Do Fluent Tickers Appeal to Investor Sentiment?

Greg Durham

College of Business
Montana State University
P.O.Box 173040
Bozeman, MT 59717
E-mail: gregdurham@montana.edu

Mukunthan Santhanakrishnan*

Cox School of Business
Southern Methodist University
Fincher Building P.O. Box 750333
Dallas, TX 75275
E-mail: muku@cox.smu.edu

Date of This Draft: August 30th, 2013

---

* Corresponding author. Please contact via e-mail (muku@cox.smu.edu) or phone (214-768-2260).

Do Fluent Tickers Appeal to Investor Sentiment?

Abstract

The objective of this study is to examine the roles of linguistic fluency and investor sentiment in asset valuation. Employing an innovative measure for fluency of ticker symbols, we examine whether stock returns differ across stocks with tickers of different fluencies, depending on the level of investor sentiment that characterizes the marketplace. We find that when incoming sentiment is high, stocks with most-fluent tickers have lower returns than stocks with least-fluent tickers have. This study contributes to the literature by documenting that stock prices are affected by certain characteristics of securities that have no bearing on stocks' underlying cash flows or required returns.

Acknowledgements: The authors especially thank Russell Jame for his thorough, helpful feedback on this paper. We also thank James Lin, Jared DeLisle, Dan Rogers, and John Settle, as well as seminar participants at Portland State University in 2010 and at the Behavioral Decision Research in Management Conference in 2012.

Keywords: fluency; sentiment, ticker symbols; behavioral finance; asset pricing

JEL Classification Codes: **G120** Asset Pricing; Trading Volume; Bond Interest Rates, and **G140** Information and Market Efficiency; Event Studies.

1. Introduction

In a frictionless market with rational investors, an asset's expected return is based solely on its expected future cash flows and its current price. Price incorporates required return, which is a function of systematic risk. Therefore, holding expected future cash flows constant, variation in expected returns on assets is solely a function of variation in systematic risks associated with the assets. Expected returns should not be influenced by the assets' identifying characteristics (such as company names or trading symbols). Recent empirical evidence, however, shows that time-varying sentiment does have a significant influence on cross-sectional variation in stock returns. In addition, a growing body of literature sheds light on the impact of fluency of firm names and ticker symbols on both asset valuation and investor base. The objective of this study is to bridge the two strands of research on the roles of investor sentiment and linguistic fluency in asset valuation. Specifically, we examine whether the effects of ticker-symbol fluency on stock returns vary, depending on the level of investor sentiment that characterizes the marketplace.

Baker and Wurgler (2006) are the first researchers to construct a unique sentiment index that encompasses six well-established proxies for investor sentiment. They document that when sentiment is high at the beginnings of periods, subsequent returns are low for stocks of young firms, stocks with higher arbitrage costs, and other hard-to-value stocks. Other studies further corroborate the role of sentiment in asset valuation. Lemmon and Portniaguina (2008) find results similar to those of Baker and Wurgler, while using consumer sentiment instead of investor sentiment. More recently, Billett *et al*. (2012) find that when incoming sentiment is high, stocks associated with high-prestige brands exhibit future returns that are lower than those on stocks of low-prestige brands.

Securities' identifying characteristics can also influence asset valuation. Rashes (2001) finds that stock prices of companies that have similar ticker symbols (or tickers, for short) tend to exhibit comovement in returns, possibly because investors get confused between the tickers. In a similar vein, Cooper, *et al*. (2001) discuss a case wherein investors, in response to an initial public offering (IPO) filing by AppNet Systems, bought stock in and, hence, increased the stock price of Appian

Technology (whose ticker APPN could potentially be inferred to belong to AppNet). Furthermore, Misra and Kadapakkam (2007) find that changes in ticker symbols are associated with changes in trading volumes and prices surrounding the effective dates. All three studies suggest that investors do devote attention to companies' ticker symbols.

Since investors are under cognitive overload when facing a multitude of investment options, they are likely to rely on mental shortcuts or heuristics when processing complicated information about the various options. Therefore, such investors are likely to prefer stocks with tickers that are familiar, easy to process, or both. Alter and Oppenheimer (2006) find that stocks with tickers that are more easily pronounceable outperformed stocks with tickers that are harder to process. Head *et al*. (2009) conduct a similar study and find that a portfolio of stocks with 'clever' tickers exhibits abnormal returns, even after controlling for the well-known Fama-French factors (Fama and French (1993)) and a momentum factor (Carhart (1997)). Green and Jame (2012) show that investor behavior is influenced by fluency of company names, documenting that firms that have fluent names have greater investor recognition and higher valuation. These studies suggest that investors are influenced by fluency of firms' identifying characteristics. The combined evidence shows that attention to tickers and fluency of tickers both impact the investment decision.

We strive to connect these bodies of research by examining the interplay among investor sentiment, fluency of tickers, and asset valuation. As noted earlier, Baker and Wurgler (2006) argue that market-level sentiment and investor demand for stocks with certain characteristics are correlated. Alter and Oppenheimer (2008) suggest that people perceive greater value in items that are more fluent than in items that are not so easily processed. We posit that sentimental investors are likely to be persuaded by the fluency of a specific salient characteristic, namely, a stock's ticker. We anticipate that in periods with high incoming sentiment, stocks with more fluent tickers will initially be valued more highly than stocks with low-fluency tickers will be, leading to lower returns on the former group of stocks during such periods. We anticipate the converse relation in periods for which incoming sentiment is low.

We employ an innovative measure of fluency for ticker symbols in this study. This measure is based on an algorithm pioneered by Travers and Olivier (1978) and also employed by Green and Jame (2012). The algorithm assigns an "Englishness" value to any given succession of letters, based on the frequency with which each given cluster of letters within the succession appears in the English language.

After establishing a fluency value for every ticker in the CRSP[1] universe of stocks from 1966 through 2010, we begin by performing our own analogous versions of two studies that have previously found relations between ticker-symbol characteristics and stock returns. Alter and Oppenheimer (2006) find statistically significant differences between one-day returns on stocks with tickers that are deemed pronounceable and returns on stocks in a complement set, using IPO dates as event dates. Similar to what Alter and Oppenheimer find, we find statistically significant differences between returns on stocks with most-fluent tickers and returns on stocks with least-fluent tickers, also focusing on IPO dates from which we measure returns. Head *et al*. (2009) find abnormal returns (beyond those explained by well-known risk factors) on a portfolio of stocks with "clever tickers". Similar to Head *et al*.'s findings, we find abnormal returns on a portfolio of stocks with most-fluent tickers.

Next, we employ a method similar to that of Baker and Wurgler (2006) and perform monthly, fluency-based sorts on the same CRSP universe of stocks. Using the monthly sorts, we form monthly zero-cost portfolios that are long in the quintile of stocks with the most-fluent tickers and short in the quintile of stocks with the least-fluent tickers. For each zero-cost portfolio, we calculate its return as the difference between the two extreme portfolios' equally-weighted monthly returns. We then regress the portfolio returns on incoming investor sentiment and on the four Fama-French factors, and we find a negative relation between incoming sentiment and monthly returns. This relation implies that when incoming sentiment is high [low], subsequent returns on stocks with highly-fluent tickers are less [greater] than returns on stocks with tickers of low fluency.

Our study advances the literature in that it is the first to jointly examine the fluency of ticker

[1] CRSP is the abbreviation for the Center for Research in Securities Prices.

symbols and overall levels of sentiment in the marketplace. We demonstrate that stock returns are related to the accessibility of a particular characteristic (i.e., the ticker symbol) that has no bearing on firms' underlying cash flows, and that the type of relation is dependent upon level of sentiment.

This paper proceeds with a literature review and a development of our hypotheses in section 2. It continues with a discussion of our dataset and variables in section 3. Section 4 follows with an explanation of our methodology and an analysis of our findings. Section 5 concludes.

2. Background and development of hypotheses

Under conditions of uncertainty and bounded rationality (Simon (1955), Simon (1979)), and when decisions become overly complex, individuals often employ cognitive shortcuts to make decisions. Kahneman (2003) identifies two generic modes of cognitive function: (1) an intuitive mode in which judgments and decisions are made instinctively, instantaneously, implicitly, and emotionally, and (2) a control mode in which decisions are made more deliberately, logically, and explicitly, with rationality and reason. Kahneman (2003) posits that in settings with uncertainty and information complexity, decisions are often made through the intuitive mode as opposed to via the controlled-reasoning mode. We suggest that in similar settings, individual investors may rely on mental shortcuts in order to simplify their menu of investment options and reduce it to a more tractable set.

Empirical evidence supports the notion that investors do rely on shortcuts when making investment decisions. For example, investors prefer to invest in companies whose headquarters are located closer to them (Coval and Moskowitz (1999)) and in companies with management teams that speak the same language as they do (Grinblatt and Keloharju (2011)). Barber and Odean (2008) demonstrate that investors are prone to buy stocks of companies that are in the news, stocks that experience abnormally-high trading volume, and stocks with very high single-day returns.

Alter and Oppenheimer (2006) find evidence consistent with cognitively-constrained investors turning to a particular salient characteristic (namely, ticker symbol) when making investment decisions. Investors appear to be more inclined to buy stocks with fluent tickers than they are to buy stocks with non-fluent tickers. Perhaps such inclination exists because tickers serve as mental shortcuts or because, as Alter and Oppenheimer (2008) suggest, investors are attracted to items that are more fluently, cognitively processed.

Head *et al*. (2007) examine stocks with what they call "clever tickers", tickers that are witty in such ways that the tickers might linger longer in investors' memories. Stocks with clever tickers might make for easier recall by an investor making a current set of investment decisions. The authors regress excess returns on a portfolio of clever-ticker stocks on three Fama-French factors (Fama and French (1993) and a momentum factor (Carhart (1997)). They find abnormal returns on the portfolio; i.e., the portfolio's returns (in excess of the risk-free rate) are beyond those that are explainable by the well-known market, size, value, and momentum factors.

Investors are likely to be influenced not only by stock-level characteristics such as fluency, but also by overall levels of sentiment in the investing marketplace. When overall market sentiment is high, investors who face cognitive constraints might focus more on assets' attributes that are familiar or easy to process. Billett *et al*. (2012) show that a product-market attribute (namely, brand prestige) appeals to investors; we suggest that tickers and their potential ease of processing might also appeal to investors.

Consequently, when sentiment is high, it can cause stocks with highly fluent tickers to trade at prices that are above fundamental values. Thus, these stocks are likely to exhibit low returns in periods subsequent to the high-sentiment periods. We anticipate that if individual investors are affected by the fluency of ticker symbols, then when incoming sentiment is high, returns on a portfolio of high-fluency-ticker stocks will be lower than returns on a portfolio of low-fluency-ticker stocks. When incoming investor sentiment is low, returns on high-fluency stocks will be greater than returns on low-fluency stocks.

3. Data and variables

For our analysis, we utilize the following variables that might impact investor behavior and stock returns: ticker symbols, fluencies of tickers, an index that captures market-wide levels of sentiment, and the four Fama-French factors that are well-documented as explaining much of the cross-sectional variation in returns. We also use stock returns to construct the dependent variables in our study.

3.1. Ticker symbols

Our study uses data from CRSP for 22,458 stocks, spanning the years 1966 through 2010. For each stock, we are interested in the set of ticker symbols attached to the stock over our sample period (recognizing that in many cases, a stock maintains the same ticker over the entire horizon). Over our evaluation horizon, the CRSP database shows 22,774 unique tickers. Due to how we specify our fluency measure (described in the next section), we must exclude all stocks for which the affiliated ticker has fewer than three characters, thereby reducing the number of unique stocks to 21,572 and the number of unique tickers to 22,287. Some stocks have been assigned more than one ticker during our sample period, and some tickers have been assigned to more than one stock, resulting in 27,753 different stock-ticker combinations.

3.2. Fluency

For every ticker affiliated with the stocks in our sample, we calculate a fluency measure using a revised version of a linguistic algorithm created by Travers and Olivier (1978) (T&O). The T&O algorithm calculates (what the authors call) Englishness for any given sequence of characters as the product of one probability and a series of conditional probabilities. The first term is the probability of observing a space. The second term is the probability of observing the first letter in the sequence of characters, conditional on the preceding character being a space. The third term is the probability of observing the second letter in the sequence, conditional on the preceding two characters being a space followed by the first letter of the sequence. Each additional term thereafter in the calculation is the conditional probability of observing a particular letter following a particular pair of characters. The final term is the conditional probability of observing a space following the final pair of characters in the sequence.

Our transformation of T&O's Englishness equation is the same as the twofold transformation performed by Green and Jame (2012). First, we replace each T&O probability with ratios of relative frequencies of trigrams and bigrams, acknowledging that this transformation does not contain analogues for the first two terms in T&O's original specification.[2] Next, we perform a logarithmic transformation of the product of relative frequencies. This calculation creates fluency values that are increasing in degree of Englishness. Our exact equation for fluency of a stock's ticker is as follows:

$$\begin{aligned} \textit{Fluency} \;=\; \{ & \log F(\#L_1L_2) + \log [\, F(L_1L_2L_3)/F(L_1L_2) \,] + \log [\, F(L_2L_3L_4)/F(L_2L_3) \,] \\ & + \log [\, F(L_3L_4L_5)/F(L_3L_4) \,] + \log [\, F(L_4L_5\#)/F(L_4L_5) \,] \}, \end{aligned} \tag{1}$$

where F( ) represents frequency and # represents a space. L_1 , L_2, L_3, L_4, and L_5 represent the first, second, third, and (as applicable) fourth and fifth characters in the ticker symbol. For a stock with a three-letter ticker, the summation involves only three terms, with the final term being the log of the ratio $F(L_2L_3\#)/F(L_2L_3)$. For a stock with a four-letter ticker, the summation involves four terms, with the fourth term being the log of the ratio $F(L_3L_4\#)/F(L_3L_4)$.

To account for all of the frequencies that are used in our numerous Englishness calculations, we rely on *The Corpus of Contemporary American English* (*CoCA*), an extensive dataset that provides estimates of frequencies of words and character strings, based on examination of over 160,000 texts published between 1990 and 2010, inclusively.[3]

Figure 1 shows the relative frequencies of trigrams to bigrams as well as the resulting values for fluency, for a representative set of tickers. We report values for the tickers GRO, INFO, INMD, TXN, and NTRT, as these tickers' fluency values are the median values for the five quintiles of stocks that emerge when the stocks are sorted by their tickers' fluency. We also report numbers for five additional tickers: THE, FOOD, EASY, IDEA, and COMMA. Not surprisingly, THE is the ticker among all three-, four-, and five-digit tickers that exhibits the very highest degree of Englishness.[4] The other four tickers are, in our opinion, simply catchy and fun. As one example of a calculation using equation 2 above, the fluency of FOOD is calculated as follows:

[2] As previously mentioned, we exclude tickers with fewer than three characters. The reason for the exclusion is that our algorithm relies on trigrams.

[3] The corpus can be accessed at http://corpus.byu.edu/coca. It was created and is maintained by Mark Davies, Professor of Corpus Linguistics at Brigham Young University.

[4] Though its details are not reported in Figure 1, HMFRV is at the other extreme from THE: it is the least-fluent ticker among the tickers in our final sample.

$$\textit{Fluency} = \{ \log F(\#FO) + \log [F(FOO)/F(FO)] + \log [F(OOD)/F(OO)] + \log [F(OD\#)/F(OD)] \}$$

$$\textit{Fluency} = \{ \log 5598204 + \log [235574/6908052] + \log [985282/4448792] + \log [932219/2604861] \}$$

$$\textit{Fluency} = 9.62541.$$

We obtain these frequency counts from the *CoCA* database in the following manner. To get the value for F(*#FO*), we sum the frequencies of all words and character strings that begin with the string FO. To get the values for F(*FOO*) and F(*FO*), we sum the corresponding frequencies of all of the words and character strings that contain the sequences FOO and FO, respectively. Our accumulation of frequencies continues in similar fashion until we reach the final ratio. F(*OD#*) is calculated by summing the corresponding frequencies of all words and character strings that end with OD, and F(OD) reflects the sum of frequencies of all of the words and character strings that contain OD.

Due to occasional missing frequency counts, we are unable to calculate this fluency measure for 2060 tickers in our sample, resulting in 20,037 unique stocks, 20,068 unique tickers, and 25,293 different stock-ticker combinations. Table 1 shows descriptive statistics for *Fluency*, a unitless variable. The mean and median values are –0.3568 and –0.2117, respectively. Values range from a minimum of –23.6383 to a maximum of 16.6095. Exhibit 1 lists the twenty tickers with the highest values for Fluency, as well as the ten tickers that are calculated as being the least fluent. Tickers with the highest fluencies include THE, AND, FOR, THER, and WAS, among others.

3.3. Investor sentiment

One key explanatory variable in our study is the incoming level of investor sentiment (*Sentiment*), which we suggest will affect stock returns in the subsequent period. For our *Sentiment* variable, we use the monthly values of the Investor Sentiment index, an index constructed, calculated, and reported by Jeffrey Wurgler.[5] The index is a composite measure that incorporates six different, well-established proxies for investor sentiment. The six specific variables include the closed-end mutual fund discount, the turnover of shares on New York Stock Exchange stocks, the number of IPOs, the average first-day return on IPOs, the proportion of new equity issues relative to all new debt and equity issues, and the dividend premium. Each of these six proxies for sentiment is described in detail by Baker and Wurgler (2006). The authors employ a principal-components analysis to identify a component within each of the six indices that is common to all. The isolated component that emerges is reasonably assumed to be the level of overall investor sentiment that is common to all six proxies.[6] In our study, we utilize monthly index values for the years 1966 through 2010. Table 1 shows descriptive statistics for the index, a unitless variable. Across our sample period, the index ranges from –2.5480 to 2.4220, with a mean of 0.0171 and median of 0.0225.

3.4. Fama-French factors

Some of our tests will employ monthly values for the three Fama-French factors, widely regarded as capturing returns associated with various types of systematic risk (Fama and French (1993)). *Market* is the market risk premium (the return on a market portfolio minus the prevailing risk-free rate), *SMB* is the size premium (the average return on stocks in the three smallest size deciles minus the average return on stocks in the three biggest deciles), and *HML* is the value premium (the average return on stocks in the two highest book-to-market-value-ratio deciles minus the average return on stocks in the two lowest deciles). Alternate specifications will also employ *UMD*, the momentum premium (initially documented by Carhart (1997) and calculated as the av-

[5] The URL for the database is http://people.stern.nyu.edu/jwurgler/data/Investor_Sentiment_Data_v23_POST.xlsx. This link is valid as of August 2012. The database was created and is maintained by Jeffrey Wurgler, Nomura Professor of Finance at the Stern School of Business at New York University and Research Associate at National Bureau of Economics Research (http://people.stern.nyu.edu/jwurgler).

[6] Recent studies on investor sentiment have employed this same index. See, for example, studies by Billett *et al*. (2012), Hribar and McInnis (2012), and Stambaugh, Yu, and Yuan (2012).

erage return on stocks in the two highest prior-period-return ("up") deciles minus the average return on stocks in the two lowest ("down") deciles).[7]

Having now established our various variables, we proceed to an explanation of our methodology and findings.

4. Methodology and Findings

We will begin by performing a pair of tests that are analogous to previous studies that have found relations between ticker-symbol characteristics and stock returns. Next, we will construct monthly zero-cost portfolios that are long in stocks with most-fluent tickers and short in stocks with least-fluent tickers. Finally, we will perform a pair of ordinary-least-squares (OLS) regression analyses, wherein the dependent variable will be these zero-cost portfolios' returns and the explanatory variables are *Sentiment* and either three or four additional risk factors, depending on the specification.

4.1. Ticker fluency and stock returns

Our first two tests are motivated by previous tests performed by Alter and Oppenheimer (2006) and Head *et al*. (2009), respectively. Alter and Oppenheimer examine whether differences in the degrees to which stock tickers are pronounceable might relate to differences in stock returns, and we focus similarly on differences in fluency and how they relate to differences in returns. In addition, while Head *et al*. highlight a portfolio of stocks with what the authors refer to as "clever tickers", our emphasis will be on a portfolio of stocks with the most-fluent tickers.

In their test, Alter and Oppenheimer select firms that completed IPOs of their stocks during their sample period covering 1990 through 2004: 665 from the New York Stock Exchange and 116 from the American Stock Exchange. Two human "coders" used "subjective impression" to sort stocks into one group with pronounceable tickers and one group with unpronounceable tickers. The authors calculate mean returns for each group over various horizons and find statistically significant differences in one-day returns across the two groups and no differences across the groups

[7] Values for each of these four factors can be found at the Data Library on the website of Professor Kenneth R. French, the Roth Family Distinguished Professor of Finance at the Tuck School of Business at Dartmouth College. The URL for the Data Library is http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ data_library.html.

for returns over longer periods.

As shown in Table 2, we perform similar analyses of differences in means. Using all of the 11,125 stocks that experienced IPOs during our sample period, we sort the stocks into quintiles based on *Fluency*. Using IPO dates as event dates, we then calculate mean one-, thirty-, and sixty-day returns for the quintile of stocks with most-fluent tickers, as well as for the quintile with least-fluent tickers. The difference in mean one-day returns is 0.1248, statistically significant at the 1% level. The differences in mean 30- and 60-day returns are 0.1031 and 0.1043, respectively; both differences are also statistically significant at the 1% level. Our findings confirm the results of Alter and Oppenheimer's (2006) similar test but with a more objective measure of fluency: investors do appear to respond differently to IPOs of stocks depending on the fluency of a stock's ticker symbol, with a preference for fluency.

Head *et al*. (2009) perform an OLS regression analysis wherein the dependent variable is the monthly excess return on a portfolio that invests equal weights in stocks that have clever tickers.[8] We proceed with an analogous test, but our dependent variable (*Return*) is the monthly excess return on an equally-weighted portfolio that rebalances monthly and invests in the quintile of stocks with the most-fluent tickers. The four explanatory variables in our specification are common to the Head *et al*. regression, as well:

$$Return = \alpha + \beta_{Market} \cdot Market + \beta_{SMB} \cdot SMB + \beta_{HML} \cdot HML + \beta_{UMD} \cdot UMD, \tag{2}$$

The explanatory variables are the three Fama-French factors and the momentum premium, as described above in section 3.4.

The two columns of results in Table 3 are for two alternate specifications of equation 2,employing the same explanatory variables as those used by Head *et al*. (2009). Our first model uses the three Fama-French factors, and the second model uses these same three factors along with the momentum factor (*UMD*). Our findings relating to fluency of stocks are similar to those of Head

[8] The excess return is calculated as the portfolio's equally-weighted return minus the prevailing rate on US Treasury bills.

et al. (2009) for clever-ticker stocks. The alpha in Model 2 equals 0.0019 and is statistically significant at the 1% level, indicating that stocks with most-fluent tickers have abnormal positive returns, returns beyond those that can be attributed to premiums associated with market risk, size, value, and momentum. The alpha in the first specification is smaller (0.0005) and statistically insignificant, perhaps driven towards zero due to it encompassing the omitted momentum variable that emerges with a statistically significant, negative coefficient in Model 2.

We are encouraged by the fact that our combined results from Tables 2 and 3 are similar to those of Alter and Oppenheimer (2006) and Head *et al*. (2009), as these similarities help to validate our chosen measure of fluency. As described in detail in section 3.2, our fluency variable is a transformation of an algorithm developed by Travers and Olivier (1978), a transformation that closely follows one performed by Green and Jame (2012). With our fluency measure now established, we proceed to examine whether investors respond differently to the fluency of ticker symbols, depending on the level of market-wide investor sentiment.

4.2. Investor sentiment, fluency, and stock returns

This section contains the highlight of our study: an examination of the interplay among investor sentiment, fluencies of tickers, and asset valuation. Baker and Wurgler (2006) find that market-level sentiment and demand for stocks with certain salient characteristics are related. Within this context, we are specifically interested in whether investors are persuaded by fluencies of ticker symbols, as well as whether the degree of persuasion is a function of the level of sentiment that describes the investing marketplace.

We anticipate that in periods defined by high incoming sentiment, stocks with more fluent tickers will initially be valued more highly than stocks with low-fluency tickers will be, resulting in lower returns on the stocks with more fluent tickers during these periods. We expect the converse relation to be true in periods for which incoming sentiment is low.

Our method is similar to that developed by Baker and Wurgler (2006). We begin by performing monthly sorts on our universe of stocks from 1966 through 2010, sorting stocks into quintiles based on *Fluency* as calculated per equation 1 in section 3.2. Next, month by month, we construct zero-cost portfolios that are long in the quintile of stocks with the most-fluent tickers and short in the quintile of stocks with least-fluent tickers. For each monthly zero-cost portfolio, we calculate its return as the difference between the two portfolios' equally-weighted monthly returns; this variable is the key dependent variable in our study.

Having established this vector of monthly portfolio returns, we next assemble values for our key explanatory variable in this study, namely, incoming investor sentiment (*Sentiment*) as described in section 3.3. We also obtain monthly values for the four Fama-French factors, already commonly known but nonetheless briefly described in section 3.4 – these variables are included as controls. We then perform a pair of OLS regression analyses, the results of which appear in Table 4. In our first analysis, the explanatory variables include *Sentiment*, as well as three Fama-French factors as control variables:

$$Return = \alpha + \beta_{Sentiment} \cdot Sentiment + \beta_{Market} \cdot Market + \beta_{SMB} \cdot SMB + \beta_{HML} \cdot HML, \qquad (3)$$

As shown in the column labeled Model 1, the coefficients on the three control variables are all statistically insignificantly different from zero, as would be expected when the dependent variable is a difference in returns across two portfolios of stocks that should not be different from each other in terms of the risk characteristics for which the control variables proxy. Most importantly to our study, the coefficient on *Sentiment* equals –0.0015 and is statistically significant at the 1% level. The interpretation of this negative, statistically significant coefficient is that when incoming sentiment is high [low], subsequent returns on stocks with highly-fluent tickers are less [greater] than returns on stocks with tickers of low fluency.

Our second specification is the same as that shown in equation 3 except for that it includes the momentum premium, *UMD*; the results for this specification are shown in Table 4 in the column labeled Model 2. Again, the coefficients on our control variables are statistically insignificant. The coefficient on *Sentiment* again equals –0.0005 and is again statistically significant at the 1% level.[9] The interpretation remains the same as for Model 1.

---

[9] We also specified our *Fluency* variable to account for the number of characters in each ticker because, by the very nature of the *Fluency* calculation, each extra character necessarily increases the numerical result from the calculation.

These findings in Table 4 are robust to our specification of the *Sentiment* variable. We alternately specify this index using lagged sentiment from the end of the preceding calendar year, as opposed to from the previous month. Additionally, we use changes in the sentiment-index values as our explanatory variable. We also employ a binary transformation of the sentiment index that depends on whether the index is positive or negative, as well as Baker's and Wurgler orthogonalized version of the sentiment index.

5. Conclusion

Baker and Wurgler (2006) have shown that variation in market-wide investor sentiment can explain variation in returns on certain categories of stocks. Various other studies have shown that investors do appear to sometimes make investment decisions based on ticker symbols; specifically, some of these studies demonstrate that investors are influenced by ticker traits such as their ease of pronunciation and their "cleverness". We are primarily interested in knowing whether the investing marketplace responds differently to the fluency of tickers depending on the level of investor sentiment that describes the market.

One of our study's innovative features is our fluency measure for ticker symbols. In constructing our fluency variable, we follow Green and Jame (2012) in performing a transformation of an Englishness equation developed by Travers and Olivier (1978), and also in then using data from *The Corpus of Contemporary American English* to calculate the fluency of every ticker symbol in our sample.

Using this fluency measure, we perform tests which are, for our database and variables, analogous to a pair of studies that document relations between ticker-symbol characteristics and corresponding stock returns. We find the same relation for our stocks with most-fluent tickers that Alter and Oppenheimer (2006) find for stocks with tickers that are easiest to pronounce, and we find the same results for our same subset of fluent-ticker stocks that Head, *et al*. (2009) find for stocks with "most-clever" tickers. These findings help to validate our choice (and construction)

Therefore, our alternate specification scales each *Fluency* calculation by the number of characters in the ticker. Results are similar to those from our original specification, albeit with slightly less statistical significance of the estimated coefficients.

of the fluency variable.

Having established this key variable, our findings culminate with our most important result, which is that stock returns differ across stocks with tickers of different fluencies, depending on the level of investor sentiment that characterizes the marketplace. We find that when sentiment is high, stocks with most-fluent tickers are valued more than stocks with least-fluent tickers are, leading to lower returns in the following period on the first group of stocks compared to the returns on the second group. We find that the converse relation exists when sentiment is low.

This paper makes no claim about any specific mechanism via which these relations emerge between the level of market-wide sentiment and differential returns across our specific fluency-based subsamples of stocks. One potential explanation is as follows. In periods of high sentiment, a greater proportion of sentimental investors populates the marketplace relative to rational traders or, alternately, any individual trader is simply characterized by a higher level of sentiment and therefore by less rationality. In turn, during these same periods, the marketplace will be characterized by investors who, knowingly or unknowingly, are more likely to be attracted to stocks with highly-fluent tickers and perhaps be averse to stocks with least-fluent tickers. To explore these potential explanations might be a particularly fruitful endeavor to any researcher who has access to individual-investor portfolio data. In support of any such endeavor, we will gladly share our database of fluency values for every ticker symbol that the CRSP database comprises.

References

Alter, A.L., Oppenheimer, D.M., 2006. Predicting short-term stock fluctuations by using processing fluency, Proceedings of the National Academy of Sciences 103, 9369-9372.

Alter, A.L., Oppenheimer, D.M., 2008. Easy on the Mind, Easy on the Wallet: The Roles of Familiarity and Processing Fluency in Valuation Judgments, Psychonomic Bulletin & Review 15(5), 985-990.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns, Journal of Finance 61(4), 1645-1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market, Journal of Economic Perspectives 21(2), 129-151.

Barber, B.M., Odean, T., 2008. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors, The Review of Financial Studies 21(2), 785-818.

Billett, M.T., Jiang, Z., Rego, L., 2012. Glamour brands and glamour stocks. Working paper, Indiana University.

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

Cooper, M.J., Dimitrov, O., Rau, P.R, 2001. A rose.com by any other name, Journal of Finance 56(6), 2371-2388.

Coval, J.D., Moskowitz, T.J., 1999. Home bias at home: Local equity preference in domestic portfolios, Journal of Finance 54(6), 2045-2073.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on bonds and stocks, Journal of Financial Economics 33(1), 3-56.

Green, T.C., Jame, R., 2012. Company name fluency, investor recognition, and firm value. Working paper, Emory University.

Grinblatt, M., Keloharju, M., 2001. How distance, language, and culture influence stockholdings and trades, Journal of Finance 56(3), 1053-1073.

Head, A., Smith, G., Wilson, J., 2009. Would a stock by any other ticker smell as sweet?, Quarterly Review of Economics and Finance 49(2), 551-561.

Hribar, P., McInnis, J.M., 2012. Investor sentiment and analysts' earnings forecast errors, Management Science 58(2), 293-307.

Kahneman, D., 2003. A perspective on judgment and choice: mapping bounded rationality, American Psychologist 58(9), 1645697-720.

Lemmon, M., Portniaguina, E., 2006. Consumer confidence and asset prices: some empirical evidence, The Review of Financial Studies 19(4), 1499-1529.

Kadapakkam, P.-R., Misra, L., 2007. What's in a nickname? Price and volume effects of a pure ticker symbol change, Journal of Financial Research 30(1), 53-71.

Rashes, M.S., 2001. Massively confused investors making conspicuously ignorant choices, Journal of Finance 56(5), 1911-1927.

Simon, H.A., 1955. A behavioral model of rational choice, The Quarterly Journal of Economics

69(1), 99-118.

Simon, H.A., 1979. Rational decision-making in business organizations, American Economic Review 69(4), 493-513.

Stambaugh, R.F., Yu, J., Yuan, Y., 2012. The short of it: investor sentiment and anomalies, Journal of Financial Economics 104(2), 288-302.

Travers, J., Olivier, D., 1978. Pronounceability and statistical 'englishness' as determinants of letter identification, American Journal of Psychology 91(3), 523-538.

Figure 1: **Representative Fluency Calculations**
This figure shows the relative frequencies of trigrams to bigrams as well as the resulting values from our calculations of *Fluency*, for a representative set of tickers. Equation 1 shows the calculation:

$$Fluency = \{ \log F(\#L_1L_2) + \log [F(L_1L_2L_3)/F(L_1L_2)] + \log [F(L_2L_3L_4)/F(L_2L_3)] + \log [F(L_3L_4L_5)/F(L_3L_4)] + \log [F(L_4L_5\#)/F(L_4L_5)] \}$$

The values of *Fluency* for GRO, INFO, INMD, TXN, and NTRT are the median values for the five quintiles of stocks that emerge when our sample of tickers are sorted by fluency. We also report *Fluency* calculations for five additional tickers: THE, FOOD, EASY, IDEA, and COMMA.

Ticker	$F(\#L_1L_2)$	$F(L_1L_2)$	$F(L_1L_2L_3)$	$F(L_2L_3)$	$F(L_2L_3L_4)$	$F(L_3L_4)$	$F(L_3L_4\#)$	*Fluency*
GRO	1672143	2809596	644309	10442621	122101	n/a	n/a	8.4082
INFO	12346290	35120364	392661	785438	269848	6908052	6611	3.8152
INMD	12346290	35120364	30434	412577	132	14728	8982	0.7360
TXN	1411	2398	8	628	12	n/a	n/a	–2.4085
NTRT	2923	13609244	858165	5391602	73	5006446	1453677	–7.2299
THE	40997512	50246860	32036866	44228826	27656742	n/a	n/a	16.6095
FOOD	5598204	6908052	235574	4448792	985282	2604861	932219	9.6245
EASY	932256	11410666	1404005	12795042	85299	642870	164571	5.2769
IDEA	312252	5484352	2239457	9919416	541964	323264	11410666	5.2850
COMMA*	8211006	10581219	2672816	8352633	84555	1188561	120451	6.0095

*So as to not disrupt the table, we report the final two frequencies for COMMA, our only five-character ticker, here: $F(L_4L_5)$ = 7722912 and $F(L_4L_5\#)$ = 182576.

Exhibit 1: **The 20 Most Fluent Tickers and the 10 Least Fluent Tickers**
This exhibit shows our sample's 20 most fluent tickers and the 10 least fluent tickers (with their respective values for *Fluency* included in parentheses).

THE (16.6095)	THIS (13.4545)	UAECB (–19.5728)
AND (15.4423)	COM (13.2482)	FDLNB (–19.8043)
FOR (14.3506)	WER (13.1175)	DGTL (–20.1175)
THER (14.2649)	SAY (13.1114)	DRNKV (–20.1981)
WAS (14.2228)	PRE (13.1101)	SCTTB (–20.8246)
HAT (13.9686)	TOR (13.0756)	LBTYB (–21.6180)
CON (13.9087)	WOR (13.0386)	FTTRV (–22.2515)
WIT (13.6782)	RES (13.0118)	GLCNV (–23.0769)
SHE (13.6407)	ALL (12.9707)	IEIBV (–23.3550)
HIS (13.5387)	IND (12.9640)	HMFRV (–23.6383)

Table 1: **Descriptive Statistics for Two Key Explanatory Variables: Ticker Fluency and Investor Sentiment**

Our final sample contains 20,068 unique tickers, as described in section 3.2 of this paper. It also contains 540 monthly observations for Baker and Wurgler's Sentiment Index, described in section 3.3. This table reports numbers of observations, means, medians, standard deviations, maximums, 95th-percentile values, 5th-percentile, and minimum values, for two key variables in our study: *Fluency Sentiment*. Both measures are unitless.

Measure	***Fluency***	***Sentiment***
N	20,068	540
Mean	–0.3568	0.0171
Median	–0.2117	0.0225
Standard Deviation	5.8675	0.9923
Maximum	16.6095	2.4220
95th Percentile	9.1144	1.8860
5th Percentile	–10.3081	–1.7320
Minimum	–23.6383	–2.5480

Table 2: **Mean Returns and Differences in Means Across Fluency Quintiles**
This table reports mean cumulative returns over horizons of different lengths. We identify every stock in our sample for which initial trading commenced at some date within our sample period. For each stock in this subsample analysis, we identify its date of initial public offering and calculate cumulative returns over 1 day, 30 days, and 60 days. We sort stocks into quintiles based on the *Fluency*, as defined per Equation 1 in the text. For the quintile of stocks with the most-fluent tickers and for the quintile of stocks with the least-fluent tickers, we calculate mean cumulative returns over the three respective horizons. The rightmost column reports differences in mean returns. (The number of observations included in each quintile calculation is reported in parentheses underneath the corresponding mean. T-statistics are reported in parentheses.)

Returns Horizon	Mean Cumulative Return on Stocks with Least-Fluent Tickers (# of Obs.)	Mean Cumulative Return on Stocks with Most-Fluent Tickers (# of Obs.)	Difference in Means (t-statistic)
1 day	0.0031 (2210)	0.1279 (2209)	0.1248 (2.63)
30 days	0.0146 (2210)	0.1176 (2209)	0.1031 (2.25)
60 days	0.0370 (2210)	0.1313 (2209)	0.1043 (2.2e5)

Table 3: **Regression Analysis of Returns on Stocks with Most-Fluent Tickers**
This table reports the results of two different ordinary-least-squares regression analyses described by equation 2 in the text:

$$Return = \alpha + \beta_{Market} \cdot Market + \beta_{SMB} \cdot SMB + \beta_{HML} \cdot HML + \beta_{UMD} \cdot UMD,$$

The monthly dependent variable is the equal-weighted return on the quintile of stocks with the most-fluent tickers in excess of the prevailing risk-free rate, as described in section 4.1. The monthly explanatory variables (*Market*, *SMB*, *HML*, and *UMD*) are the three Fama-French factors and the momentum premium, as described in section 3.4. The specification referred to as Model 1 below employs the three Fama-French factors; Model 2 also employs the momentum factor. (T-statistics are reported in parentheses under their corresponding regression coefficients. The number of observations, as well as the R-squared value, are reported for each regression.)

	Model 1	Model 2
Intercept	0.0005 (0.64)	0.0019 (2.71)
Market	1.0108 (58.29)	0.9805 (60.e42)
SMB	0.8817 (36.24)	0.8814 (39.41)
HML	0.2420 (9.25)	0.1895 (7.70)
UMD		–0.1570 (9.95)
N	540	540
R-squared	92.37%	93.55%

Table 4: **Regression Analysis of Returns on Long-Short Portfolios Based on Fluency**
This table reports the results of two different ordinary-least-squares regression analyses described by equation 3 in the text:

$$Return = \alpha + \beta_{Sentiment} \cdot Sentiment + \beta_{Market} \cdot Market + \beta_{SMB} \cdot SMB + \beta_{HML} \cdot HML + \beta_{UMD} \cdot UMD,$$

The monthly dependent variable is the return on a zero-cost portfolio that is long in the quintile of stocks with the most-fluent tickers and short in the quintile of stocks with the least-fluent tickers. The key explanatory variable, *Sentiment*, is the Investor Sentiment index, described in section 3.3. The other explanatory variables (*Market*, *SMB*, *HML*, and *UMD*) are the three Fama-French factors and the momentum premium, as described in section 3.4. The specification referred to as Model 1 below employs the three Fama-French factors; Model 2 also employs the momentum factor. (T-statistics are reported in parentheses under their corresponding regression coefficients. The number of observations, as well as the R-squared value, are reported for each regression.)

	Model 1	Model 2
Intercept	–0.0045 (42.91)	–0.0045 (41.80)
Market	0.0022 (0.91)	0.0016 (0.66)
SMB	0.0035 (1.02)	0.0035 (1.01)
HML	–0.0015 (0.41)	–0.0025 (0.67)
UMD		–0.0029 (1.21)
Sentiment	–0.0005 (4.37)	–0.0005 (4.35)
N	540	540
R-squared	3.65%	3.74%