ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Master Thesis Financial Economics

The Fluency Effect: Evidence on the Existence, the Persistence and the Profitability

Abstract

Recent papers in behavioral finance show the existence of the fluency effect. The establishment and explanation of this effect was their priority. This paper expands the literature about the fluency effect by investigating the development of the fluency effect over time and the profitability of trading on this effect. I find that, using the companies of the S&P 500 between 1970 and 2017, the fluent companies persistently outperform the market in the total sample period and all sub-sample periods. Although the fluency effect slightly diminishes over time, the last years it remained relatively stable with an outperformance of 0.4% per month. The zero-cost trading strategy of buying fluent and selling disfluent stocks yields no significant positive returns.

Keywords: Fluency, Stock Returns, Efficient Market Hypothesis, Zero-cost Trading Strategy

Name Student: Thijs Jan Trouwborst
Student Number: 411038

Supervisor: Dr. M. Montone
Second Assessor: Dr. J.J.G. Lemmen

Date Final Version: 20-06-2018

Table of contents

1. Introduction

The investment decision is a complicated process for investors. Investors are influenced by a lot of factors, for example the economic cycle, the investment budget and the investment opportunities. These factors are accompanied by a wide range of potential investment biases, such as the home bias, overconfidence, the disposition effect and many more. In addition, recent studies have shown the potential impact of the name of a company name on the investment decision. The effect of the company name on this investment decision can be captured in the word fluency, that relates to processing stimuli and memory recall. Fluency is the ease with which people process information and remember a word, name or code (Green & Jame, 2013).

Studying the influence of fluency on stock performance contributes to the growing impact of behavioral finance in the existing finance literature, which states that some financial phenomena can be better understood using models in which agents are not fully rational, according to Barberis and Thaler (2003). The existence of the fluency effect has already been established by several other papers. Alter and Oppenheimer (2006) investigate the difference in performance between fluent named stocks and disfluent named stocks. Green and Jame (2013) investigate the effects of company name fluency on breadth of ownership, liquidity and firm value, finding that companies with fluent have higher breadth of ownership, greater share turnover and a larger firm value. These results are contradictions to the traditional efficient market hypothesis (Fama, 1998). The efficient market hypothesis states that the prices on the stock market incorporate all publicly available information, so no one can outperform the market just with public information. Since company names and ticker codes are publicly available, these factors cannot lead to an outperformance of the market according to traditional efficient market theories. In this paper, I will answer the following research question:

Do fluent stocks persistently outperform the market and does investing based on this fluency pays off?

I will answer this question with the companies from the S&P 500 index between 1970 and 2017. Not only the existence of the fluency effect will be investigated, but this research will fill a gap in the literature by also examining the persistence and the profitability of the fluency effect over time. These three different investigations will also form the structure of my methodology.

To analyze the existence of the fluency effect I compare the average returns of portfolios based on fluency and calculated the outperformance of the fluent stocks with a 4-factor regression model between 1970 and 2017. I find a significant outperformance of the fluent stocks of 0.39% a month. When analyzing the persistence of this outperformance, I find that in all different sample periods the outperformance remains positive and significant. Finally, when examining the profitability of a zero-cost trading strategy, I find no significant risk-adjusted returns for investing based on fluency. Based on these results, the first part of the research question can be confirmed, because the fluent stocks persistently outperform the market. Only, I find no evidence on the profitability of the fluency effect. This research only tests one investment strategy, so there might be other strategies that are profitable. Overall, this research delivers new important insights on the development and the profitability of the fluency effect, because if the fluency effect persists over time, there might profitable investment strategies for investors and reasons for managers to choose their company name wisely.

The remainder of the paper will have the following structure: Section 2 is the theoretical framework and will discuss the most important related papers. The data of this research will be discussed in section 3. In section 4, I will explain the used methods. Section 5 presents the empirical results and in section 6 and 7, I will conclude and discuss possible shortcomings and recommendations for further research.

2. Theoretical Framework

In this theoretical framework, I will discuss the most important existing papers about the effect of fluency. I will elaborate on their concepts, theories and most important results. I will discuss the scientific and economic relevance of my research, based on the existing literature. Some parts of my paper will amplify the existing literature and other parts will fill in a gap in the literature about fluency. I will discuss the relevant literature chronologically.

Because fluency is a relatively recent topic in academic research and behavioral finance, the first papers on the effects of fluency on stock returns arose in the 21^{st} century. Alter and Oppenheimer (2006) were the first to analyze the effect of fluency on firm performance. Before this paper, fluency was only used as a purely psychological concept, but Alter and Oppenheimer (2006) were the first to link this psychological fluency concept with finance. Their description of fluency is: “That people tend to prefer easily processed information”. They used the stocks from the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX). They performed three different studies where the panel estimations of firm performance, based on fluency, were compared to their actual performance. In the first study, a laboratory study, the participants predicted significantly higher returns for fluent stocks. In the second study, with real-world stock market data, IPOs of fluent stocks significantly outperformed that of disfluent stocks after 1 day and 1 week, but not significantly in the longer term. In the third study, IPOs with pronounceable tickers significantly outperformed unpronounceable tickers after the first trading day, but not significantly in the longer term. With these three studies combined, Alter and Oppenheimer (2006), for the first time, show that fluency influences the stock performance.

The second relevant paper, that appeared three years later, is the paper from Head, Smith and Wilson (2009). They examined the relation between 'clever tickers' and their stock returns on companies from the CRSP database from 1984 till 2005. Clever tickers are ticker codes that are related to the firm's key business and are pronounceable. This clever ticker can give an indication of management's creativity and intelligence. In this way clever tickers can be connected to fluency, because tickers with a meaning and are pronounceable are easier to process for investors. They used a survey to determine what is a clever ticker and what is not. When they compared the clever ticker portfolio with the market portfolio they found an annual outperformance of 11.5% of the clever ticker portfolio. They also tested whether this outperformance can be explained by the systematic risk factors of Fama and French (1993) augmented by a momentum factor (Carhart, 1997). They found that this outperformance cannot be explained by these factors because of a positive and significant alpha that yields an annual excess return of 12.3%. Head, Smith and Wilson (2009) concluded that this outperformance could be explained by the fact that a clever ticker is an indicator of the managers' ability or by the fact that the clever ticker is memorable and has an influence on investor buying. So, this paper is a second proof that fluency, now in ticker codes, could influence stock performance, even when taking into account the systematic risk factors.

A more recent paper about the effect of fluency is the paper from Green and Jame (2013), which I will also use as my reference paper. They extended the fluency effect to other firm characteristics, like breadth of ownership, liquidity and firm value. They find that companies with fluent names have higher levels of breadth of ownership, greater share turnovers, smaller transaction price impacts, trade at significant premiums relative to less fluent companies and have higher firm value. To deal with the concern that fluency proxies for an omitted time-invariant firm characteristic, they investigated the effects of name changes within firms. This resulted in even stronger relations of the characteristics, which mitigates this concern. To link this paper more to my intended research on the effect on stock prices, Green and Jame (2013) investigated the effect of fluency on market-to-book ratios. Here, they find that a one unit increase in their fluency measure, increased the market-to-book ratio by 2.53%. This means a $3.75 million increase in market value for their median firm size. So, this paper shows that fluency not only significantly affects stock performance, but also other firm characteristics.

Finally, the most recent paper about the effect of fluency on stock returns appeared in February 2018, written by van den Assem, Montone and Zwinkels (2018). They examined the possible reasons why fluent stocks outperform disfluent stocks. The two possible hypotheses they give are that fluency conveys information on the quality of the firm or that this fluency grabs the attention of unsophisticated investors. To analyze these hypotheses, they used the sample period and fluency scores from Green and Jame (2013). They find evidence for the former hypothesis, because fluent stocks have higher risk-adjusted returns, and these returns are highest in times when noise trader demand is high. So, this paper shows the positive effect of fluency on stock performance that lacked in the paper of Green and Jame (2013) and proved the origin of this outperformance.

The contribution of my paper consists of some interfaces with existing papers and some extensions in this field of research. Because this topic is only investigated for approximately 10 years, it is important to find more evidence for this fluency effect. The prior researches had the priority to prove and explain the existence of the fluency effect over their total sample period. The establishment of this effect was in the most cases the priority of the authors, but since fluency has gained more attention the last years, it is also scientifically relevant to investigate the development of this effect over time. The second addition to existing literature will be the analysis to see if there is any mispricing where investors can make a profitable strategy on. So, I will not only prove if there is a fluency effect, but I will make things practical by investigating whether it is profitable to invest in a zero-cost strategy by buying the most fluent stocks and selling the least fluent stocks. If this zero-cost strategy yields a persistent positive return, it means that arbitrage cannot take this mispricing away. So, to analyze the extent of this potential profitability it is, again, important to investigate the development of this mispricing over time. So, by analyzing these new topics, I will clarify the size of the fluency effect over multiple decades and a whether trading on this effect yields significant returns over these periods.

3. Data

In this section I will elaborate on the data that I will use in this research. First, I will explain how the fluency scores are constructed. Then, I will explain the chosen sample of companies and the sample period. After that, I will discuss the variables that are needed to perform the methods I will use. Finally, I will discuss the data needed to perform robustness checks based on the fluency measure and the sample period from Green and Jame (2013).

3.1 Fluency Measure

To analyze the effect of fluency on stock performance, I have to construct a fluency measure that gives a reliable score based on the average investor. Because of the large number of companies, it is hard to create a fluency score for every company based on a survey. This means I have to construct an algorithm based on purely linguistic criteria. This linguistic algorithm I will use is comparable to that of Green and Jame (2013). First, I will manipulate the company names as follows: all abbreviations that CRSP has made are expanded to the original company name, I remove last expressions of the company names that most investors ignore (Co, Inc, Ltd etc.), articles, conjunctions (a, the, etc.) and the operating states in company names.

My measure is based on name length, number of syllables and Englishness. Based on length, every company will get a score based on the number of words in the company name. Company names containing one word get a score of 4, two words get a score of 3, three words get a score of 2 and more than three words get a score of 1. To score the names based on the number of syllables, company names with two or less syllables get a score of 4, three or four syllables get a score of 3, five or six syllables a score of 2 and seven or more than six syllables get a score of 1. Finally, Englishness will be checked by the Microsoft spell-check. If all words in the company name pass the check, it gets a score of 3. If one word does not pass the check, the company gets a score of 2. A company name with two or more words that do not pass the check gets a score of 1. Finally, all company names will end up with an aggregate fluency score with a maximum of 11 and a minimum of 3, which will be the fluency measure.

3.2 Sample and Time Span

The companies I will analyze in this research are the companies of the Standard and Poor's (S&P) 500 that were constituents between 1970 and 2017. These companies must have share code number 10 or number 11, that contain ordinary common shares. There are two reasons why I choose this sample. The first one is the English language. Because fluency is a relatively subjective measure, it is better to keep your sample based on one country with one native language. The native language of the United States is English, which is the reason why I can use Englishness as one of the fluency measures.

The second reason in the number of companies of this sample. The high number of companies creates a more reliable view on the effect of fluency. When considering only a small sample with the well-known companies, the investors will base their investment on familiarity and likeability rather than on fluency. So, it is important to also incorporate smaller and less known companies to prevent biased results.

I will analyze the effect from fluency on stock performance from 1970 till 2017. I choose this timespan, because fluency is relatively recent topic in behavioral finance. So, it is important to incorporate the most recent years in this research. I choose 1970 as my starting point, because the total number of months in this sample give an appropriate number of observations to create reliable results and to perform the different methods I will discuss later.

3.3 Other Variables

I will analyze the effect of fluency on the monthly stock returns of the companies. These monthly holding period returns are retrieved from CRSP database. I choose monthly returns, because these returns are less influenced by the bid-ask spread and are more stable over time. These returns are gathered for the period from January 1970 till December 2017.

To obtain the Fama and French factors I download the size, book-to-market and momentum factors from the website of Kenneth R. French. I will discuss these factors in more detail in the methodology section. I also need a size measure to analyze the fluency effect over various size quintiles. The size measure will be calculated by multiplying the numbers of shares outstanding by the share price, taking into account share splits and share dividends. This measure is called the market capitalization of the firm. Observations with negative stock prices will be removed, because these prices are based on the bid/ask average if there is no closing price available. The variables to create this market capitalization are also retrieved from the CRSP database.

Finally, I need a proxy for the market return and for the risk-free rate. My proxy for the market return is the equally-weighted CRSP return inclusive dividends. The return of the total CRSP database is a proper proxy, because now I can compare the returns of the fluent stocks in the S&P 500 with the returns of the biggest index in the United States. I will use the 30 days US treasury bill rate as my risk-free rate, because this is a short-term rate with less default risk. Both proxies are retrieved from the CRSP database. Both proxies will also be used to create the market factor. The market factor is the market return minus the risk-free rate and will be discussed in more detail in the methodology section. Eventually, my panel dataset will contain 1,483 companies and 402,365 monthly observations.

3.4 Robustness Check Data

I will also perform robustness checks to test the reliability of my results on the existence of the fluency effect. I will do this by using the fluency measure and sample period of Green and Jame (2013). In the first robustness check, I will use my own fluency measure on the sample period of Green and Jame (2013). They used the sample period between 1982 and 2009. In this way, I can check whether the results change when using a fluency measure on a different sample period.

For the second robust check, I will construct the fluency measure just like Green and Jame (2013) did. I will use their fluency measure on my companies and sample period to examine whether the change of the fluency measure in my sample period influences the results. Green and Jame (2013) construct their fluency scores slightly different than I do. Based on name length, names containing one word get a score of 3, two words get a score of 2 and more than two words get a score of 1. Based on pronounceability, they used the linguistic algorithm developed by Travers and Olivier (1978). This algorithm is based on the probability that a certain letter follows on the former two letters in a word. All company names will get an Englishness score based on this algorithm. All firms in the bottom quintile of Englishness will get a score of 0 and all other firms will get a score of 1. Finally, Englishness will also be scored on the Microsoft spell-check. Firm names with all words passing the spell-check will get a score of 1 and all others will get a score of 0. The final fluency measure is the aggregate score of these criteria with a maximum of 5 and a minimum of 1. So, I will also use this fluency measure on my sample period to check whether there are any differences in results. All results of these robustness checks can be found in Appendix B.

4. Methodology

I will use different methods to analyze the existence, the persistence and the profitability of the fluency effect. In all parts, I will use portfolio sorting and the three-factor model from Fama and French (1993) augmented with the momentum factor from Carhart (1997). First, I will explain the portfolio sorting method. Second, I will explain the model from Fama, French and Carhart I will use to calculate the abnormal returns.

Using these two methods, I will calculate the abnormal returns of all portfolios over the total timespan to check whether the fluency effect exists in my sample period (part 1). Then, I will use the rolling window regression method to analyze the persistence and size of the outperformance over time (part 2). Finally, I will investigate whether there is significant mispricing based on fluency, and whether trading on this mispricing pays off (part 3). I will also perform the robustness checks, mentioned in the data section, in part 1, to compare the results of the existence of the fluency effect with different fluency measures and different sample periods. I will mention the differences or similarities of these checks in the results section and put the tables of the robustness checks in appendix B.

4.1 Portfolio Sorting

Portfolio sorting is a methodological tool that helps dividing your sample into different groups based on a certain characteristic. In this way, I can perform the analyses on all portfolios and compare the results. The most important portfolio sorting in this research, is the sorting based on the fluency measure. I will divide my sample into five quintiles based on the fluency score. The firms with the highest fluency scores will end up in quintile 5, and firms with the lowest fluency scores will end up in quintile 1. These portfolios are being resorted every month, because every month there can be new firms that enter the sample. The next step is to create an average monthly return for every portfolio. In this way, I can compare the returns of the fluent stock portfolios with the returns of the disfluent stock portfolios.

I will also use this method to create the size quintiles based on the market capitalization. So, for every size quintile, the sample will again be divided into 5 fluency quintiles. This means the total sample of companies will contain double sorted portfolios, based on size and fluency. Now I can analyze whether the fluency effect is different for the various size quintiles.

4.2 Regression Model

To calculate the abnormal returns of all portfolios, I will use the three-factor model of Fama and French (1993), augmented by the momentum factor from Carhart (1997). So, the total regression function consists of four factors, a constant and an error term. The constant will capture the part of the returns that could not be explained by the factors. The factors of Fama and French are systematic risk factors that should explain all differences in returns. In combination with the short-term momentum effect of Carhart, the efficient market hypothesis (Fama, 1970) predicts that these four factors will capture all variation in returns, which will lead to an insignificant constant of approximately zero. The fluency effect predicts that firms with fluent names achieve significant abnormal returns, which means the constant is positive and significant. In this case, the factors in the model cannot explain the returns of the fluent stocks. The final regression function looks as follows:

$$R_{\mathrm{x,t}} = \alpha + \beta 1 * MKT_{\mathrm{t}} + \beta 2 * SMB_{\mathrm{t}} + \beta 3 * HML_{\mathrm{t}} + \beta 4 * UMD_{\mathrm{t}} + \varepsilon_{\mathrm{t}} \qquad (1)$$

Where $R_{\mathrm{x,t}}$ is the return of portfolio x minus the risk-free rate, MKT_{t} is the market return minus the risk-free rate, SMB_{t} is the size factor, HML_{t} is the book-to-market factor, UMD_{t} is the momentum factor and ε_{t} is the error term. SMB stands for small minus big and is created by subtracting the three portfolios with the highest market capitalization from the three portfolios with the lowest market capitalization. HML stands for high minus low and is created by subtracting the two portfolios with growth firms (low book-to-market) from the three portfolios with the value firms (high book-to-market). UMD stands for up minus down and is created by subtracting the two portfolios with the lowest (2-12 months) prior returns form the two portfolios with the highest (2-12 months) prior returns. The regressions will be tested on both heteroskedasticity (Breusch-Pagan/Cook-Weisberg test) and autocorrelation (Breusch-Godfrey LM test). I will use regressions that are robust for these problems if these tests confirm heteroskedasticity or autocorrelation. I will use this regression model to analyze the existence and the persistence of the fluency effect.

4.3 Existence of the Fluency Effect

To analyze whether the fluency effect exists in this sample, I will perform multiple tests. All tests will be based on analyzing the fluency effect over the total sample period. First, I will check whether the various quintiles have significantly different average returns than the market portfolio. Then, I will test whether the most fluent portfolio has a significantly higher average return than the most disfluent fluency portfolio. I will also compare the fluency effect over the various size quintiles, to analyze whether the fluency effect is different for different sized companies. I will test the significance of the differences in the average return by using a matched-paired t-test. This test looks as follows:

$$t = (\bar{X} - \mu) / \, (s \sqrt{n}) \tag{2}$$

Where $\bar{X}$ is the mean of the differences between two groups, μ will be equal to zero, s is the standard deviation of the differences and n is the number of observations.

Maybe the differences are due to an industry effect or a few fluent outliers. If this is the case, the effect of fluency could be zero or marginal. To check this, I will analyze the distribution of the SIC-codes of the fluent companies, and the number and percentage of fluent companies that could have caused the outperformance of the most fluent quintile, like Head, Smith and Wilson did (2009). After this analysis, I will get a first impression about the outperformance of the fluent firms.

Second, I will analyze whether the differences in the average returns could be explained by the four factors of Fama and French (1993) and Carhart (1997). It can be, that the outperformance of the fluent stocks can be explained by these systematic risk factors, and it has nothing to do with fluency. So, when comparing the alphas of the different portfolios, I can examine whether the fluent stocks also have higher abnormal returns. When this is the case, fluency could really have an effect on the stock performance.

After performing both tests, I will be able to discuss the first hypothesis: *Fluent stocks outperform disfluent stocks and the market, and this outperformance cannot be explained by the systematic risk factors of Fama, French and Carhart.*

4.4 Persistence of the Fluency Effect

To analyze the development of the fluency effect over time, I will split my sample into multiple time periods. To do this, I will use the rolling window regression method. Here, I will also use the 4-factor model of Fama, French (1993) and Carhart (1997). This method uses a moving timespan, by which I will use a movement of one year after every regression. Every single regression contains a sample period of 20 years. So, the first regression will capture the years 1970 till 1989. The second regression will capture the period from 1971 till 1990, and so on. In this way, I can analyze the development of the outperformance, alpha, over time. Because there are overlapping years in multiple periods, the standalone effect of a single year is spread out. When I use standalone sample periods, the effect of one year could influence the results of a total sample period. Now, when using the rolling window regression, the results are less influenced by outliers in certain years or months. So, with this method, I can analyze the development of the fluency effect gradually.

After performing this analysis, I will be able to discuss the second hypothesis: *There is a significant outperformance of the fluent stocks in all sub-periods.*

4.5 Profitability of the Fluency Effect

In this final analysis, I will be more practical by examining whether trading on fluency pays off and how much. By analyzing the average excess returns of the most fluent and the most disfluent portfolios, I will check whether there is persistent mispricing throughout the years. I will do this for the full sample period, and for subperiods of 5 years. I will test the mispricing by subtracting the risk-adjusted returns of most disfluent stock portfolios from the risk-adjusted returns of the most fluent stock portfolios. I will test the mispricing by analyzing the signs and the significance of these differences in excess returns. Again, I will use the matched-pair t-test to determine the significance of the mispricing. The risk-adjusted returns will be calculated, like Jegadeesh and Titman (1993) did for calculating the profitability of the momentum factor, with the following formula:

$$R_{x,t} - r_{ft} = \alpha + \beta^*(MKT_t - r_{ft}) + \varepsilon_t \qquad (3)$$

Where $R_{x,t}$ is the return on portfolio x, r_{ft} is the risk-free rate and MKT_t is the market return. The excess returns will be captured in the constant alpha.

After determining whether there is significant mispricing or not, I will investigate the profitability of a zero-cost trading strategy. I will analyze the profitability of a zero-cost trading strategy by buying the most fluent stocks and selling the most disfluent stocks.

After performing these analyses, I will be able to discuss the third hypothesis: *There is persistent mispricing, and trading on this mispricing leads to significant excess returns.*

5. Results

This result section will also be divided based on the different parts of my research. First, I will discuss the results of the existence of the fluency effect in the total sample period. Second, I will discuss the results of the persistence of this fluency effect over time. Finally, I will discuss the profitability of the fluency effect by implementing an investment strategy on fluency. In every part I will discuss the most important tables in the result section and I will put the other tables and figures in the appendix.

5.1 Existence of the Fluency Effect

The first test to analyze the existence of the fluency effect is to compare the average returns of the different fluency quintiles. These tests, using the matched-paired t-test, are performed for the full sample, the biggest size quintile and the smallest size quintile. In table 1 you can find the results of the full sample, where Q1 is the most disfluent quintile and Q5 is the most fluent quintile. The results of the biggest and smallest size quintiles can be found in table A1 and A2 in the appendix.

Table 1: Comparison of fluency quintiles with matched-paired t-test, full sample

	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Ret	0.0142	0.0146	0.0138	0.0142	0.0147	
St. Dev.	0.0512	0.0534	0.0530	0.0536	0.0534	
Market	0.0111	0.0111	0.0111	0.0111	0.0111	
Difference	0.0031	0.0035	0.0028	0.0031	0.0036	0.0005
t-statistic	2.84***	3.25***	2.68***	3.01***	3.52***	0.82
# of obs.	576	576	576	576	576	576

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01

The most important result to derive from this table is that the most fluent quintile has the highest average return of all quintiles. This means, the fluent stocks performed best on average over the sample period from 1970 till 2017. It also has the highest significance on outperforming the market return. Surprising is that, on average, all fluency quintiles of the S&P 500 outperformed the equally-weighted CRSP market return. There is no clear pattern of the average returns across the quintiles, because the most disfluent quintile did not perform worst over the total sample period. This fact can also be seen in the matched-paired t-test between the most fluent and the most disfluent quintile in the last column of table 1. The most fluent quintile has a higher average return, but this difference is not statistically significant.

When comparing these results with the results from the biggest and smallest firms in the sample (Table A1 and A2 in the appendix), the most fluent quintile also has the highest average return of all quintiles. The biggest firms, on average, have a higher average return than the smallest firms. For the biggest companies, the most fluent quintile significantly outperforms the market and the most disfluent quintile. For the smallest firms, although the most fluent quintile has the highest average return, this quintile does not significantly outperform the market and the most disfluent quintile.

To check whether the outperformance of the fluent stocks is not due to an industry effect or a few fluent outliers, I made a distribution of all SIC codes of the fluent companies and checked how many of these companies performed better than the average fluent stock. This list can be found in the appendix, table A3. There are 81 different major industry groups, the two-digit SIC codes. The distribution of the fluent companies in my sample span 62 of these major groups. The highest number of fluent companies are concentrated in group 35 (Industrial Machinery & Equipment). There are 38 companies (8.5%) in this group, by which 22 of the 38 companies in this group performed worse than the average fluent stock. The second highest concentration in an industry is 33 companies (7.4%) in group 28 (Chemical & Allied Products), where 16 of the 33 companies performed worse than the average fluent stock. The other industries from the distribution of table A3 and potential outliers do also not lead to notable concerns about the results.

The second test to examine the existence of the fluency effect is to check whether the outperformance of the fluent stocks can be explained by the risk factors from Fama and French (1993) and the momentum factor from Carhart (1997). The output of this regression can be found in table 2. I test these regressions for heteroskedasticity and autocorrelation. Because of the tested heteroskedasticity, I use robust standard errors. The regression output without controlling for heteroskedasticity can be found in table A4.

When looking at table 2, the alphas of all fluency quintiles are positive and highly significant. This means that the factors in the regression model cannot fully explain the returns of the quintiles. When looking more closely at the different alphas, the alpha of the most fluent quintile has the highest size (0.0039) and significance ($t = 4.03$). This means the most fluent quintile yields an excess return of approximately 0.39% a month and 4.7% a year. When comparing this with the most disfluent quintile, this quintile yields an excess return of approximately 0.34% a month and 4.1% a year.

Table 2: 4-factor regressions on full sample period of all fluency quintiles

	(1) Ret_Q1	(2) Ret_Q2	(3) Ret_Q3	(4) Ret_Q4	(5) Ret_Q5
MKT	1.0065***	1.0284***	1.0338***	1.0401***	1.0359***
	(43.42)	(44.14)	(47.95)	(48.84)	(44.64)
SMB	-0.5393***	-0.5003***	-0.5142***	-0.5124***	-0.4897***
	(-14.08)	(-12.55)	(-13.68)	(-14.40)	(-12.08)
HML	0.0940*	0.0534	0.0197	-0.0190	-0.0167
	(1.94)	(1.09)	(0.44)	(-0.43)	(-0.38)
UMD	0.0129	0.0097	0.0059	0.0103	0.0366
	(0.30)	(0.22)	(0.16)	(0.24)	(0.92)
Constant	0.0034***	0.0038***	0.0032***	0.0036***	0.0039***
	(3.65)	(3.72)	(3.42)	(3.66)	(4.03)
Observations	576	576	576	576	576
R^2	0.85	0.84	0.86	0.86	0.86
Adjusted R^2	0.85	0.84	0.86	0.86	0.86

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01

When analyzing the other factors in the model, the market factor coefficient is around 1 and significant. The size factor is negative and significant, which is not in line with the theory. Maybe this factor is disturbed because of the relatively high sized companies in the S&P 500. Finally, the book-to-market and momentum are not significant.

The results of the robustness checks for the existence of the fluency effect, stored in appendix B, confirm the results discussed above. When testing the existence of the fluency effect with the sample period of Green and Jame (2013) and my fluency measure, table B1, the results are comparable. Also, the robustness check with using the fluency measure of Green and Jame (2013) on my sample period, table B2, the results are approximately the same. At these robustness checks the most fluent quintile also outperforms the market significantly and outperforms the most disfluent quintile insignificantly. For the results of the 4-factor model, the robustness checks also give approximately the same results. The most fluent quintile has the highest alpha that is highly significant, for a different sample period, table B3, and a different fluency measure, table B4.

Overall, the size and the significance of the alphas are hard to reconcile with the efficient market hypothesis, because the market, size, book-to-market and momentum factor cannot explain the results of all quintiles. Especially not that of the most fluent quintile, which yields the highest and most significant monthly excess return. With the results of the match-paired t-tests and the regression analyses, I cannot reject the first hypothesis: *Fluent stocks outperform disfluent stocks and the market, and this outperformance cannot be explained by the systematic risk factors of Fama, French and Carhart.*

5.2 Persistence of the Fluency Effect

To investigate the development of the outperformance of the fluent stocks, I use the rolling window regression method. Every sample period contains 20 years, which means 240 months of observations. For every sample period, shifting 1 year every regression, I calculate the alpha for every fluency quintile. These alphas are the excess returns that cannot be explained by the market, size, book-to-market and momentum factor. The development of these excess returns for the most fluent quintile can be seen in figure 1 below. The development of the alphas from the other quintiles can be found in figure A1 in the appendix.

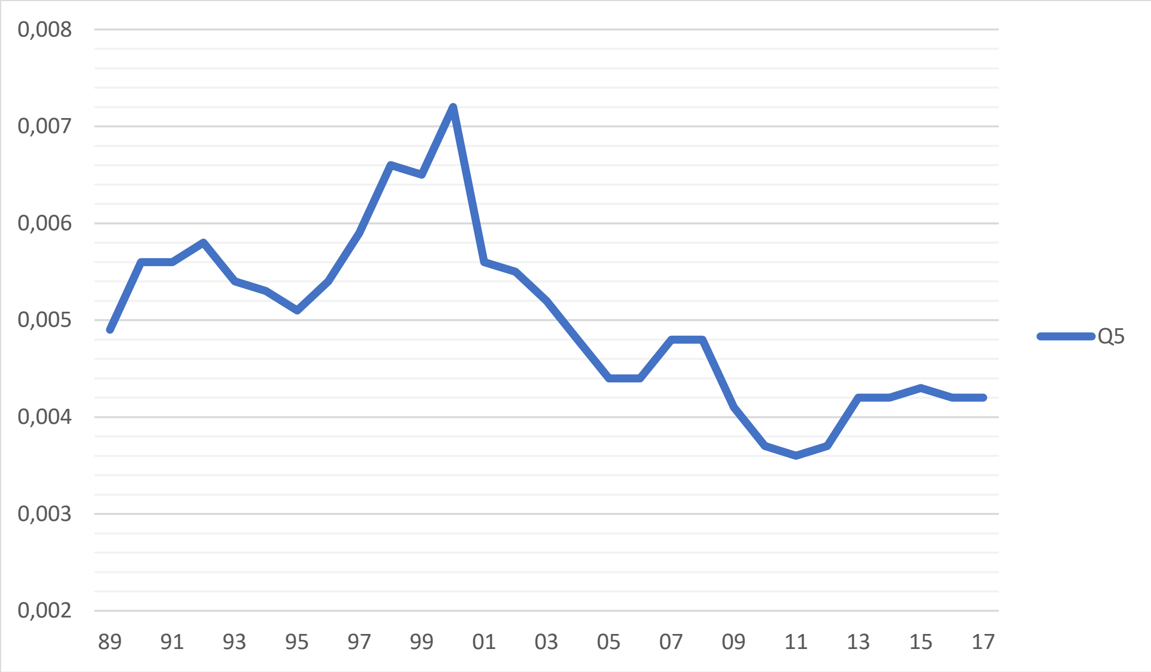

Figure 1. **Development of the Excess Returns (alphas) of the Most Fluent Quintile**

To interpret the figure, the years on the horizontal axis are the last years of every sample period of 20 years. So, 89 on the left side of the horizontal axis means the alpha of the sample period from 1970 till 1989. On the right side of the horizontal axis, 17, means the alpha of the sample period from 1998 till 2017. All alphas of the different sample periods are positive for the most fluent quintile, with a minimum alpha of 0.0036 (0.36%) in 2011 and a maximum alpha of 0.0072 (0.72%) in 2000. Also, all alphas are significant on a 1% significance level. This means, in all sample periods, the excess returns of the most fluent quintile cannot be explained by the four factors.

When looking at the development of the size of the alphas, the size diminishes slightly over the total sample period. There was a peak around end year 2000, and after this peak the alpha reduces to around 0.0045 (0.45%) in end year 2005. The final 12 years the alpha remained relatively stable around 0.004 (0.4%). The same pattern of the alphas can be seen for the other quintiles (figure A1). The most fluent quintile has on average the highest alpha, but the highest peak can be seen for quintile number 2, with a maximum of 0.008 (0.8%). Because of the stable alphas in most recent sample periods, the fluency effect is not likely to fade away in the coming years.

With these results I can discuss the second hypothesis: *There is a significant outperformance of the fluent stocks in all sub-periods.* Because all alphas of the fluent quintile are positive and significant on a 1% significance level, I cannot reject this hypothesis.

5.3 Profitability of the Fluency Effect

To investigate the profitability of trading on fluency, I calculated the risk-adjusted returns of a zero-cost trading strategy. The risk-adjusted returns mean that I calculate the excess returns, considering the risk-free rate and the market return. The zero-cost trading strategy means that the investor buys the portfolio with the stocks in the most fluent quintile and sells the portfolio with the stocks in the most disfluent quintile. I calculated these risk-adjusted returns for the full sample period and for subperiods of 5 years (8 years for the last sample period). The returns can be seen in table 3 and are the monthly averages over the stated sample period. I also calculated the non-risk-adjusted returns, this means just the raw returns, of the zero-cost trading strategy. These returns can be found in table A5 in the appendix.

Table 3: Risk-adjusted returns of the zero-cost trading strategy

	Q5	Q1	Q5 - Q1
70 - 17	0.0046***	0.0045***	0.0001
70 -74	0.0048**	0.0049*	-0.0002
75 -79	-0.0035	-0.0039	0.0004
80 - 84	0.0046	0.0055*	-0.0009
85 - 89	0.0101***	0.0097***	0.0004
90 - 94	0.0041	0.0028	0.0014
95 - 99	0.0086***	0.0084**	0.0003
00 - 04	0.0044	0.0061	-0.0018
05 - 09	0.0038	0.0043*	-0.0005
10 - 17	0.0052***	0.0036	0.0016

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01

When looking at the results in table 3, over the total sample period the average monthly risk-adjusted return is an insignificant 0.0001. Over the total period of 48 years, this means a total risk-adjusted return of 5.96%. The results of the various 5-year sub-periods are also not that convincing. Only 5 out of 9 sub-periods deliver a positive risk-adjusted return and none of the monthly returns are significant. In the best case, over the period between 2010 and 2017, the average monthly return is 0.0016. This means a total return of 15.53% over those 8 years. In the worst case, over the period between 2000 and 2004, the average monthly return of the zero-cost trading strategy is -0.0018. Over 5 years, this means a return of -10.63%. So, the zero-cost trading strategy of buying the most fluent stocks and selling the most disfluent stocks, will not always lead to positive risk-adjusted returns. This is also the case when looking at the raw returns in table A5, without considering the market return and the risk-free rate.

Especially, when taking into account there are transaction costs in the real world, the profits are even lower. So, although the average monthly return of this strategy is positive over a period of 48 years, the profits are not too high, and the possible negative returns make it a risky investment strategy. Overall, I can reject the third hypothesis: *There is persistent mispricing, and trading on this mispricing leads to significant excess returns.*

6. Conclusion

In this section I will revisit the most important results and discuss my research question. My research consists of three parts, by which I examine the existence, the persistence and the profitability of the fluency effect. By analyzing these characteristics of the fluency effect and their hypotheses, I am able to discuss the research question: *Do fluent stocks persistently outperform the market and does investing based on this fluency pays off?*

In the first part, the fluent stocks significantly outperform the market portfolio over the total sample period and this outperformance cannot be explained by the 4-factor regression model. The most fluent quintile has a higher excess return than the other quintiles and yields 4.7% per year. This is in line with the existing literature that examined the existence of the fluency effect. In the second part, this outperformance can be found in every sub-sample period when using the rolling window regression method. The excess returns of the fluent stocks diminished slightly over time, but the last decades the excess returns of the fluent stocks are relatively stable around 0.4% monthly. Regarding the research question, the results of these parts show that fluent stocks significantly outperform the market and that this outperformance persists in all periods between 1970 and 2017.

In the third part, the zero-cost investment strategy of buying the most fluent portfolio and selling the most disfluent portfolio do not systematically yield positive returns. Sometimes the profits are even negative, and all results are insignificant. So, although the fluent stocks yield the highest returns, the zero-cost trading strategy is not a riskless profitable strategy. Maybe other investment strategies, based on fluency, do pay off, but with these results I can not conclude that investing based on fluency pays off. These results on the persistence and the profitability of the fluency effect deliver new important insights in academic research, because the fluency effect seems to survive over time, but the zero-cost trading strategy does not give an opportunity to exploit this outperformance.

7. Discussion and Recommendations for Further Research

In this research there are a few points that leave room for discussion or that could have been done differently. The first one is about the fluency measure. Fluency is a relatively subjective measure, and it is difficult to perfectly capture the opinion of the average investor into a fluency measure. I, now, only use purely linguistic criteria to construct my fluency measure. Although the results were approximately the same with the Englishness measure of Travers and Olivier (1978), there may be some omitted factors that should have been incorporated in constructing the fluency measure. Optimal would be a clearly explained survey with a sufficient number of respondents and companies. Only, also surveys can have some disadvantages regarding the respondents and the survey itself. So, choosing a fluency measure leaves room for discussion and maybe my fluency can be improved in further research with other than only linguistic factors.

A second discussion is the chosen company sample, the S&P 500 companies that were constituent between 1970 and 2017. These companies approximately belong to the 500 largest companies in the United States. A risk, based on fluency, is that investor rather base their investment decision based on familiarity and likeability than on fluency. This familiarity and likeability can exist because of bigger media coverage, higher number of employees and other factors. Therefore, the fluency effect may be smaller for larger companies. This also leaves room for improvements in further research. A sample with large and small companies would improve the reliability of the results, and also comparing the fluency effect for large and small companies will deliver new insights.

Finally, it is important to repeat this research with more recent data to analyze the further development of the fluency effect. Because of the growing attention of the fluency effect, more investors will try to exploit this outperformance and the fluency effect could change. Also, extrapolating this research to other areas would be scientifically relevant. Maybe the fluency effect in English speaking countries is different than in other, non-English speaking countries, because English is a widely spoken language across the world. Also, comparing the results to results from countries with two main languages, like Canada and Belgium, can deliver new insights.

8. References

Alter, A., Oppenheimer, D., 2006. Predicting short-term stock fluctuations by using processing fluency. *PNAS* 103, 9369-9372.

Assem, M.J., Montone, M., Zwinkels, R.C.J., 2018. Company Name Fluency and Stock Returns. *Working paper*. http://dx.doi.org/10.2139/ssrn.3123201.

Barberis, N. and R. Thaler, 2003. A survey of behavioral finance. *Handbook of the Economics of Finance* 1, 1053-1128.

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

Fama, E.F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance, 25*, 383-417.

Fama, E.F., French, K., 1993. Common risk factors in the returns on bonds and stock. *Journal of Financial Economics,* 33, 3–53.

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

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

Jegadeesh, H., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance,* 48, 65-91.

Travers, J., Olivier, D., 1978. Pronounceability and statistical ‘Englishness’ as determinants of letter identification. *American Journal of Psychology* 91, 523-538.

9. Appendix A

Table A1: Comparison of fluency quintiles with matched-paired t-test, biggest firms

	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Ret	0.0131	0.0144	0.0170	0.0162	0.0204	
St. Dev.	0.0471	0.0473	0.0503	0.0486	0.0649	
Market	0.0111	0.0114	0.0107	0.0113	0.0118	
Difference	0.0020	0.0031	0.0062	0.0049	0.0086	0.0053
t-statistic	1.12	1.95*	3.07***	2.40**	3.11***	1.90**
# of obs.	576	574	511	546	325	325

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01

Table A2: Comparison of fluency quintiles with matched-paired t-test, smallest firms

	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Ret	0.0078	0.0058	-0.0034	0.0011	0.0086	
St. Dev.	0.0946	0.1180	0.1217	0.0929	0.0911	
Market	0.0109	0.0111	0.0110	0.0111	0.0111	
Difference	-0.0031	-0.0053	-0.0144	-0.0100	-0.0025	-0.0008
t-statistic	-1.04	-1.30	-3.33***	-3.76***	-0.87	-0.22
# of obs.	556	576	449	576	576	556

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01

Table A3: Industry distribution of the most companies

SIC code	Industry	Number of firms	%	better	worse
09	Fishing, Hunting & Trapping	1	0.2%	1	0
10	Metal, Mining	3	0.7%	1	2
12	Coal Mining	1	0.2%	0	1
13	Oil & Gas Extraction	11	2.5%	3	8
14	Nonmetallic Minerals, Except Fuels	1	0.2%	1	0
15	General Building Contractors	5	1.1%	3	2
16	Heavy Construction, Except Building	1	0.2%	0	1
17	Special Trade Contractors	1	0.2%	0	1
20	Food & Kindred Products	17	3.8%	6	11
21	Tobacco Products	3	0.7%	1	2
22	Textile Mill Products	4	0.9%	2	2
23	Apparel & Other Textile	7	1.6%	4	3
24	Lumber & Wood Products	6	1.3%	1	5
25	Furniture & Fixtures	3	0.7%	1	2
26	Paper & Allied Products	7	1.6%	1	6
27	Printing & Publishing	5	1.1%	1	4
28	Chemical & Allied Products	33	7.4%	17	16
29	Petroleum & Coal Products	9	2.0%	2	7

30	Rubber & Miscellaneous Plastics Products	4	0.9%	3	1
31	Leather & Leather Products	3	0.7%	0	3
32	Stone, Clay & Glass Products	8	1.8%	6	2
33	Primary Metal Industries	10	2.2%	2	8
34	Fabricated Metal Products	10	2.2%	5	5
35	Industrial Machinery & Equipment	38	8.5%	16	22
36	Electronic & Other Electric Equipment	27	6.0%	12	15
37	Transportation Equipment	20	4.5%	10	10
38	Instruments & Related Products	12	2.7%	8	4
39	Miscellaneous Manufacturing Industries	7	1.6%	1	6
40	Railroad Transportation	1	0.2%	1	0
41	Local & Interurban Passenger Transit	1	0.2%	0	1
42	Trucking & Warehousing	2	0.4%	0	2
44	Water Transportation	1	0.2%	1	0
45	Transportation by Air	3	0.7%	0	3
47	Transportation Services	1	0.2%	1	0
48	Communications	13	2.9%	5	8
49	Electric, Gas & Sanitary Services	11	2.5%	4	7
50	Wholesale Trade - Durable Goods	5	1.1%	3	2
51	Wholesale Trade - Nondurable Goods	10	2.2%	8	2
52	Building Materials & Gardening Supplies	1	0.2%	1	0
53	General Merchandise Stores	11	2.5%	4	7
54	Food Stores	5	1.1%	2	3
55	Automative Dealers & Service Stations	1	0.2%	1	0
56	Apparel & Accessory Stores	7	1.6%	6	1
57	Furniture & Homefurnishings Stores	2	0.4%	2	0
58	Eating & Drinking Places	6	1.3%	2	4
59	Miscellaneous Retail	8	1.8%	4	4
60	Depository Institutions	7	1.6%	1	6
61	Nondepository Institutions	2	0.4%	1	1
62	Security & Commodity Brokers	5	1.1%	3	2
63	Insurance Carriers	16	3.6%	8	8
64	Insurance Agents, Brokers & Service	2	0.4%	0	2
65	Real Estate	3	0.7%	1	2
67	Holding & Other Investment Offices	19	4.2%	16	3
70	Hotels & Other Lodging Places	2	0.4%	1	1
72	Personal Services	3	0.7%	1	2
73	Business Services	28	6.3%	22	6
78	Motion Pictures	1	0.2%	1	0
79	Amusement & Recreation Services	1	0.2%	0	1
80	Health Services	1	0.2%	1	0
87	Engineering & Management Services	6	1.3%	4	2
89	Services, Not Elsewhere Classified	1	0.2%	1	0
99	Non-Classifiable Establishments	5	1.1%	2	3
Total		448	100%	48.21%	51.79%

Table A4: 4-factor regressions full sample period on all fluency quintiles, without controlling for heteroskedasticity

	(1) Ret_Q1	(2) Ret_Q2	(3) Ret_Q3	(4) Ret_Q4	(5) Ret_Q5
MKT	1.0065***	1.0284***	1.0338***	1.0401***	1.0359***
	(48.97)	(47.07)	(51.35)	(50.69)	(49.86)
SMB	-0.5393***	-0.5003***	-0.5142***	-0.5124***	-0.4897***
	(-15.21)	(-13.27)	(-14.80)	(-14.47)	(-13.66)
HML	0.0940***	0.0534*	0.0197	-0.0190	-0.0167
	(3.11)	(1.66)	(0.66)	(-0.63)	(-0.55)
UMD	0.0129	0.0097	0.0059	0.0103	0.0366*
	(0.62)	(0.43)	(0.29)	(0.49)	(1.73)
Constant	0.0034***	0.0038***	0.0032***	0.0036***	0.0039***
	(3.92)	(4.06)	(3.67)	(4.05)	(4.40)
Observations	576	576	576	576	576
R^2	0.85	0.84	0.86	0.86	0.86
Adjusted R^2	0.85	0.84	0.86	0.86	0.86

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01

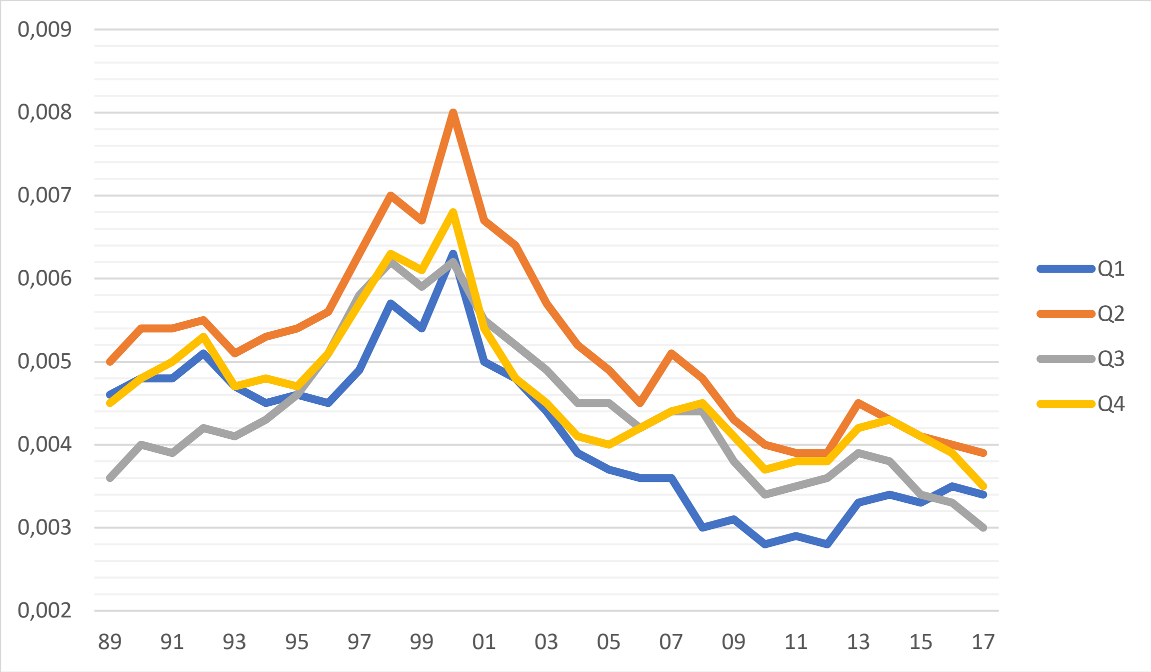

Figure A1. **Development of the Excess Returns (Alphas), Quintile 1 to 4**

Table A5: Raw returns of the zero-cost trading strategy

	Q5	Q1	Q5 - Q1
70 - 17	0.0147	0.0142	0.0005
70 -74	-0.0021	-0.0005	-0.0015
75 -79	0.0250	0.0211	0.0039*
80 - 84	0.0183	0.0183	-0.0001
85 - 89	0.0202	0.0191	0.0011
90 - 94	0.0136	0.0120	0.0016
95 - 99	0.0220	0.0211	0.0009
00 - 04	0.0138	0.0151	-0.0013
05 - 09	0.0077	0.0084	-0.0007
10 - 17	0.0142	0.0137	0.0005

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01

10. Appendix B: Robustness Checks

Table B1: Comparison of fluency quintiles with matched-paired t-test, 1982-2009 (Green & Jame, 2013)

	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Ret	0.0156	0.0159	0.0155	0.0147	0.0162	
St. Dev.	0.0513	0.0528	0.0535	0.0521	0.0554	
Market	0.0112	0.0112	0.0112	0.0113	0.0112	
Difference	0.0045	0.0048	0.0043	0.0034	0.0051	0.0006
t-statistic	2.82***	2.99***	2.88***	2.33**	3.02***	0.53
# of obs.	336	336	336	290	336	336

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01

Table B2: Comparison of fluency quintiles with matched-paired t-test, 1970-2017 (Green & Jame, 2013)

	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Ret	0.0132	0.0140	0.0140	0.0149	0.0146	
St. Dev.	0.0544	0.0517	0.0519	0.0545	0.0574	
Market	0.0111	0.0111	0.0111	0.0111	0.0111	
Difference	0.0021	0.0029	0.0029	0.0038	0.0036	0.0014
t-statistic	1.44***	2.61***	2.68***	3.96**	2.96***	1.09
# of obs.	576	576	576	576	576	576

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01

Table B3: 4-factor regressions on full sample period of all fluency quintiles, 1982-2009 (Green & Jame, 2013)

	(1) Ret_Q1	(2) Ret_Q2	(3) Ret_Q3	(4) Ret_Q4	(5) Ret_Q5
MKT	1.0100***	1.0020***	1.0240***	1.0223***	1.0069***
	(33.18)	(27.76)	(33.04)	(29.22)	(24.90)
SMB	-0.5688***	-0.5271***	-0.5250***	-0.5266***	-0.5041***
	(-11.44)	(-10.13)	(-11.01)	(-8.76)	(-8.13)
HML	0.1179*	0.0459	0.0100	-0.1119	-0.0856
	(1.83)	(0.66)	(0.17)	(-1.57)	(-1.30)
UMD	0.0509	0.0183	0.0159	0.0171	0.0232
	(0.86)	(0.30)	(0.30)	(0.43)	(0.36)
Constant	0.0041***	0.0049***	0.0045***	0.0037***	0.0057***
	(3.05)	(3.26)	(3.16)	(2.79)	(3.28)
Observations	336	336	336	290	336
R^2	0.82	0.80	0.83	0.85	0.77
Adjusted R^2	0.82	0.80	0.83	0.85	0.77

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01

Table B4: 4-factor regressions on full sample period of all fluency quintiles, 1970-2017 (Green & Jame, 2013)

	(1) Ret_Q1	(2) Ret_Q2	(3) Ret_Q3	(4) Ret_Q4	(5) Ret_Q5
MKT	0.9772***	1.0074***	1.0236***	1.0474***	1.0309***
	(29.00)	(40.06)	(46.88)	(50.12)	(31.93)
SMB	-0.5526***	-0.5122***	-0.5429***	-0.4624***	-0.4605***
	(-10.39)	(-12.57)	(-14.64)	(-13.14)	(-9.56)
HML	0.2088***	0.1284**	0.0698	-0.0606	-0.1762***
	(3.29)	(2.45)	(1.55)	(-1.42)	(-3.59)
UMD	-0.0642	0.0242	0.0103	0.0150	-0.0218
	(-1.39)	(0.49)	(0.25)	(0.42)	(-0.54)
Constant	0.0027**	0.0030***	0.0032***	0.0043***	0.0048***
	(2.09)	(2.89)	(3.42)	(4.69)	(4.09)
Observations	576	576	576	576	576
R^2	0.73	0.83	0.86	0.87	0.80
Adjusted R^2	0.72	0.83	0.86	0.87	0.80

t statistics in parentheses, * p < 0:10, ** p < 0:05, *** p < 0:01