have been widely used in previous work. The most common style classifications are "small-cap funds", "mid-cap funds", "growth funds", "growth and income funds", "equity income funds", "technology sector funds" and "real-estate sector funds".

[11]In order to mitigate the impact of outliers and to be closer to the normality assumption, we winsorize the distribution at the 2.5% and 97.5% level in each month separately. Using alternative cut-offs (1%, 5%, 95%, 99%) does not change inferences.

analysis in the stock market. We rely on the fund name reported by CRSP.[12]

The selection of control variables is guided by the large amount of previous work on the determinants of fund flows. The computation of these variables is explained in detail in table 11. At the fund-level, we consider the following characteristics: size, age, expense ratio, turnover ratio, past return and risk, and number of share classes. To control for the well-established convex relationship between fund performance and fund flows (e.g. Chevalier and Ellison (1997), Barber et al. (2005)), we take additional style-specific performance measures into account. Specifically, for each year and each of the 19 investment objectives separately, we determine the relative performance position of each fund. The resulting performance rank variable is evenly distributed between 0 and 1. We allow for a non-linear effect by also including the squared value of the performance rank. Finally, as a parsimonious approach inspired by Green and Jame (2013), we compute the length of the fund name defined as the number of letters, after dropping share-class information and incorporation terms when applicable.

At the fund family-level, we compute the total assets under management as well as the number of offered funds. We also include a dummy for the three largest families in each month. At the style-level, we compute growth rates from the value-weighted flows of all funds with the same investment objective in a given month. In addition, all regressions include style-year fixed effects. Standard errors are double-clustered by fund and month.

In untabulated findings, we have verified that the position measures are weakly correlated with any of the other explanatory variables. Moreover, there is virtually no relation between position measures and future fund returns. Thus, the regression setting seems well-suited to isolate the impact of pure ordering effects induced by alphabetization of

[12]Typical names are "AAL Capital Growth Fund", "Buffalo Small Cap Fund", or "Navellier Performance Funds: Navellier Aggressive Micro Cap Portfolio". Sometimes, CRSP reports a "the" in front of the actual fund name. A typical example is "The Glenmede Fund, Inc.: Glenmede Large Cap Value Portfolio". In these cases, we drop the "the", which however has little impact on our findings.

fund names. Table 11 shows the main findings.

Please insert table 11

We run regressions separately for all funds (models 1 and 2) as well as for small funds (models 3 and 4) and large funds (models 5 and 6), determined by a monthly median split. Model 1 reveals that the impact of *position continuous* is as expected: funds with names at the beginning of the alphabet generate ceteris paribus about 0.16% higher inflows each month. This effect is statistically significant and comparable to moving from the 50th percentile of average mutual fund flows to the 54th percentile.

Models 3 to 6 uncover that the effect is driven by small funds with names at the very beginning of the alphabet. This is in line with the limited attention hypothesis as small funds are likely to be less visible than large funds (see e.g. the argumentation in Sirri and Tuffano (1998)). The first 5% of smaller than median funds benefit by far the most. As model 4 shows, they achieve more than 0.5% higher inflows per month (or more than 6% higher inflows per year) than small funds towards the end of the alphabet. This highly significant effect is equivalent of moving from the 50th percentile of the distribution of flows for below median-sized sample funds to the 60th percentile. Even in this subsample, the average fund still has about $80 million under management. Thus, findings are important from an economic perspective.

The online appendix shows that inferences remain qualitatively unchanged if we use alternative ways to control for style, fund family, and time effects. In the overall picture, our insights from the stock market can thus be transferred to the mutual fund market. Being listed early in the alphabet leads to economically substantial benefits for names early in the alphabet.

5. Conclusion

Sorting names alphabetically is an omnipresent convention. For many settings such as academia, economics, or politics, this widespread practice has been shown to yield an advantage to those positioned early in the alphabet. This paper is the first to analyze this type of "alphabetic bias" in financial markets.

We find that a higher alphabetic ranking provides stocks with higher share turnover, investors with lower transaction costs, and firms to some extent with broader ownership, higher valuations, and higher recognition among analysts and the local media. Digging deeper, we find that alphabetic bias is most pronounced for stocks which are otherwise less visible or which are disproportionately traded by less sophisticated investors. The cross-section of mutual fund flows and of stock-level trading activity in international markets lends further support to the idea that alphabetical ordering matters for investor decision making and has economically important consequences for market outcomes.

These novel findings might have implications for several actors in financial markets. For instance, in an attempt to affect consumer choice, many firms are willing to pay for their products to be displayed in an attention-grabbing matter (e.g. Amstrong et al. (2009)). In the context of financial markets, our findings reveal that a seemingly minor detail such as essentially the first letter of a name can have a similar impact free of charge.

References

Amihud, Y., 2002, "Illiquidity and stock returns: Cross-section and time-series effects," *Journal of Financial Markets*, 5, 31–56.

Amstrong, M., J. Vickers, and J. Zhou, 2009, "Promince and consumer search," *RAND Journal of Economics*, 40, 209–233.

Ang, J., A. Chua, and D. Jiang, 2010, "Is A better than B? How affect influences the marketing and pricing of financial securities," *Financial Analysts Journal*, 66, 40–54.

Atkins, A. B., and E. A. Dyl, 1997, "Market structure and reported trading volume: NASDAQ versus the NYSE," *Journal of Financial Research*, 20, 291–304.

Baker, H. K., G. E. Powell, and D. G. Weaver, 1999, "Does NYSE listing affect firm visibility?," *Financial Management*, 28, 46–54.

Bakker, E., and A. Lijphart, 1980, "A crucial test of alphabetic voting: the elections at the University of Leiden, 1973-1978," *British Journal of Political Science*, 10, 521–525.

Bandarchuk, P., and J. Hilscher, 2013, "Sources of momentum profits: Evidence on the irrelevance of characteristics," *Review of Finance*, 17, 809–845.

Barber, B., and T. Odean, 2001, "Boys will be boys: gender, overconfidence, and common stock investment," *Quarterly Journal of Economics*, 116, 261–292.

Barber, B. M., and T. Odean, 2000, "Trading is hazardous to your wealth," *Journal of Finance*, 55, 773–806.

——— , 2008, "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors," *Review of Financial Studies*, 21, 785–818.

——— , 2013, "The behavior of individual investors," *Handbook of the Economics of Finance*, Elsevier, edited by Constantnides, Harris, and Stulz, 1533–1569.

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

Barberis, N., and A. Shleifer, 2003, "Style investing," *Journal of Financial Economics*, 68, 161–199.

Barberis, N., A. Shleifer, and J. Wurgler, 2005, "Comovement," *Journal of Financial Economics*, 75, 283–317.

Boyer, B. H., 2011, "Style related comovement: Fundamentals or labels?," *Journal of Finance*, 66, 307–332.

Brogaard, J., J. Engelberg, and C. A. Parsons, 2014, "Networks and productivity: Causal evidence from editor rotations," *Journal of Financial Economics*, 111, 251–270.

Byers, A., 2006, "Jeff Bezos: The Founder of Amazon.com (Internet Career Biographies)," *Rosen Publishing Group*.

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

Carney, D. R., and M. R. Banaji, 2012, "First is best," *PLoS ONE*, 7, 1–5.

Chen, H., G. Noronha, and V. Singal, 2004, "The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation," *Journal of Finance*, 59, 1901–1929.

Chevalier, J., and G. Ellison, 1997, "Risk taking by mutual funds as a response to incentives," *Journal of Political Economy*, 105, 1167–1200.

Chicago Tribune, 1992, "'A alphabetical advantage' helps business listed in yellow pages," *November 08, 1992*.

Chordia, T., S.-W. Huh, and A. Subrahmanyam, 2007, "The cross-section of expected trading activity," *Review of Financial Studies*, 20, 709–740.

Cohen, L., and A. Frazzini, 2008, "Economic links and predictable returns," *Journal of Finance*, 63, 1977–2011.

Cooper, M. J., O. Dimitrov, and R. Rau, 2001, "A rose.com by any name," *Journal of Finance*, 56, 2371–2388.

Cooper, M. J., H. Gulen, and P. R. Rau, 2005a, "Changing names with style: Mutual fund name changes and their effect on fund flows," *Journal of Finance*, 60, 2825–2858.

Cooper, M. J., A. Khorana, I. Osobova, A. Patel, and P. R. Rau, 2005b, "Managerial actions in response to a market downturn: valuation effects of name changes in the dot.com decline," *Journal of Corporate Finance*, 11, 319–335.

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

Daniel, K., D. Hirshleifer, and A. Subrahmanyam, 1998, "Investor psychology and security market under- and overreactions," *Journal of Finance*, 53, 1839–1885.

DellaVigna, S., and J. M. Pollet, 2009, "Investor inattention and Friday earnings announcements," *Journal of Finance*, 64, 709–749.

DeMaria, R., and J. L. Wilson, 2002, "High score! The illustrated history of electronic games," *Osborne/McGraw-Hill.*

Diether, K. B., C. J. Malloy, and A. Scherbina, 2002, "Differences of opinion and the cross section of stock returns," *Journal of Finance*, 57, 2113–2141.

Dimson, E., 1979, "Risk measurement when shares are subject to infrequent trading," *Journal of Financial Economics*, 5, 263–278.

Dorn, D., G. Huberman, and P. Sengmueller, 2008, "Correlated trading and returns," *Journal of Finance*, 63, 885–920.

Economist, 2001, "As easy as ZYX," *Vol. 360, No. 8237, p. 13.*

Einav, L., and L. Yariv, 2006, "What's in a surname? The effects of surname initials on academic success," *Journal of Economic Perspectives*, 20, 175–187.

Engelberg, J., and C. A. Parsons, 2011, "The causal impact of media in financial markets," *Journal of Finance*, 66, 67–97.

Engelberg, J., C. Sasseville, and J. Williams, 2012, "Market madness? The case of mad money," *Management Science*, 58, 351–364.

Falkenstein, E. G., 1996, "Preferences for stock characteristics as revealed by mutual fund holdings," *Journal of Finance*, 51, 111–135.

Fama, E. F., and K. R. French, 1992, "The cross-section of expected stock returns," *Journal of Finance*, 47, 427–465.

——— , 1997, "Industry costs of equity," *Journal of Financial Economics*, 43, 153–193.

Fama, E. F., and J. D. MacBeth, 1973, "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, 81, 607–636.

Fang, L., and J. Peress, 2009, "Media coverage and the cross-section of stock returns," *Journal of Finance*, 64, 2023–2052.

Gil-Bazo, J., and P. Ruiz-Verdu, 2009, "The relation between price and performance in the mutual fund industry," *Journal of Finance*, 64, 2153–2183.

Green, T. C., and B.-H. Hwang, 2009, "Price-based return comovement," *Journal of Financial Economics*, 93, 37–50.

Green, T. C., and R. Jame, 2013, "Company name fluency, investor recognition, and firm value," *Journal of Financial Economics*, 109, 813–834.

Greenwood, R., 2008, "Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights," *Review of Financial Studies*, 21, 1153–1186.

Grinblatt, M., and M. Keloharju, 2001, "How distance, language, and culture influence stockholdings and trades," *Journal of Finance*, 56, 1053–1073.

Grullon, G., G. Kanatas, and J. P. Weston, 2004, "Advertising, breadth of ownership, and liquidity," *Review of Financial Studies*, 17, 439–461.

Gurun, U. G., and A. W. Butler, 2012, "Don't believe the hype: Local media slant, local advertising, and firm value," *Journal of Finance*, 67, 561–597.

Harris, M., and A. Raviv, 1993, "Differences of opinion make a horse race," *Review of Financial Studies*, 6, 473–506.

Haslem, J. A., 2012, "Mutual funds and individual investors: Advertising and behavioral issues," *Unpublished working paper, University of Maryland.*

Heston, S. L., and R. Sadka, 2008, "Seasonality in the cross-section of stock returns," *Journal of Financial Economics*, 87, 418–445.

Hillert, A., H. Jacobs, and S. Müller, 2014, "Media makes momentum," *Review of Financial Studies, forthcoming.*

Himmelberg, C., G. Hubbard, and D. Palia, 2002, "Understanding the determinants of managerial ownership and the link between ownership and performance," *Journal of Financial Economics*, 53, 353–384.

Hirshleifer, D., S. S. Lim, and S. H. Teoh, 2009, "Driven to distraction: Extraneous events and underreaction to earnings news," *Journal of Finance*, 64, 2289–2325.

Ho, D. E., and K. Imai, 2008, "Estimating causal effects of ballot order from a randomized natural experiment," *Public Opinion Quarterly*, 72, 216–240.

Hong, H. G., J. D. Kubik, and J. C. Stein, 2008, "The only game in town: Stock-price consequences of local bias," *Journal of Financial Economics*, 90, 20–37.

Hong, H. G., T. Lim, and J. C. Stein, 2000, "Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies," *Journal of Finance*, 55, 265–295.

Hortacsu, A., and C. Syverson, 2004, "Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds," *Quarterly Journal of Economics*, 119, 403–456.

Hou, K., 2007, "Industry information diffusion and the lead-lag effect in stock returns," *Review of Financial Studies*, 20, 1113–1138.

Huberman, G., 2001, "Familiarity breeds investment," *Review of Financial Studies*, 14, 659–680.

Huddart, S., M. Lang, and M. H. Yetman, 2009, "Volume and price patterns around a stock's 52-week highs and lows: Theory and evidence," *Management Science*, 55, 16–31.

Jacobs, H., and M. Weber, 2012, "The trading volume impact of local bias: Evidence from a natual experiment," *Review of Finance*, 16, 867–901.

Jain, P. C., and J. S. Wu, 2000, "Truth in mutual fund advertising: Evidence on future performance and fund flows," *Journal of Finance*, 55, 937–958.

Jurajda, S., and D. Münich, 2010, "Admission to selective schools, alphabetically," *Economics of Education Review*, 29, 1100–1109.

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

Kahneman, D., 1973, "Attention and effort," *Prentice Hall, New Jersey.*

Kawasaki, G., 2006, "The name game," $blog.guykawasaki.com/2006/02/the_name_game.html$.

Koehler, J. J., and M. Mercer, 2009, "Selection neglect in mutual fund advertisements," *Management Science*, 55, 1007–1121.

Kumar, A., 2009, "Who gambles in the stock market?," *Journal of Finance*, 64, 1889–1933.

Kumar, A., and C. C. Lee, 2006, "Retail investor sentiment and return comovements," *Journal of Finance*, 61, 2451–2486.

Kumar, A., A. Niessen-Ruenzi, and O. Spalt, 2014, "What is in a name? Mutual fund flows when managers have foreign sounding names," *Review of Financial Studies, conditionally accepted*.

Lee, C. M., A. Shleifer, and R. H. Thaler, 1991, "Investor sentiment and the closed-end fund puzzle," *Journal of Finance*, 46, 75–109.

Lee, C. M., and B. Swaminathan, 2000, "Price momentum and trading volume," *Journal of Finance*, 55, 2017–2069.

Lemmon, M., and E. Portniaguina, 2006, "Consumer confidence and asset prices: some empirical evidence," *Review of Financial Studies*, 19, 1499–1529.

Llorente, G., R. Michaely, G. Saar, and J. Wang, 2002, "Dynamic volume-return relation of individual stocks," *Review of Financial Studies*, 15, 1005–1047.

Lo, A. W., and J. Wang, 2000, "Trading volume: Definitions, data analysis, and implications of portfolio theory," *Review of Financial Studies*, 13, 257–300.

Lohse, G. L., 1997, "Consumer eye movement pattern on yellow pages advertising," *Journal of Advertising*, 26, 61–73.

Loughran, T., and P. Schultz, 2004, "Weather, stock returns, and the impact of localized trading behavior," *Journal of Financial and Quantitative Analysis*, 39, 343–364.

———, 2005, "Liquidity: Urban versus rural firms," *Journal of Financial Economics*, 78, 341–374.

Meer, J., and H. S. Rosen, 2011, "The ABCs of charitable solicitation," *Journal of Public Economics*, 95, 363371.

Menzly, L., and O. Ozbas, 2010, "Market segmentation and cross-predictability of returns," *Journal of Finance*, 65, 1555–1580.

Merritt, D. J., 1999, "Calling professor AAA: how to visit at the school of your choice," *Journal of Legal Education*, 49, 557–563.

Merton, R. C., 1987, "A simple model of capital market equilibrium with incomplete information," *Journal of Finance*, 42, 483–510.

Michel, J.-B., Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, The Google Books Team, J. P. Picket, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden, 2011, "Quantitative analysis of culture using millions of digitized books," *Science*, 331, 176–182.

Milwaukee Journal, 1978, "Don't look back; alphabetism may be gaining on you," *December 10, 1997.*

Mullainathan, S., 2002, "Thinking through categories," *Unpublished working paper, MIT.*

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

———, 1994, "Automatic lag selection in covariance matrix estimation," *Review of Economic Studies*, 61, 631653.

Niessen-Ruenzi, A., and S. Ruenzi, 2013, "Sex matters: Gender and prejudice in the mutual fund industry," *Unpublished working paper, University of Mannheim.*

Odean, T., 1998, "Are investors reluctant to realize their losses?," *Journal of Finance*, 53, 1775–1798.

——— , 1999, "Do investors trade too much?," *American Economic Review*, 89, 1279–1298.

Peng, L., and W. Xiong, 2006, "Investor attention, overconfidence and category learning," *Journal of Financial Economics*, 80, 563–602.

Peress, J., 2008, "Media coverage and investors' attention to earnings announcements," *Unpublished working paper, Insead.*

Petersen, M. A., 2009, "Estimating standard errors in finance panel data sets: Comparing approaches," *Review of Financial Studies*, 22, 435–480.

Rashes, M. S., 2001, "Massively confused investors making conspicuously ignorant choices (MCI-MCIC)," *Journal of Finance*, 56, 1911–1927.

Richardson, M. L., 2010, "Alphabetic bias in the selection of reviewers for the American Journal of Roentgenology," *American Journal of Roentgenology*, 191, 213–216.

Seasholes, M. S., and G. Wu, 2007, "Predictable behavior, profits, and attention," *Journal of Empirical Finance*, 14, 590–610.

Sirri, E. R., and P. Tuffano, 1998, "Costly search and mutual fund flows," *Journal of Finance*, 53, 1589–1622.

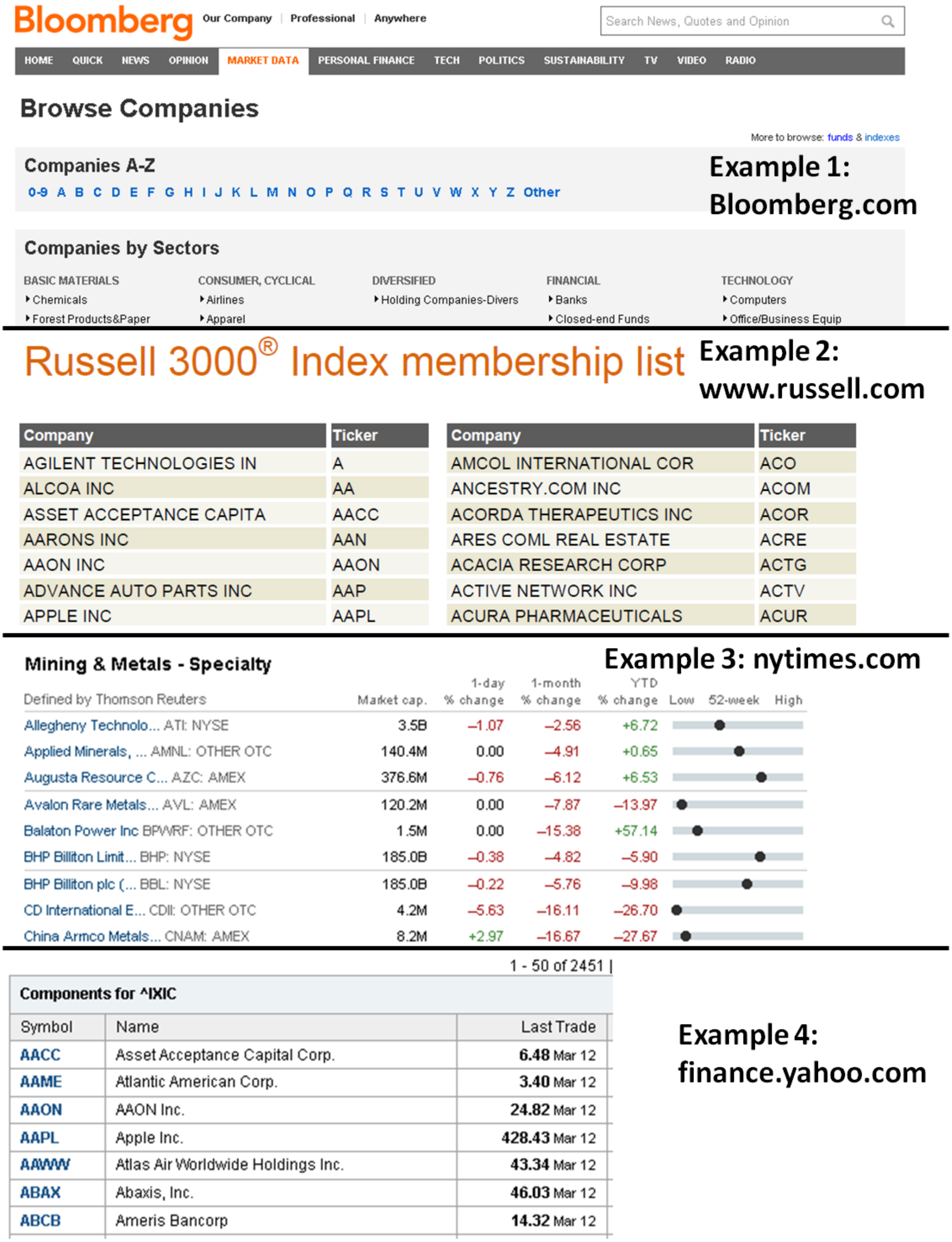
Solomon, D. H., E. F. Soltes, and D. Sosyura, 2014, "Winners in the spotlight: Media coverage of fund holdings as a driver of flows," *Journal of Financial Economics*, 113, 53–72.

Statman, M., S. Thorley, and K. Vorkink, 2006, "Investor overconfidence and trading volume," *Review of Financial Studies*, 19, 1531–1565.

Tetlock, P. C., 2011, "All the news that's fit to reprint: Do investors react to stale information?," *Review of Financial Studies*, 24, 1481–1512.

van Praag, C. M., and B. M. van Praag, 2008, “The benefits of being economics professor A (rather than Z),” *Economica*, 75, 782–796.

Figure 1: Motivating examples: Alphabetical ordering of stock names on popular finance websites



Bloomberg Our Company | Professional | Anywhere

Search News, Quotes and Opinion

HOME QUICK NEWS OPINION MARKET DATA PERSONAL FINANCE TECH POLITICS SUSTAINABILITY TV VIDEO RADIO

Browse Companies

More to browse: funds & indexes

Companies A-Z

0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Other

Example 1: Bloomberg.com

Companies by Sectors

BASIC MATERIALS	CONSUMER, CYCLICAL	DIVERSIFIED	FINANCIAL	TECHNOLOGY
Chemicals	Airlines	Holding Companies-Divers	Banks	Computers
Forest Products&Paper	Apparel		Closed-end Funds	Office/Business Equip

Russell 3000® Index membership list

Example 2: www.russell.com

Company	Ticker	Company	Ticker
AGILENT TECHNOLOGIES IN	A	AMCOL INTERNATIONAL COR	ACO
ALCOA INC	AA	ANCESTRY.COM INC	ACOM
ASSET ACCEPTANCE CAPITA	AACC	ACORDA THERAPEUTICS INC	ACOR
AARONS INC	AAN	ARES COML REAL ESTATE	ACRE
AAON INC	AAON	ACACIA RESEARCH CORP	ACTG
ADVANCE AUTO PARTS INC	AAP	ACTIVE NETWORK INC	ACTV
APPLE INC	AAPL	ACURA PHARMACEUTICALS	ACUR

Mining & Metals - Specialty

Example 3: nytimes.com

Defined by Thomson Reuters	Market cap.	1-day % change	1-month % change	YTD % change	Low 52-week High
Allegheny Technolo... ATI: NYSE	3.5B	–1.07	–2.56	+6.72	
Applied Minerals, ... AMNL: OTHER OTC	140.4M	0.00	–4.91	+0.65	
Augusta Resource C... AZC: AMEX	376.6M	–0.76	–6.12	+6.53	
Avalon Rare Metals... AVL: AMEX	120.2M	0.00	–7.87	–13.97	
Balaton Power Inc BPWRF: OTHER OTC	1.5M	0.00	–15.38	+57.14	
BHP Billiton Limit... BHP: NYSE	185.0B	–0.38	–4.82	–5.90	
BHP Billiton plc (... BBL: NYSE	185.0B	–0.22	–5.76	–9.98	
CD International E... CDII: OTHER OTC	4.2M	–5.63	–16.11	–26.70	
China Armco Metals... CNAM: AMEX	8.2M	+2.97	–16.67	–27.67	

1 - 50 of 2451 |

Components for ^IXIC

Example 4: finance.yahoo.com

Symbol	Name	Last Trade
AACC	Asset Acceptance Capital Corp.	6.48 Mar 12
AAME	Atlantic American Corp.	3.40 Mar 12
AAON	AAON Inc.	24.82 Mar 12
AAPL	Apple Inc.	428.43 Mar 12
AAWW	Atlas Air Worldwide Holdings Inc.	43.34 Mar 12
ABAX	Abaxis, Inc.	46.03 Mar 12
ABCB	Ameris Bancorp	14.32 Mar 12

Figure 2: Percentage difference in trading characteristics for alphabetically sorted firms

This figure compares trading activity (as measured by monthly stock-level turnover) and costs of trading (as measured by the monthly Amihud (2002) illiquidity ratio) for different groups of firms trading on the NYSE or AMEX. Firms are sorted alphabetically (in ascending order) by their firm name into five groups: First 5%, 5%-25%, 25%-50%, 50%-75%, and the last 25%, which serve as a benchmark. Findings display the percentage difference in trading activity and costs of trading for the first four groups relative to the fifth group. Values represent coefficients obtained from corresponding dummy variables in multivariate predictive panel regressions. Standard errors are double-clustered by firm and month. Additionally, year dummies are included in the regression. The sample period ranges from February 1995 to December 2009. Control variables are lagged (at least) one month and correspond to model 3 explained in section 2.2. The model controls for signed stock returns, book-to-market ratio, market capitalization, firm age, nominal share price, leverage, beta, number of analysts, analyst forecast dispersion, earnings surprises, earnings volatility, returns in t-12 and over t-7 to t-1, 52-week high, squared market capitalization, index membership (S&P 500, Dow Jones 30), idiosyncratic volatility, alpha, advertising expenditures, research and development expenditures, urban and rural regions, number of employees, industry classification, national and local press coverage (in total 45 newspapers), ten variables for the number of different types of 8-K filings, firm headquarter state, and company name fluency. Statistical significance at the ten, five and one-percent level is indicated by *, **, and ***, respectively.

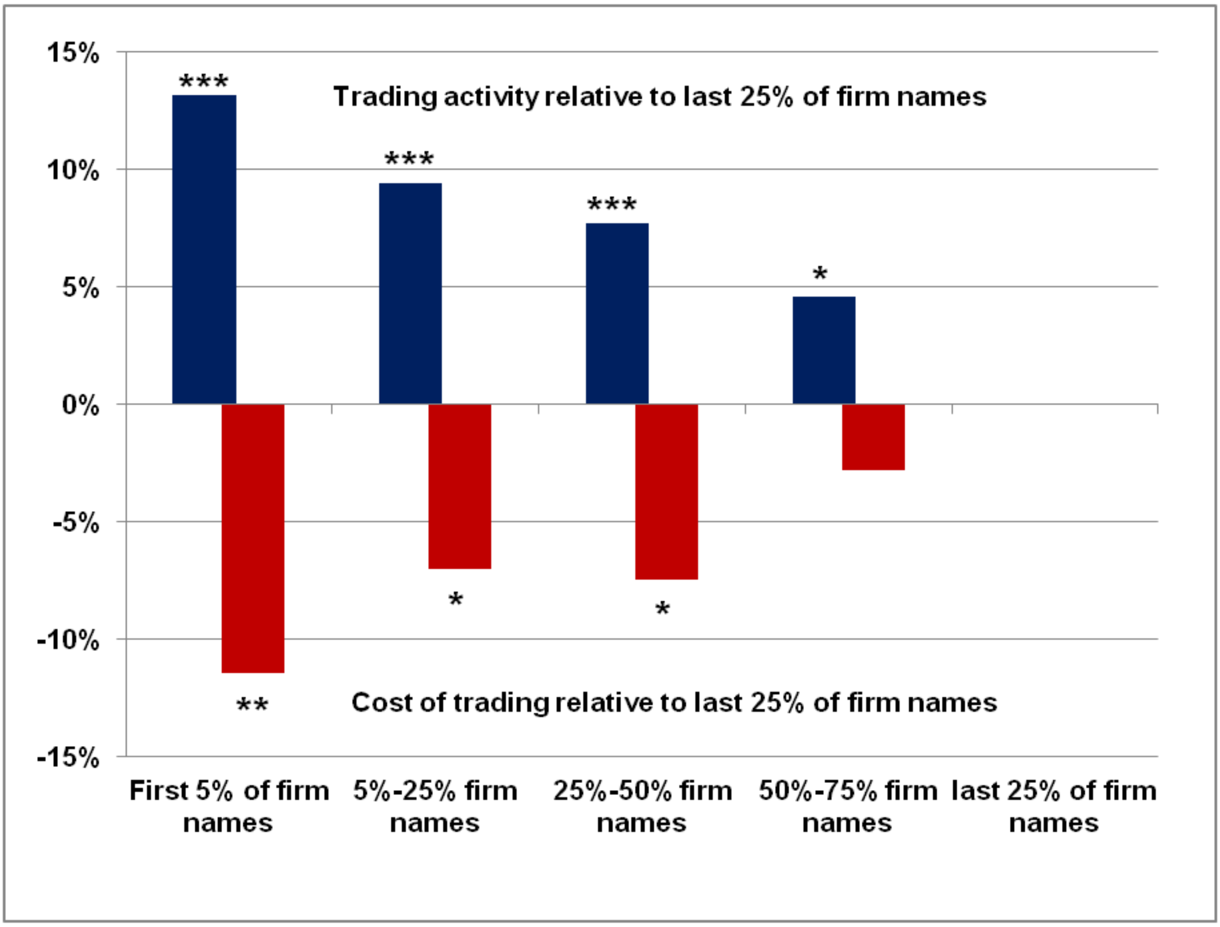


Figure 3: First letter of firm names vs fist letter of words in general

The upper half of the figure shows the relative frequency of firm names starting with a given letter in comparison to the respective frequencies for words in general. The distribution of firm names is based on NYSE/AMEX/Nasdaq stocks trading between 1983 and 2011 which meet the data requirements for model 1 described in detail in section 2.2. The distribution of words in general is derived from the application "Google Ngram" (see Michel et al. (2011) for more details). The corpus is based on the words contained in several hundreds of thousands of digitalized books written in American English and published between 1983 and 2009. The lower half of the figure shows the cumulative distribution functions based on the numbers displayed in the upper half of the figure.

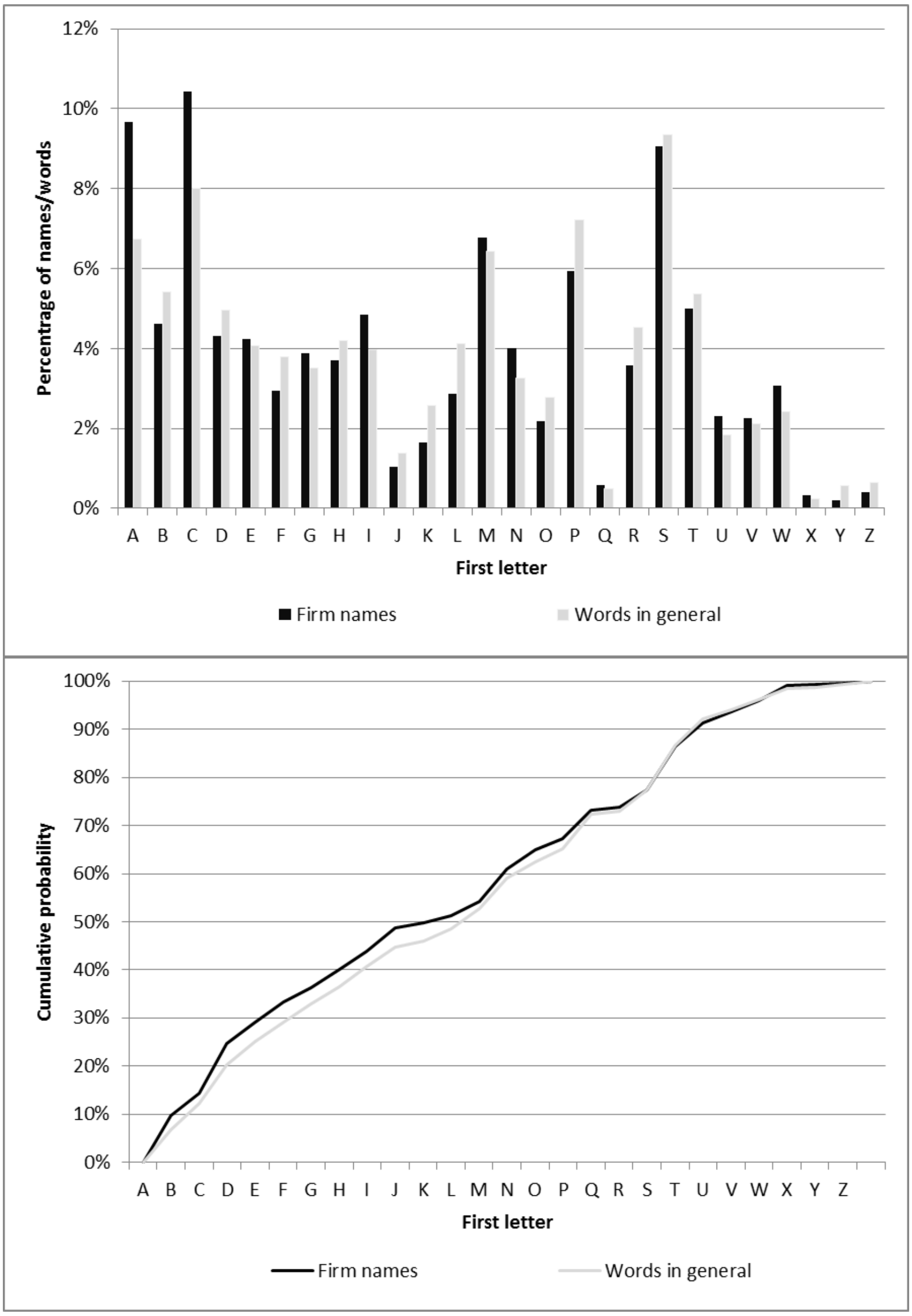


Table 1: Appendix: Description of control variables used in the stock market analysis

Variable	Description	Data source
Turnover	Turnover is defined as the number of monthly shares traded divided by the number of shares outstanding. Turnover is logarithmized as it is naturally skewed. To avoid the problem of zero turnover, we follow previous literature (e.g. Llorente et al. (2002)) in adding a small constant to the turnover before taking logs. In the tests across exchanges, we follow the method suggested in e.g. Loughran and Schultz (2005) by doubling trading volume for NYSE and AMEX stocks, pooling them with Nasdaq stocks, and adding a Nasdaq dummy to the regression.	CRSP
Illiquidity ratio	Inspired by Amihud (2002), the ratio is computed as the monthly average of daily illiquidity measures, defined as the absolute daily stock return divided by daily trading volume in millions of dollars. In the tests across exchanges, we follow the method suggested in e.g. Loughran and Schultz (2005) by doubling trading volume for NYSE and AMEX stocks, pooling them with Nasdaq stocks, and adding a Nasdaq dummy to the regression.	CRSP (daily)
Ret+ and Ret-	To proxy for rebalancing needs (e.g. Chordia et al. (2007)) or investor recognition (e.g. Merton (1987), Barber and Odean (2008)), we include two variables defined as the stock return in the previous month if positive (negative) and zero otherwise. This distinction is motivated by possible asymmetric effects caused by short-selling constraints or the disposition effect (e.g. Odean (1998), Grinblatt and Keloharju (2001)).	CRSP
Beta	We use Dimson (1979)-betas to account for nonsynchronous trading. To this end, a firm's excess return is first regressed on the current market excess return and the previous month's market excess return. We employ rolling three year window regressions (month t-37 to t-1), for which we require at least 24 valid monthly observations. Beta is then calculated as the sum of the coefficients for the current and the lagged market excess return.	CRSP
Price, age, and firm size	We include lagged logarithmized versions of the firm's market capitalization, age, and nominal share price to control for potential differences in ownership structure, visibility, or trading costs (Falkenstein (1996), Kumar and Lee (2006), Merton (1987)). While these three variables tend to be substantially correlated, our findings are very robust to specifications in which we only include one or two of them. Firm size is ln(market capitalization in million USD), measured in month t-1. Age ist computed as ln(age in months measured in month t-1), where age is determined as min (month in which the firms PERMCO first appears in CRSP; IPO date from Compustat). Nominal share price is defined as ln(abs(prc)) measured in month t-1.	CRSP and Compustat

[continued overleaf]

Book-to-market ratio	We employ the book-to-market ratio to account for differences in fundamentals and in investor attention between value and growth stocks. The book value of equity is defined as the sum of common/ordinary equity and deferred taxes. Following the standard in the literature (e.g. Fama and French (1992)), the market to book ratio is updated once every year at the end of June. Market capitalization (from CRSP) is measured as of the previous December, while book equity is based on the last fiscal year ending in the previous calendar year.	CRSP and Compustat
Leverage	Following Chordia et al. (2007) and the references therein, we also use a firm's leverage. It is defined as book debt scaled by total assets, where book debt is the sum of current liabilities, long-term debt, and preferred stock. For the computation of preferred stock, we use redemption value or, if not available, the liquidating value. For firms with missing preferred stock, the value is set to zero. We use lagged leverage with the same timing as specified for book equity, i.e. we update the value once every year at the end of June.	Compustat
Earnings surprise	Earnings surprises might be considered to be a proxy estimation uncertainty about fundamental values. Earnings surprises are computed as the difference between the actual earnings in quarter t (primary earnings per share before extraordinary items, epspxq/ajexq) and expected earnings (earnings four quarters ago). The difference is scaled by the price per share at the end of quarter (obtained from Compustat). In the regressions, we use the most recent lagged value of earnings surprises. To exclude stale observations, we impose the additional constraint that the maximum number of calendar days between the earnings announcement and the date used in the regression is 125 days.	I/B/E/S and Compustat
Earnings volatility	Earnings volatility might be considered to be a proxy estimation uncertainty about fundamental values. As in Chordia et al. (2007), it is defined as the standard deviation of the most recent eight quarterly earnings. We require at least two valid observations to compute this value. In the regressions, we use the most recent lagged value of earnings surprises. To exclude stale observations, we impose the additional constraint that the maximum number of calendar days between the earnings announcement and the date used in the regression is 250 days.	Compustat
Analyst coverage	We use the lagged logarithmized number of analysts following a stock as a proxy for information diffusion, the mass of informed agents and the firm's visibility (e.g. Hong et al. (2000)). Coverage is computed as ln(1+number of analysts providing fiscal year estimates in month t-1). We use old values to fill up single months with missing analyst coverage.	I/B/E/S

[continued overleaf]

Analyst forecast dispersion	We use analyst forecast dispersion as a proxy for differences of opinion (e.g. Diether et al. (2002)). As in e.g. Bandarchuk and Hilscher (2013), forecast dispersion is defined as the standard deviation of earnings per share forecasts scaled by the absolute value of the mean forecast. In case we have less than two analysts covering a given stock in a given month, we set forecast dispersion to zero. We use old values to fill up single months with missing forecast dispersion. We use analyst dispersion lagged by one month (t-1) in the regressions.	I/B/E/S
Squared market capitalization	The squared value of ln (market capitalization in million USD), measured in month t-1, is included to better control for nonlinearities in size	CRSP
S&P 500 membership, Dow Jones 30 membership, Nasdaq 100 membership	To control for index effects (e.g. Chen et al. (2004)), we construct a dummy variable which takes on a value of 1 (0) if a firm is (not) a member of the S&P 500 (Dow Jones 30, Nasdaq 100) in month t-1	Compustat
Advertising expenditures, research and development expenditures	Grullon et al. (2004) show that advertising expenditures are positively related to investor recognition. Research and development expenditures (R&D) might have a similar effect. We use both variables (xad, xrd) and scale them by total sales. We scale advertising expenditures (xad) by sales (sale). We use lagged values with the same timing as specified for book equity, i.e. we update the value once every year at the end of June. We follow Himmelberg et al. (2002) and Green and Jame (2013) in setting missing values to 0 and including a dummy that equals one when advertising expenditures are missing, and zero otherwise. We employ the same timing and treatment of missing values for research and development expenditures	Compustat
Employees	Following Loughran and Schultz (2005), we consider the (logarithmized) number of employees as reported by Compustat (emp). We employ the same timing and treatment of missing values as described in detail for advertising expenditures.	Compustat
Momentum formation	We employ the cumulative return over t-7 to t-1 to account for effects related to momentum as discussed in e.g. Statman et al. (2006) or Lee and Swaminathan (2000).	CRSP
Seasonal Momentum	We employ the return twelve months ago to account for effects related to seasonal momentum as discussed in e.g. Heston and Sadka (2008).	CRSP
52 week high dummy	Inspired by e.g. Seasholes and Wu (2007) or Huddart et al. (2009), we consider a 52-week high dummy which takes a value of 1 (0) if the price at the end of month t-1 was less than 2% below the highest price observed at the end of at any day in months t-13 to t-1	CRSP (daily)
Idiosyncratic volatility	Idiosyncratic volatility is defined as the standard deviation of the residual obtained from rolling monthly regressions of a firm's excess return on a Carhart (1997) four factor model. We run the rolling regressions from months t-37 to t-1 and require at least 24 valid observations.	CRSP and Kenneth French's Data Library

[continued overleaf]

Alpha	Alpha is defined as the intercept obtained in the regression framework for idiosyncratic volatility specified above. Alpha has been argued to contain a premium related to liquidity or heterogeneous information (e.g. Lo and Wang (2000)).	CRSP and Kenneth French's Data Library
Urban firm	Loughran and Schultz (2005) show that firms located in urban areas are traded heavier due to the larger local investor base (see also Hong et al. (2008)). Following the literature on local bias, we use a company's headquarter as a proxy for its location. A stock is defined as an urban stock if the headquarter is in one of the ten largest US metropolitan areas according to the 2000 census (New York City, Los Angeles, Chicago, Washington-Baltimore, San Francisco, Philadelphia, Boston, Detroit, Dallas, Houston)	Compustat and US Census
Rural firm	Loughran and Schultz (2005) show that firms located in urban areas are traded heavier due to the larger local investor base (see also Hong et al. (2008)). A company is defined as rural if its headquarter is at least 100 miles from the center of any of the 49 U.S. metropolitan areas of one million or more people according to the 2000 census	Compustat and US Census
Industry classification	We use the 48 industries defined in Fama and French (1997), and group stocks which are not assigned to any industry in category 49.	Kenneth French's homepage
Firm name fluency	We rely on the company name fluency measure developed in Green and Jame (2013). Values are available at a yearly frequency and are lagged by one year.	Russel Jame's homepage
National media coverage	Several studies suggest that media coverage is a powerful proxy for firm-specific news and investor recognition (e.g. Barber and Odean (2008), Engelberg and Parsons (2011), Peress (2008), Tetlock (2011)). We quantify national media coverage as Ln (1+number of press articles about a given firm in the New York Times, USA Today, Wall Street Journal, and Washington Post) in month t-1. In total, we gather roughly 261,000 firm-specific articles.	LexisNexis
Local media coverage	Recent research (e.g. Gurun and Butler (2012)) has uncovered that a special role for local media. We thus gather in total about 888,000 firm-specific news articles from 41 local newspapers (see Hillert et al. (2014) for an overview). We quantify local media coverage as Ln (1+number of press articles about a given firm in any local newspaper) in month t-1.	LexisNexis
Firm events	In order to better control for firm-specific news, we gather in total about 421,000 8-K filings which are individually accessible via the EDGAR database. The SEC (http://www.sec.gov/answers/form8k.htm) defines an 8-K filing as a "'current report' companies must file with the SEC to announce major events that shareholders should know about". For instance, potential events include changes in executive management or ownership, entry into a material definitive agreement, matters related to accountants and financial statements, or asset redeployments. Along pre-defined dimensions, firms have to specify the type of reportable event(s) in the main body of an 8-K filing. We thus write a parsing algorithm to identify the exact type of news. On this basis, we construct in total ten variables with the logarithmized number of different firm events in the previous month.	EDGAR

Table 2: Stock-level turnover, illiquidity and alphabetic bias: Predictive panel regressions

This table summarizes the main results of eighteen predictive panel regressions which differ in the dependent variable (natural logarithm of firm turnover in month t, logarithmized Amihud (2002) illiquidity ratio in month t), the firm universe (NYSE/AMEX, Nasdaq, all), and the control variables/sample periods (models 1-3). In all regressions, the explanatory variable of interest is *position continuous*, defined as the relative position of a firm's name within the alphabetically sorted firm universe in the previous month. Thus, *position continuous* takes on values between zero and one. All control variables are lagged at least by one month. If we run regressions across all exchanges, trading volume for NYSE/AMEX firms is doubled, and a Nasdaq dummy is included. In panel A, model 1 controls for signed stock returns, book-to-market ratio, market capitalization, firm age, nominal share price, leverage ratio, and beta. In panel B, model 2 additionally controls for the number of analysts, analyst forecast dispersion, earnings surprises, earnings volatility, returns in t-12 and over t-7 to t-1, 52-week high, squared (logarithmized) market capitalization, index membership (S&P 500, Dow Jones 30, Nasdaq 100 (all where applicable)), idiosyncratic volatility, alpha, advertising expenditures, research and development expenditures, dummies for urban and rural regions, number of employees, and industry classification (49 Fama and French (1997) industries). Finally, model 3 additionally controls for national and local press coverage (in total 45 newspapers), all 8-K filings (ten different dummies which proxy for different firm events), firm headquarter state, and company name fluency. In all regressions, standard errors are double-clustered by firm and month. Additionally, year dummies are included in the regression. Statistical significance at the ten-, five- and one-percent level is indicated by *, **, and ***, respectively.

Panel A: Model 1						
Firm universe	NYSE/AMEX		Nasdaq		All	
Sample period	7/1963-12/2011		1/1983-12/2011		1/1983-12/2011	
Stock turnover: Coefficient (t-stat) of *position continuous*	-0.100***	(-3.05)	-0.069**	(-2.06)	-0.090***	(-3.34)
Stock turnover: N / R^2	846,810	0.427	653,855	0.376	1,152,541	0.392
Amihud illiquiditiy: Coefficient (t-stat) of *position continuous*	0.078**	(2.10)	0.102**	(2.45)	0.093***	(2.82)
Amihud illiquidity: N / R^2	845,929	0.912	640,063	0.817	1,138,194	0.878
Panel B: Model 2						
Firm universe	NYSE/AMEX		Nasdaq		All	
Sample period	2/1980-12/2011		1/1983-12/2011		1/1983-12/2011	
Stock turnover: Coefficient (t-stat) of *position continuous*	-0.088***	(-2.87)	-0.089***	(-3.08)	-0.098***	(-4.35)
Stock turnover: N / R^2	516,629	0.485	517,225	0.439	981,361	0.445
Amihud illiquidity: Coefficient (t-stat) of *position continuous*	0.070*	(1.77)	0.107***	(2.76)	0.089***	(2.98)
Amihud illiquidity: N / R^2	516,559	0.918	515,810	0.847	979,879	0.894
Panel C: Model 3						
Firm universe	NYSE/AMEX		Nasdaq		All	
Sample period	2/1995-12/2009		2/1995-12/2009		2/1995-12/2009	
Stock turnover: Coefficient (t-stat) of *position continuous*	-0.123***	(-3.46)	-0.081**	(-2.49)	-0.108***	(-4.21)
Stock turnover: N / R^2	223,577	0.530	302,991	0.457	526,568	0.464
Amihud illiquidity: Coefficient (t-stat) of *position continuous*	0.111**	(2.24)	0.105**	(2.49)	0.107***	(3.15)
Amihud illiquidity: N / R^2	223,570	0.922	302,942	0.870	526,512	0.906

Table 3: Determinants of alphabetic rank (column 1) and firm-level turnover (columns 2-3)

This table shows coefficients of predictive regressions of *position continuous* (in column 1) or the natural logarithm of firm turnover in month t (columns 2-3) on a number of explanatory variables. The firm universe is NYSE/AMEX/Nasdaq. *Position continuous* is defined as the relative position of a firm's name within the alphabetically sorted firm universe. Turnover for NYSE/AMEX firms is doubled, and a Nasdaq dummy is included. All control variables are lagged at least by one month. Control variables correspond to the ones used in model 3 (see section 2.2 for an overview and the table 1 for a detailed description). In columns 1 and 2, we run predictive panel regressions and double-cluster standard errors by firm and month. Additionally, year dummies are included in the regression. In column 3, we run predictive Fama/MacBeth regressions which rely on heteroskedasticity- and autocorrelation-consistent (HAC) Newey and West (1987) standard errors with automatic lag length selection (see Newey and West (1994)). In all columns, statistical significance at the ten-, five- and one-percent level is indicated by *, **, and ***, respectively.

Specification	1		2		3	
Dependent variable	Position Continuous		Turnover		Turnover	
Econometric approach	Predictive panel regression		Predictive panel regression		Fama/MacBeth regression	
Position Continuous			**-0.1080*** **	**(-4.21)**	**-0.0907*** **	**(-7.51)**
Ret+	-0.003	(-0.70)	0.896***	(14.25)	1.071***	(8.97)
Ret-	-0.007	(-0.86)	-1.516***	(-22.76)	-1.654***	(-15.09)
Beta	-0.002	(-0.97)	0.062***	(9.53)	0.079***	(7.81)
Share Price	-0.012*	(-1.93)	0.110***	(6.99)	0.063***	(2.97)
Age	0.002	(0.41)	-0.081***	(-6.65)	-0.034***	(-2.97)
Market Capitalization	0.023**	(2.35)	0.364***	(14.44)	0.540***	(9.22)
Book to Market	0.008*	(1.94)	0.007***	(2.76)	0.032***	(3.98)
Leverage	-0.012	(-1.40)	0.038**	(2.49)	0.038**	(2.49)
Advertising	-0.008	(0.50)	0.001	(0.57)	0.101**	(2.06)
Missing Advertising	-0.003	(-0.43)	-0.023	(-1.38)	-0.040***	(-4.72)
Research & Development	-0.001	(-0.07)	0.016	(1.25)	-0.009	(-0.65)
Missing R&D	0.020**	(2.07)	-0.039*	(-2.10)	-0.043***	(-3.24)
Earnings Volatility	-0.043	(-1.52)	-0.002	(-1.11)	0.050	(0.88)
Earnings Surprise	0.001*	(1.94)	0.005*	(1.76)	0.278***	(6.34)
Number Analysts	-0.015**	(-2.19)	0.478***	(27.72)	0.583***	(42.86)
Analyst Forecast Dispersion	0.000	(0.13)	0.003	(1.17)	0.001	(0.21)
S&P 500	-0.029	(-1.62)	-0.121***	(-3.75)	-0.0560	(-1.11)
Dow Jones 30	-0.049	(-0.95)	-0.353***	(-3.87)	-0.142	(-1.61)
Nasdaq 100	-0.028	(-1.30)	-0.028	(-0.73)	0.046*	(1.96)
Nasdaq Membership	0.012	(1.22)	-0.396***	(-16.56)	-0.440***	(-6.08)
Squared Market Capitalization	-0.001	(-0.75)	-0.019***	(-10.34)	-0.036***	(-10.37)
Momentum Formation	-0.000	(-0.22)	0.087***	(6.24)	0.105***	(3.56)
52 Week High	-0.001	(-0.48)	-0.002	(-0.13)	0.084**	(2.04)
Number Employees	0.003	(0.79)	-0.026***	(-3.50)	-0.036***	(-4.58)
Employees Missing	0.004	(0.17)	0.036	(0.56)	-0.039	(-0.62)
Urban Firm	0.012	(0.93)	0.014	(0.59)	-0.004	(-0.22)
Rural Firms	0.009	(0.42)	-0.054	(-1.33)	-0.047*	(-1.71)
Idiosyncratic Volatility	-0.018	(-0.41)	2.856***	(12.58)	2.835***	(8.61)
Alpha	-0.084	(-0.98)	0.115	(0.53)	0.613	(1.40)
Name Fluency	0.005	(1.02)	0.011*	(1.71)	0.009*	(1.76)
Seasonal Momentum	-0.014	(-1.03)	0.059**	(2.23)	0.070***	(5.18)
National Media Coverage	0.003	(0.43)	-0.085***	(-5.65)	0.005	(0.99)
Local Media Coverage	-0.013**	(-2.00)	0.045***	(4.39)	0.031***	(4.56)
Year Controls	yes		yes		yes	
Firm Event Controls (ln(1+ 8-K filings))	yes		yes		yes	
State Controls	yes		yes		yes	
Industry Controls	yes		yes		yes	
(Average) R^2	0.025		0.46		0.51	
(Average) N	526,506		526,506		2,942	

Table 4: Turnover, (il)liquidity and alphabetic bias: Further evidence

This table summarizes the main results from six predictive panel regressions which differ in the dependent variable (Panel A: logarithmized monthly stock-level turnover, Panel B: logarithmized Amihud (2002) monthly illiquidity ratio), and the model specification (model 1-3 as described in detail in section 2.2). Displayed are the coefficients and t-statistics for four dummy variables, which indicate the position of a given firm name within the NYSE/Amex/Nasdaq universe. *Position dummy* 5 takes on a value of 1 if the company name is among the first 5% in a given month and zero otherwise. *Position dummy* 25 is 1 if the firm name is between percentiles 5 and 25. Finally, *position dummy* 50 (*position dummy* 75) is 1 if the name is between percentiles 25 and 50 (50 and 75). Consequently, findings are benchmarked against the firms near the end of the alphabet (percentile 75 and above). Standard errors are double-clustered by firm and month, and year dummies are included in the regression. The last three rows of each panel provide p-values from F-tests of coefficient equality. Statistical significance at the ten-, five- and one-percent level is indicated by *, **, and ***, respectively.

Panel A: Logarithmized turnover as dependent variable			
Firm universe	All exchanges	All exchanges	All exchanges
Model specification	Model 1	Model 2	Model 3
Coefficient of position dummy 5	0.0977***	0.1069***	0.1074***
t-stat of position dummy 5	2.59	3.46	3.19
Coefficient of position dummy 25	0.0754***	0.0740***	0.0806***
t-stat of position dummy 25	3.31	3.87	3.61
Coefficient of position dummy 50	0.0459**	0.0516***	0.0626***
t-stat of position dummy 50	2.03	2.80	3.02
Coefficient of position dummy 75	0.0208	0.0121	0.0133
t-stat of position dummy 75	0.97	0.67	0.64
p-value position dummy 5 = position dummy 25/50/75	0.55/0.16/0.04**	0.28/0.07*/0.00***	0.43/0.17/0.00***
p-value position dummy 25 = position dummy 50/75	0.17/0.01**	0.21/0.00***	0.37/0.00***
p-value position dummy 50 = position dummy 75	0.25	0.02**	0.01**
Panel B: Logarithmized illiquidity ratio as dependent variable			
Firm universe	All exchanges	All exchanges	All exchanges
Model specification	Model 1	Model 2	Model 3
Coefficient of position dummy 5	-0.1249***	-0.1156***	-0.1169***
t-stat of position dummy 5	-3.02	-3.12	-2.85
Coefficient of position dummy 25	-0.0613**	-0.0531**	-0.0623**
t-stat of position dummy 25	-2.11	-2.05	-2.07
Coefficient of position dummy 50	-0.0443*	-0.0433*	-0.0642**
t-stat of position dummy 50	-1.68	-1.76	-2.32
Coefficient of position dummy 75	-0.0014	0.0132	0.0101
t-stat of position dummy 75	-0.05	0.54	0.36
p-value position dummy 5 = position dummy 25/50/75	0.13/0.05**/0.00***	0.10*/0.04**/0.00***	0.19/0.19/0.00***
p-value position dummy 25 = position dummy 50/75	0.55/0.04**	0.70/0.00***	0.95/0.02**
p-value position dummy 50 = position dummy 75	0.09*	0.01**	0.00***

Table 5: Turnover, (il)liquidity and alphabetic bias: Fama/MacBeth regressions and name changes

This table summarizes the main results from Fama and MacBeth (1973)-type regressions (in panel A and B), the alphabetic distribution of firm name changes (in panel C), and individual firm fixed effect panel regressions (in panels D and E) which explore the consequences of name changes. The dependent variable in the regressions is either the natural logarithm of monthly firm turnover (in panel A and D) or the logarithmized monthly Amihud (2002) illiquidity ratio (in panels B and E). Control variables and sample periods are described in table 2. In all specifications, the explanatory variable of interest is *position continuous*, defined as the relative position of a firm's company name within the alphabetically sorted firm universe in the previous period t-1. Thus, *position continuous* takes on values between zero and one. The Fama and MacBeth (1973)-type regressions rely on heteroskedasticity- and autocorrelation-consistent (HAC) Newey and West (1987) standard errors with automatic lag length selection as in Newey and West (1994). For the analysis in panels C to E, we eliminate blank spaces in firm names reported in CRSP (see also footnote 4). In panels D and E, we condition on the five year periods both before and after a name change identifiable in the NYSE/Amex/Nasdaq universe of model 2. We require the name change to alter *position continuous* by at least 25%. The month of the name change is excluded, and we require at least two years of valid observations both before and after the name change. We then run panel regressions as in our baseline analysis (see table 2) with all controls of model 2, but we additionally include firm fixed effects, a name change dummy, and average *position continuous* both before and after the name change. Month to month changes in *position continuous* are captured by *small change* which is one (minus one) in the case of a position earlier (later) in the alphabetical ranking and zero otherwise. The standard errors in the firm-fixed regressions are clustered by firm. Statistical significance at the ten-, five- and one-percent level is indicated by *, **, and ***, respectively.

Panel A: Fama/MacBeth regressions with turnover as dependent variable (model 3)			
Model 3	NYSE/AMEX	Nasdaq	All
Coefficient of *position continuous*	-0.120***	-0.069***	-0.091***
t-stat of *position continuous*	-5.83	-3.66	-7.51
Panel B: Fama/MacBeth regressions with Amihud illiquidity as dependent variable (model 3)			
Model 3	NYSE/AMEX	Nasdaq	All
Coefficient of *position continuous*	0.109***	0.090***	0.095***
Coefficient of *position continuous*	7.31	3.93	7.32
Panel C: Alphabetic distribution of name changes (model 2)			
Strength of name changes	Number of name changes	% earlier in the alphabetic ranking (p-value)	Average change in *position continuous* (t-stat)
Absolute change of at least 25%	774	51.2% (0.26)	-2.33% (-1.22)
Absolute change of at least 50%	333	53.2% (0.12)	-5.33% (-1.42)
Absolute change of at least 75%	92	55.4% (0.15)	-9.94% (-1.13)
Absolute change of at least 90%	14	71.4% (0.05*)	-39.29% (-1.68)
Panel D: Name changes (firm fixed effect regressions) with turnover as dependent variable (model 2)			
Coefficient of *position continuous*	-0.142**	Coefficient of *small change*	0.0057
t-stat of *position continuous*	-2.08	t-stat of *small change*	0.79
Panel E: Name changes (firm fixed effect regressions) with Amihud illiquidity as dependent variable (model 2)			
Coefficient of *position continuous*	0.150*	Coefficient of *small change*	-0.0060
t-stat of *position continuous*	1.70	t-stat of *small change*	-0.70

Table 6: Breadth of ownership, firm valuation, and alphabetic bias: Panel regressions

This table summarizes the main results from panel regressions of breadth of ownership (panel A) and firm valuation (panel B) on measures of the relative position of a firm's name in alphabetical ordering and a number of controls. Breadth of ownership (firm valuation) is computed as the logarithmized number of shareholders (logarithmized market-to-book ratio) in a given year, averaged across all months in that year. In the case of market-to-book ratios, we exclude negative values as well as firms from the banking and insurance sector, and moreover winsorize the data at the 99th percentile in each year. Each panel reports findings from nine regressions, which differ in the alphabetical ordering measure and the control variables (model 1-3 as described in detail in section 2.2). In some regressions, we additionally include the stock's Amihud illiquidity ratio or/and the logarithmized number of shareholders as a control variable. In all regressions, the firm universe consists of stocks trading at NYSE, Amex, or Nasdaq. Standard errors are double-clustered by firm and year. Additionally, year dummies are included in the regression. Statistical significance at the ten-, five- and one-percent level is indicated by *, **, and ***, respectively.

Panel A: Logarithmized number of shareholders as dependent variable			
Model specification	Model 1	Model 2	Model 3
N	95,246	82,045	45,663
Panel A1: Impact of position continuous			
Coefficient of position continuous	-0.0879**	-0.0810**	-0.0865*
t-stat of position continuous	-2.23	-2.02	-1.70
Controlling for Amihud illiquidity	-0.0829**	-0.0726*	-0.0737
t-stat of position continuous	-2.10	-1.82	-1.41
Panel A2: Impact of position dummies			
Coefficient of position dummy 5	0.0862	0.1020*	0.0987
t-stat of position dummy 5	1.56	1.72	1.22
Coefficient of position dummy 25	0.0879***	0.0713**	0.0783*
t-stat of position dummy 25	2.60	2.05	1.72
Coefficient of position dummy 50	0.0387	0.0292	0.0592
t-stat of position dummy 50	1.21	0.90	1.39
Coefficient of position dummy 75	0.0615*	0.0385	0.0330
t-stat of position dummy 75	1.87	1.18	0.075
Panel B: Logarithmized market-to-book-ratio as dependent variable			
Model specification	Model 1	Model 2	Model 3
N	93,052	80,216	44,550
Panel B1: Impact of position continuous			
Coefficient of position continuous	-0.0537**	-0.0121	-0.0301
t-stat of position continuous	-2.27	-0.85	-1.45
Controlling for illiquidity/ownership	-0.0510**	-0.0101	-0.0261
t-stat of position continuous	-2.11	-0.54	-1.25
Panel B2: Impact of position dummies			
Coefficient of position dummy 5	0.1002***	0.0241	0.0584**
t-stat of position dummy 5	3.04	0.89	2.04
Coefficient of position dummy 25	0.0135	-0.0020	0.0093
t-stat of position dummy 25	0.84	-0.27	0.51
Coefficient of position dummy 50	0.0370*	-0.0037	0.0041
t-stat of position dummy 50	1.85	-0.25	0.21
Coefficient of position dummy 75	0.0012	-0.0090	-0.0088
t-stat of position dummy 75	0.09	-0.67	-0.51

Table 7: Firm visibility and alphabetic bias

This table summarizes main results from pooled predictive panel regressions of stock-level turnover (panel A) or of the stock-level Amihud illiquidity ratio (panel B) on *position continuous*, a proxy for a firm's visibility, the interaction term (the coefficient of primary interest) and control variables (see table 2 and section 2.2 for details). The regressions are all based on model 3 and NYSE/AMEX/Nasdaq stocks, which represents our most comprehensive specification. The sample period ranges from February 1995 to December 2009. All firm visibility proxies are lagged by at least one month. *Analyst coverage* refers to ln(1+number of analysts providing fiscal year estimates in month t-1). *Sales* refers to ln(firm sales), winsorized at the 99% level. *Industry market share* refers to the ln(one month lagged relative market capitalization within the 49 Fama and French (1997) industries), wincorized at the 99% level. *Advertising expenditures* refers to ln(advertising expenditures), winsorized at the 99% level. In this regression, only firm months with non-missing advertising expenditures are taken into account. In both panels, standard errors are double-clustered by firm and month. Additionally, year dummies are included in the regression. Statistical significance at the ten-, five- and one-percent level is indicated by *, **, and ***, respectively.

Visibility proxy	Interaction *position continuous* x visibility proxy	t-stat	Implied size of *position continuous* at the 25th pctl of the visibility proxy distribution	at the 75th pctl of the visibility proxy distribution
Panel A: Logarithmized turnover as dependent variable				
Analyst coverage	0.082***	(2.99)	-0.172	-0.048
Sales	0.013	(1.16)	-0.125	-0.088
Industry market share	0.073***	(2.72)	-0.154	-0.070
Advertising expenditures	1.976**	(2.06)	-0.229	-0.162
Panel B: Logarithmized Amihud illiquidity ratio as dependent variable				
Analyst coverage	-0.131***	(-3.90)	0.207	0.010
Sales	-0.022*	(-1.68)	0.139	0.075
Industry market share	-0.089**	(-2.40)	0.163	0.061
Advertising expenditures	-1.171	(-1.01)	0.163	0.123

Table 8: Investor sophistication and alphabetic bias

This table summarizes main results from pooled predictive panel regressions of stock turnover (Panel A) or the Amihud illiquidity ratio (Panel B) on *position continuous* (the coefficient of interest) and control variables (see table 2 and section 2.2 for details). We run subsample regressions based on model 3 and NYSE/AMEX/Nasdaq stocks. The sample period ranges from February 1995 to December 2009. Small firms (large firms) are firms with a below (above) median market capitalization in the previous month. Lottery firms (non-lottery firms) are firms with above (below) median idiosyncratic volatility, above (below) median idiosyncratic skewness, and below (above) median nominal share price. Firms with high (low) media coverage are firms with above (below) median residual media coverage, defined as the residual from monthly cross-sectional regressions of the logarithmized number of firm-specific articles in the New York Times, the USA Today, the Wall Street Journal, and the Washington Post on logarithmized lagged market capitalization. Fluent (nonfluent) firm names are firms with a fluency score of at least three (not more than two). High (low) institutional ownership refers to firms with above (below) median institutional ownership lagged by one quarter. Standard errors are double-clustered by firm and month. Additionally, year dummies are included in the regression. Statistical significance at the ten-, five- and one-percent level is indicated by *, **, and ***, respectively.

Panel A: Logarithmized turnover as dependent variable			Panel B: Logarithmized illiquidity ratio as dependent variable		
	Coefficient	t-stat		Coefficient	t-stat
Small stocks (N=263,192)	-0.1446***	-4.03	Small stocks (N=263,136)	0.1835***	3.96
Large stocks (N=263,314)	-0.0684**	-2.48	Large stocks (N=263,314)	0.0374	0.91
Difference	-0.0762*	-1.78	Difference	0.1461**	2.47
Lottery stocks (N=125,116)	-0.1122***	-2.79	Lottery stocks (N=125,099)	0.1427***	2.78
Non-lottery stocks (N=122,015)	-0.0923***	-3.14	Non-lottery stocks (N=122,007)	0.0909*	1.94
Difference	-0.0199	-0.39	Difference	0.0518	0.76
High media coverage (N=263,286)	-0.1321***	-3.95	High media coverage (N=263,233)	0.1502***	3.63
Low media coverage (N=263,330)	-0.0775***	-2.68	Low media coverage (N=263,217)	0.0578	1.42
Difference	-0.0546	-1.44	Difference	0.0924*	1.90
Fluent firm name (N=215,304)	-0.1757***	-4.54	Fluent firm name (N=215,276)	0.2054***	3.92
Nonfluent firm name (N=311,202)	-0.0712**	-2.11	Nonfluent firm name (N=311,174)	0.0584	1.32
Difference	-0.1045**	-2.05	Difference	0.1470**	2.16
Low institutional ownership (N=263,184)	-0.1714***	-4.45	Low institutional ownership (N=263,133)	0.1914***	4.06
High institutional ownership (N=263,322)	-0.0266	-1.10	High institutional ownership (N=263,317)	0.0146	0.35
Difference	-0.1448***	-3.30	Difference	0.1768***	2.97

Table 9: Alphabetic bias among different groups of market participants

This table summarizes main results from pooled predictive panel regressions of stock turnover and breadth of ownership (panel A and panel B) or coverage (panel C) on *position continuous* (the coefficient of interest) and (lagged) control variables (see table 2 and section 2.2 for details). Panel A compares the aggregate behavior of retail investors from a large discount broker with the behavior of the overall market and the aggregate behavior of mutual fund managers, measured over the same period, with the same controls, and the same econometric approach. Regressions are either based on model 1 or 2. The sample period ranges from January 1991 to November 1996. Retail turnover is computed as the logarithmized ratio of the monthly number of shares traded by individual investors and the number of shares outstanding. We consider all stocks which meet the data requirements of model 1 or 2, respectively. Retail breadth of ownership for a given stock in a given month is ln (1+number of individual investors who own the stock). Both retail turnover and retail breadth of ownership are winsorized each month at the 95th percentile. Panel B compares the behavior of fund managers with the behavior of the overall market. Fund manager turnover and breadth of ownership is based on (changes) in the quarterly Thomson Reuters Mutual Fund Holdings. We average values over the three months of a given quarter, and moreover winsorize fund turnover at the 99th percentile. In Panel C, the dependent variable is either analyst coverage (measured as ln(1+number of analysts providing fiscal year estimates)), national media coverage (measured as ln(number of monthly articles in New York Times, USA Today, Wall Street Journal, and Washington Post)) or local media coverage (measured as ln(number of monthly articles in 41 local newspapers)). In all panels, standard errors are double-clustered by firm and month. Additionally, year dummies are included in the regression. Statistical significance at the ten-, five- and one-percent level is indicated by *, **, and ***, respectively.

	Panel A: Retail investors			Matched comparison (model 1-2)	
	Model 1 (1991-1996)	Model 2 1991-1996)		Fund managers	Market
Turnover	-0.1649*	-0.1395		[-0.077, -0.063]	[-0.087, -0.080]
t-stat	-1.71	-1.50		[-1.54, -1.00]	[-2.44, -1.80]
Ownership	-0.0776**	-0.0646		[-0.060, -0.053]	[-0.061, -0.041]
t-stat	-1.98	-1.60		[-2.65, -2.42]	[-1.45, -1.10]
	Panel B: Mutual fund managers			Matched comparison (model 1-3)	
	Model 1 (1983-2011)	Model 2 (1983-2011)	Model 3 (1995-2009)		Market
Turnover	-0.0669*	-0.0867***	-0.0710**		[-0.120, -0.103]
t-stat	-1.85	-3.06	-2.39		[-4.21, -3.79]
Ownership	-0.0330**	-0.0334**	-0.0301*		[-0.086, -0.081]
t-stat	-1.98	-2.38	-1.81		[-2.23, -1.67]
	Panel C: Analysts and the Media				
Analyst coverage		National media coverage		Local media coverage	
Model 3	+ liquidity/ownership	Model 3	+ liquidity/ownership	Model 3	+ liquidity/ownership
-0.0363**	-0.0252	-0.0112	-0.0104	-0.0403*	-0.0375*
-2.11	-1.57	-0.90	-0.84	-1.79	-1.68

Table 10: International evidence on ordering effects: Predictive panel regressions

This table explores ordering effects in the cross-section of monthly stock-level turnover (Panel A) or of the monthly stock-level Amihud illiquidity ratio (Panel B) in pooled Western stock markets (Australia, Canada, France, Germany, UK) and in the Japanese stock market. For each country, we gather daily stock market data from Datastream, accounting data from Worldscope, and analyst data from I/B/E/S. The sample period ranges from January 2004 to December 2013. Details about the sample construction are provided in the appendix. Displayed are the main results of predictive multivariate panel regressions of the natural logarithm of firm turnover or illiquidity on a set of lagged control variables and a lagged measure of the relative position of a stock in a given ordering scheme. Ordering schemes are either based on alphabetic ordering of firm names (as of December 2013) or, in the case of Japan, based on the local stock code rank assigned by the Securities Identification Code Committee (the national numbering agency). More specifically, for each month and each Datastream level 4 industry, we compute the relative position of a firm's stock code. Thus, both the alphabetic ordering measure and the stock code ordering measure are uniformly distributed over the interval (0,1]. We refer to the resulting variables as *position continuous*. Moreover, *position dummies* (separately for each ordering scheme) are defined as in table 4. Lagged controls include two variables for the return in the previous month, firm size, firm age, stock price, book-to-market ratio, leverage, beta, fifty-two week high, idiosyncratic volatility, alpha, squared market capitalization, momentum formation period return, return twelve months ago, analyst coverage, and analyst forecast dispersion. All regressions also include month and industry dummies. In the case of the Western stock markets, we pool the data, add country fixed effects, and also interact these effects with all control variables (except the industry and month dummies). In all regressions, standard errors are double-clustered by firm and month. Statistical significance at the ten-, five- and one-percent level is indicated by *, **, and ***, respectively.

Panel A: Logarithmized monthly turnover as dependent variable			
Countries	Western Countries	Japan	Japan
Ordering scheme	Alphabetic ordering	Alphabetic ordering	Stock code ordering
N	248,850	269,587	269,587
R^2	0.67	0.50	0.50
Panel A1: Impact of position continuous			
Coefficient of position continuous	-0.121***	0.0102	-0.207***
t-stat of position continuous	(-3.34)	(0.31)	(-5.75)
Panel A2: Impact of position dummies			
Coefficient of position dummy 5	0.164***	0.0557	0.202***
t-stat of position dummy 5	(4.32)	(1.42)	(3.66)
Coefficient of position dummy 25	0.0716**	-0.0403	0.133***
t-stat of position dummy 25	(2.29)	(-1.42)	(4.34)
Coefficient of position dummy 50	0.0239	0.0002	0.0813***
t-stat of position dummy 50	(0.84)	(0.01)	(3.06)
Coefficient of position dummy 75	0.0005	-0.0495*	0.0475*
t-stat of position dummy 75	(0.02)	(-1.76)	(1.88)

Panel B: Logarithmized monthly Amihud illiquidity ratio as dependent variable			
Countries	Western Countries	Japan	Japan
Ordering scheme	Alphabetic ordering	Alphabetic ordering	Stock code ordering
N	248,850	269,587	269,587
R^2	0.89	0.87	0.87
Panel B1: Impact of position continuous			
Coefficient of position continuous	0.121***	-0.0121	0.157***
t-stat of position continuous	(2.86)	(-0.35)	(4.18)
Panel B2: Impact of position dummies			
Coefficient of position dummy 5	-0.178***	-0.0429	-0.183***
t-stat of position dummy 5	(-3.93)	(-1.02)	(-3.23)
Coefficient of position dummy 25	-0.0693*	0.0408	-0.0931***
t-stat of position dummy 25	(-1.89)	(1.36)	(-2.91)
Coefficient of position dummy 50	-0.0218	0.0010	-0.0540*
t-stat of position dummy 50	(-0.64)	(0.03)	(-1.95)
Coefficient of position dummy 75	-0.0127	0.0532*	-0.0267
t-stat of position dummy 75	(-0.39)	(1.81)	(-0.99)

Table 11: Mutual fund flows and alphabetic bias

This table summarizes the main results from predictive regressions of monthly mutual funds flows (winsorized each month at the 2.5% and 97.5% level) on measures of the relative position of fund names within the alphabetically sorted fund universe as well as on a number of control variables. The sample period ranges from January 1992 to December 2012. We focus on domestic open-end U.S. equity mutual funds. In specifications 1 and 2, we use all funds. Models 3 and 4 (5 and 6) only consider funds with total net assets below or equal to (above) the median-sized fund in a given month. *Position continuous* as well as the four *Position dummies* are defined analogously to the analysis in the stock market (see e.g. table 1 and 5). *Fund size* is the natural logarithm of one-month lagged total net assets of the fund (in million USD). *Fund age* is ln(age in months). *Fund expense ratio* and *fund turnover* are lagged by one year and logarithmized. *Length of fund name* is the number of letters of the fund name, after dropping share-class information and incorporation terms where applicable. *Fund return* is the cumulative return (net of fees) over the previous twelve months. *Fund risk* is the standard deviation of the previous twelve monthly return observations. *Fund performance rank* indicates the relative performance of the fund in its market segment (one of 19 styles identified by the CRSP style code) in a given month. At the family-level, *total net assets* and *no. of funds* are lagged by one month and logarithmized. *Top* 3 *fund families* is a dummy which takes on a value of one (zero) if the fund belongs to one of the three largest fund families based on one-month lagged total net assets. *Style growth* is the current growth rate of the fund's market segment (one of 19 styles) due to aggregated fund flows. All regressions contain dummies for each *style − year observation*. Standard errors are double-clustered by fund and month. Statistical significance at the ten-, five- and one-percent level is indicated by *, **, and ***, respectively.

Fund universe	All funds		Small funds		Large funds	
Model specification	(1)	(2)	(3)	(4)	(5)	(6)
Position continuous	-0.162**		-0.267***		-0.0012	
	(-2.51)		(-2.95)		(-0.14)	
Position dummy 5		0.325***		0.515***		0.0978
		(3.72)		(4.31)		(1.03)
Position dummy 25		0.0244		0.0934		-0.0925
		(0.44)		(1.20)		(-1.40)
Position dummy 50		0.0625		0.0321		0.0538
		(1.15)		(0.41)		(0.88)
Position dummy 75		-0.121**		-0.110		-0.153**
		(-2.26)		(-1.55)		(-2.41)
Fund size	-0.0465***	-0.0465***	-0.184***	-0.188***	0.00593	0.0073
	(-2.72)	(-2.73)	(-6.10)	(-6.26)	(0.26)	(0.32)
Fund age	-0.824***	-0.825***	-0.975***	-0.975***	-0.689***	-0.692***
	(-21.86)	(-21.90)	(-19.29)	(-19.28)	(-16.48)	(-16.60)
Fund expense ratio	0.0391	0.0409	0.0271	0.0300	0.0884	0.0959
	(0.64)	(0.68)	(0.36)	(0.40)	(1.10)	(1.19)
Fund turnover rate	-0.0428*	-0.0413*	-0.0125	-0.0129	-0.0814***	-0.0771***
	(-1.76)	(-1.70)	(-0.41)	(-0.42)	(-2.92)	(-2.77)
No. share classes	-0.231*	-0.205	-0.181	-0.151	-0.417***	-0.415***
	(-1.83)	(-1.62)	(-1.08)	(-0.90)	(-3.11)	(-3.08)
Length of fund name	-0.0045***	-0.0046***	-0.0075***	-0.0080***	-0.0012	-0.0014
	(-2.92)	(-2.99)	(-3.50)	(-3.70)	(-0.68)	(-0.78)
Fund return	3.977***	3.973***	4.420***	4.414***	3.643***	3.637***
	(10.75)	(10.75)	(10.37)	(10.37)	(10.27)	(10.26)
Fund risk	0.441	0.471	1.995	2.045	-0.320	-0.320
	(0.22)	(0.24)	(0.87)	(0.89)	(-0.15)	(-0.15)
Fund performance rank	0.266	0.268	0.181	0.184	0.459**	0.472**
	(1.54)	(1.55)	(0.82)	(0.84)	(2.38)	(2.46)
(Fund performance rank)^2	0.826***	0.824***	1.074***	1.070***	0.418**	0.404**
	(4.96)	(4.95)	(4.88)	(4.87)	(2.34)	(2.26)
Fund family: Total net assets	0.173***	0.175***	0.197***	0.200***	0.119***	0.122***
	(7.77)	(7.86)	(6.49)	(6.61)	(3.51)	(3.59)
Fund family: No. of funds	-0.279***	-0.284***	-0.316***	-0.318***	-0.237***	-0.247***
	(-7.25)	(-7.34)	(-5.24)	(-5.27)	(-5.24)	(-5.40)
Top 3 fund families	0.421***	0.409***	0.706***	0.733***	0.387***	0.334***
	(5.27)	(5.01)	(3.75)	(3.87)	(4.49)	(3.73)
Style growth	91.38***	91.39***	91.46***	91.49***	90.82***	90.84***
	(24.06)	(24.07)	(21.46)	(21.48)	(19.79)	(19.80)
Constant	1.858***	1.820***	2.768***	2.668***	2.108**	2.271**
	(2.85)	(2.78)	(3.71)	(3.56)	(2.33)	(2.45)
Style-year fixed effects	yes	yes	yes	yes	yes	yes
Observations	374,566	374,566	187,359	187,359	187,207	187,207
R-squared	0.133	0.133	0.125	0.125	0.164	0.165