Sentiment & Attention

Liam Connor Murray
(S100000)
Karl Olof Vincent Lindberg
(S129103)

May 2023

Master Thesis
MSc in Economics & Business Administration
Finance & Investments

Supervisor: *Mads Gjedsted Nielsen*

COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Date of Submission:
Number of pages:
Number of characters:

Contents

Abstract

ONE CLEAN CONCISE PARAGRAPH TO BE WRITTEN AT VERY END.

1 Introduction

1.1 Background

There exists a number of definitions for investor sentiment found in the behavioral finance literature, although the vast majority of them refer to investor sentiment at an aggregate, market-wide level and not at a firm-specific level. Authors Malcom Baker and Jeffrey Wurlger[12] offer up two definitions of market-wide sentiment in their frequently cited journal article titled "Investor Sentiment and the Cross-Section of Stock Returns" published in 2004. The first definition of investor sentiment refers to the propensity of an investor or group of investors to speculate, while the second definition refers to an investor or group of investors' optimism or pessimism towards stocks in general. Authors Seok et al[13] define investor sentiment as the resulting demand shocks generated by uninformed, noisy investors that lead to persistent mispricings in asset prices (Seok et al, 2019). In the theoretical study titled "Market liquidity as a sentiment indicator", authors Malcom Baker and Jeremy C. Stein[14] identifies "sentiment shocks" occurring as a result of overconfident, uninformed investors over-weighting the strength of their private signals (either positive buying signals or negative selling signals), and then (perhaps irrationally) trading on these signals accordingly.

Despite the existence of the ever-growing literature demonstrating that stock returns are not completely explained by their respective fundamentals, there exists no clear-cut, unanimously agreed upon method of measuring and/or quantifying investor sentiment at either a market-wide or firm-specific level. Authors Baker and Wurgler[12] state that "there are no definitive or uncontroversial measures" when referring to investor sentiment (Baker and Wurgler, 2006). In the realm of factor investing, there exists a well-known and often cited framework known as the Fama-French 3 factor model often used to explain future expected returns[55]. The Fama-French 3 factor model includes market risk; measured by the market return in excess of the risk free rate, size; measured by the difference in returns of smaller capitalization stocks and large capitalization stocks, and value; measuring the difference in returns of firms with high book-to-market ratios (referred to as value stocks) and firms with low book-to-market stocks (referred to as growth stocks). Authors Jones and Bandopadhyaya (2008) state that "non-economic factors such as investor sentiment are increasingly being recognized as explanatory variables for analyzing asset prices" and "as the literature grows, so too does the array of competing measures" (Bandopadhyaya & Jones, 2008). Fast forward to today at the present time of writing, there is still no well-documented, unanimously agreed upon framework to quantify or measure investor sentiment, at either a market-wide or firm-specific level.

1.2 Investor Sentiment Vs. Investor Attention

Investor attention and investor sentiment are two distinct concepts in finance that can provide valuable insights into the behavior, tendencies, and general proclivities of market participants. We define investor attention as the level of awareness that investors possess towards a specific stock, sector/industry or market as a whole. Investor attention can be measured and quantified through a variety of different metrics such as the volume of trades[14], the level of media coverage[17], the number of research reports or the aggregate count of analyst buy/sell recommendations for a specific stock, industry or market. Investor sentiment on the other hand refers to the general mood or emotion investors and market participants convey towards current market conditions[12]. Investor sentiment is indicative of the collective level of confidence (skepticism) and optimism (pessimism) that market participants currently feel towards market expectations at some arbitrary point in the future. We recognize the dimensionality of each concept as the most important distinguishing factor between sentiment and attention. Investor sentiment is three dimensional and can be quantified as relatively positive (bullish), relatively negative (bearish), or relatively neutral. Investor attention on the other hand is two dimensional and is measurable on a single axis from relatively high to relatively low.

In addition to the differences in dimensionality, investor attention must lead investor sentiment. For an investor to feel either bearish or bullish about a specific stock, sector/industry or market, an investor must first be aware and have allocated a portion of their attention accordingly prior to having the capacity of developing some degree of sentiment. Investor attention is limited and scarce in nature[18], and investors must determine how to allocate their attention across the tens of thousands of listed stocks worldwide[19]. As mentioned in the Journal of Finance article titled "In Search of Attention" by authors Da, Engelberg and Gao published in 2011, there does not exist a direct proxy of investor attention[17]. It is virtually impossible to measure and/or quantify what fraction of an individual's mental capacity is consumed by a specific stock at a specific point in time. Instead, a variety of indirect proxies are used to relay an understanding of the degree of attention investors allocate toward a specific stock, sector/industry or market.

Indirect proxies of investor attention include but are not limited to trading volume[14], market turnover rate[12], frequency of mention on social media platforms[26] (i.e. Twitter, Reddit etc.), frequency of mention in financial news articles and/or headlines[17] (i.e. Dow Jones, Wall Street Journal, Reuters etc.) and online search volume frequency[17] (i.e. Google Trends). It is important to note that the sole mention of a stock, sector/industry or market in prominent financial news media does not guarantee an allocation of investor attention, which is "especially true in the so-called information age where a wealth of information creates a poverty of attention" (Da et al, 2011). Trading volume and market turnover can fluctuate for a number of reasons besides investor attention, and "a news article in the Wall Street Journal does not guarantee attention unless investors actually read it" (Da et al, 2011). If all investors who stumble upon a financial news article concerning a stock enter a trade upon finishing reading, the aggregate count of published financial news articles in a specific period of time would be a strong quantifier of investor attention. Perhaps all investors prefer to search a stock's ticker online prior to making a trading decision. Investors may want to see recent price development or confirm if the tone of other financial news outlets is aligned with the general tone of the initial article. In this case, the aggregate count of a stock's ticker as an online search term may serve as a better proxy for investor attention[17].

1.3 Investor Sentiment Proxies

There exists today a number investor sentiment gauges at a market-wide level with the most frequently cited being the VIX. The VIX, or more formally the CBOE Volatility Index[20], is an index that measures the expected volatility of the S&P 500 index over the following 30 days. Financial news media frequently refers to the VIX as the "fear index" as it tends to rise during periods of heightened uncertainty and fall during periods of market stability and strengthening investor confidence. According to the CBOE's Vix Volatility Suite web page[20] "the VIX index is a calculation designed to produce a measure of constant 30-day expected volatility of the U.S. stock market, derived from real-time, mid-quote prices of S&P 500 Index (SPX) call and put options" (Chicago Board Options Exchange, 2023). In addition to the VIX, a very similar indicator known as the VXN provides a similar insight, instead using the Nasdaq-100 index as the underlying index. The Nasdaq-100 is heavily weighted towards technology, and growth oriented stocks, which tend to be more difficult to value, more difficult to arbitrage, and as a result, more susceptible to sentiment-based price movement[12]. For a technology focused investor, the VXN is a more appropriate, sector specific gauge of the 30-day implied volatility.

In addition to using the implied volatility of 30-day index options, investor sentiment can be imperfectly measured using a variety of other indicators such as the bullish percent index[21], high-low index[21] and put-call ratios[16]. The bullish percent index, often referred to as the BPI, is a market breadth indicator commonly used by market technicians to gauge the overall health of the market. The intuition behind the BPI is fairly straightforward, albeit rather rudimentary. The BPI is calculated by identifying the number of stocks forming point-and-figure and bullish reversal patterns, and simply dividing the number of stocks in a bullish trend by the total number of stocks in the index or sector of interest[21]. A BPI value ranging from 0-49% is indicative of negative market sentiment. A BPI value ranging from 51-100% is indicative of positive market sentiment. Lastly, a BPI value of exactly 50% is indicative of neutral market sentiment.

The high-low index (HLI) is another frequently used technical analysis indicator that intends to provide a gauge of investor sentiment[21]. The HLI records the difference between the highest high and lowest low of a stock or market index over a specified look back period, ultimately measuring the strength or weakness of a trend or trend reversal[21]. Calculating the high-low index involves subtracting the lowest low from the highest high, and dividing the difference by the sum of the lowest low and the highest high to generate a value ranging from -100 to 100. Positive values indicate an uptrend in the underlying, while negative values indicate a downtrend in the underlying. Although indicators such as the BPI or HLI intend to gauge investor sentiment through using stock price data, these indicators provide an indirect proxy for investor sentiment. Indicators using price data measure the second order effect of investor sentiment on a stock's price as opposed to providing a direct measurement of investor sentiment.

A more direct proxy for investor sentiment is the put-call ratio, calculated by dividing the total trading volume in put options by the total trading volume in call options for a specific stock[16]. The put-call ratio is often cited by practitioners when referring to general market sentiment and is frequently used as a contrarian indicator[16]. Authors Bandopadhyaya and Jones conclude that in the period of January 2004 through April 2006, the put-call ratio is a statistically significant determinant of the S&P 500 index's variation not explained by the previous day's level. An American put option gives an investor the right, but not the obligation, to sell an underlying asset at a specific price (the strike price) within a specified time period[22]. An American call option gives an investor the right, but not the obligation, to buy an underlying asset at a specific price (the strike price) within a specified time period[22]. A put-call ratio equal to 1 indicates that investors purchased an equal number a stock's put and call options in specified period. A ratio greater (less) than 1 indicates investors purchased more put (call) options, suggesting a relatively bearish (bullish) sentiment within the options market for a specific stock. A put-call ratio provides a more direct, first order measurement of investor sentiment since price data is not used in its calculation and may reflect aggregate investor sentiment in a more timely manner than price-based sentiment indicators such as the BPI or HLI outlined previously.

1.4 US Market Environment

In the second half of 2009, shortly after the tail end of the global financial crisis of 2008, most major stock market indices worldwide began recovering their losses. Many governments and central banks around the world took significant measures to stabilize the global economy and stimulate economic growth. The United States emerged from the global financial crisis stronger than ever as "the U.S. stock market rose by 250 percent in the 2010s, nearly four times the average gain in other national stock markets" (Ruchir Sharma, 2020). For over an entire decade, U.S. equity markets enjoyed a utopian-like bull market primarily due to the low-interest rate environment. From December of 2008 until November 2015, the United States Federal Reserve's effective rate was held close to 0%[24].

In addition to a prolonged low interest rate environment, quantitative easing (QE) measures by the United States Federal Reserve also contributed significantly to the steady, unwavering economic growth of the 2010s. QE is one of the many tools used by a central bank to conduct monetary policy and stimulate a slowing economy. QE refers to central banks purchasing large amounts of financial assets such as government bonds from commercial banks and other financial institutions, effectively increasing the amount of money in circulation and providing substantial amounts of liquidity to financial markets[28]. Additionally, increased amount of money in circulation has a downward effect on interest rates rendering it easier for individuals and businesses to borrow money to reinvest in the economy. As a result of the favorable economic conditions in the 2010s, the S&P 500 (often used as the benchmark index for the U.S economy) averaged an annual return of 13.5%[25] (Washington Post, 2019).

1.5 FAANG+M

FAANG+M refers to a group of US technology stocks including Facebook (META), Apple (AAPL), Amazon (AMZN), Netflix (NFLX), Google (GOOGL) and Microsoft (MSFT)[49]. FAANG+M represents the 6 most well-known technology stocks in the US[47]. At the peak of the bull market rally of the 2010s, FAANG+M stocks accounted for over 23% of the market capitalization of the S&P 500 index[48]. Throughout the period of January 1, 2012 until April 21, 2023, the S&P 500 index returned 175.7% while the

FAANG+M index returned 591.6%[48]. The return of the S&P 500 index ex-FAANG+M over the same period would have been 137.3%[48]. FAANG+M stocks have consumed the minds of investors over the course of the last decade[49], making them ripe for a study exploring the relationship between investor sentiment, investor attention and stock returns.

1.6 Bloomberg Social Velocity Factors

In late 2014, Bloomberg added social sentiment analytics to their popular and widely used Bloomberg trading terminals[26]. Within the Bloomberg terminal, the social sentiment analytics data is referred to as the Bloomberg Social Velocity Factors. These factors will be referred to as the "BSV factors" throughout this paper. In the Bloomberg Professional Service Offering sentiment analysis white paper titled "Embedded Value In Bloomberg News & Social Sentiment Data", the author begins by echoing the findings of previous investor sentiment literature reiterating that "when rational arbitrageurs have limited risk-bearing capacity and time horizons, the actions of irrational noise traders can affect asset prices" and subsequently, "such actions can be interpreted as being driven by fluctuating investor sentiment" (Bloomberg, 2014). Through the use of supervised machine-learning techniques to process mass amounts of textual information, Bloomberg constructs both news and social sentiment values for a specified stock ticker. It is important to note that social sentiment values are limited to the social media platform Twitter.

Based on the methodology outlined in the white paper[26], a human expert initially assigns scores (1 for positive sentiment , 0 for neutral sentiment, -1 for negative sentiment) to a financial news article or tweet. The score labeling is explicitly based on the question "if an investor having a long position in the security mentioned were to read this news or tweet, is he/she bullish, bearish or neutral on his/her holdings?" (Bloomberg, 2014). The annotated scores are then fed into machine-learning models, whereby the model automatically assigns a confidence interval that a news article or tweet exhibits positive, neutral or negative sentiment concerning a specific stock.

The sentiment analytics data consists of six different count metrics including Twitter Publication Count (TC), Twitter Positive Count (TPC), Twitter Negative Count (TNC), News Publication Count (NC), News Positive Count (NPC) and News Negative Count (NNC). TC (NC) provides the total number of tweets (news articles) mentioning a stock in a specified time period. TPC and TNC (NPC and NNC) provide the total number of tweets (news articles) receiving positive and negative sentiment scores respectively in a specified time period. According to the white paper methodology, news sentiment data is “recomputed every two minutes with an eight hour rolling window” while Twitter sentiment data is “recomputed every minute with a 30-minute rolling window”[26]. In an official blog post published on March 14, 2011[27], Twitter revealed that the average number of tweets sent per day was 50 million (Twitter, 2011). Fast forwarding to the time of writing in 2023, the average number of tweets sent per day is likely far greater. It is assumed that the recomputing window for Twitter sentiment data is half that of news sentiment data since the volume of published tweets is far greater than the volume of financial news articles published in any arbitrary time period.

The Bloomberg white paper[26] also explores a number of trading strategies using sentiment analytics including a daily sentiment-strategy, a daily earnings event-driven strategy and an intraday sentiment-driven strategy, which are effectively confirming the efficient market hypothesis[11]. According to the back tested results, Bloomberg concludes that the “sentiment strategies outperform the corresponding benchmark index ETFs significantly, which strongly demonstrates the value embedded in Bloomberg News & Social Sentiment data” (Bloomberg, 2014). Bloomberg reports a variety of metrics for each strategy such as annualized return, annualized volatility, Sharpe ratio and average number of long and short positions.

The daily earnings event-driven strategy was back tested for S&P 500 stocks, Russell 3000 stocks and Russell 2000 stocks in order of descending average market capitalization throughout the period of January 2, 2015 to August 31, 2016. Based on the results of these backtests, Bloomberg concludes that the proposed trading strategies based on the Bloomberg News & Social Sentiment data is optimal for S&P 500 stocks because “S&P 500 companies attract more attention and analyst coverage, so their average sentiment from news and social sources just before earnings are reported is more likely to contain earnings-related information” (Bloomberg, 2014). In addition to these findings, Bloomberg also reports that the highest Sharpe ratios are achieved between 5 minute and 30 minute trading intervals, reporting that beyond 30 minute trading intervals, little to no meaningful returns are to be found[26], again reflecting the highly efficient nature of US financial markets[11].

1.7 Purpose of Study

The purpose of this study is to explore the predictive power of composite investor sentiment and investor attention proxies on FAANG+M stock returns. Building upon the previous literature outlining the construction of a composite market-wide investor sentiment index, this exploratory study attempts aggregate individual proxies of investor attention and investor sentiment into two firm-specific investor sentiment and attention indices denoted $SENT$ and $ATTN$ respectively.

1.7.1 Research Questions

The three primary questions this study will attempt to answer are the following:

1. Can individual proxies of investor sentiment be aggregated into a single, composite investor sentiment index for FAANG+M stocks?
2. Can individual proxies of investor attention be aggregated into a single, composite investor attention index for FAANG+M stocks?
3. Do these single, composite investor sentiment and investor attention proxies possess any predictive power for next period returns of FAANG+M stocks at the daily, weekly and monthly interval?

2 Literature Review

In the following section, the most relevant literature contributing to the inception and intuition behind our study's research questions is thorough reviewed. The literature review is split into two sections, the first concerning investor investor sentiment and second concerning investor attention.

2.1 Investor Sentiment

2.1.1 Investor Sentiment and the Cross-Section of Stock Returns

Published in August of 2006, the paper titled "Investor Sentiment and the Cross-Section of Stock Returns" written by authors Baker and Wurgler explores how investor sentiment affects the cross-section of stock returns[12]. The idea behind the paper stems from the assumption that investor sentiment related shocks are more observable in securities that are difficult to value and as a result, difficult to arbitrage[12]. A security whose price is difficult to value may be a newly listed firm with limited earnings history, a firm with an abstract or disruptive product that is difficult to value, a non-dividend paying firm or a firm whose valuation tends to be highly subjective. Authors Baker and Wurgler initially provide a theoretical prediction which states that investor sentiment shocks have cross-sectional effects, or more simply put do not impact all stock prices equally in a uniform fashion[12].

In order for authors Baker and Wurgler to empirically measure and quantify investor sentiment, a number of proxies are considered and used to create various time-series conditioning variables that are then aggregated into a composite sentiment index based first principal components[12]. To test the cross-section of stock returns and the composite sentiment index, using monthly stock returns in the period of 1963 to 2001, the authors form equal-weighted decile portfolios based on several firm characteristics such as market capitalization, age of firm, return volatility, profitability and dividend paying vs. non-dividend paying firms[12]. The authors report that "when sentiment is low, subsequent returns are higher on very young (newly listed) stocks than older stocks, high-return volatility than low-return volatility stocks, unprofitable stocks than profitable ones, and non-payers than dividend payers" (Baker and Wurgler, 2006), and that "when sentiment is high, these patterns completely reverse" (Baker and Wurgler, 2006). The intuition behind these findings is that during a bubble period where sentiment is irrationally high, the propensity to speculate on firms with highly subjective valuations is high, whereas an older firm with long earnings history is less subjective and as a result "less likely to be affected by fluctuations in the propensity to speculate" (Baker and Wurgler, 2006).

Authors Baker and Wurlger offer up the definition of investor sentiment as "optimism or pessimism about stocks in general", which is a very simplistic, concise and all-encompassing definition to say the least. In the third section of the paper titled "Empirical Approach and Data", the methodology is outlined on the basis of understanding that investor sentiment frequently leads to patterns of mispricing in securities, specifically securities that are difficult to value and as a result, difficult to arbitrage[12]. With regards to quantifying investor sentiment, the authors use six proxies while stating that "there are no definitive or uncontroversial measures". The six proxies include the "closed-end fund discount, NYSE share turnover, the number and average first-day returns on IPOs, the equity share in new issues, and the dividend premium". Using principal component analysis, the authors isolate the sentiment component from the idiosyncratic/non-sentiment component from each of the six proxies. It is important to highlight that the proxies selected intend to measure the aggregate, market-wide sentiment of investors at a specific point in time. No firm-specific measures of investor sentiment are incorporated in this study.

In the empirical results, authors Baker and Wurgler present findings from regressing an orthogonalized sentiment index on a number long-short portfolios sorted on a variety of factors[12]. The primary empirical finding is that "the cross-section of future stock returns is conditional on beginning-of-period proxies for sentiment" (Baker and Wurgler, 2006), revealing that when sentiment is high, stocks more attractive to speculative investors and tend to exhibit lower subsequent future returns. The authors conclude by suggesting that "descriptively accurate models of prices and expected returns need to incorporate a prominent role for investor sentiment" (Baker and Wurgler, 2006).

2.1.2 Measures Of Investor Sentiment: A Comparative Analysis Put-Call Ratio Vs. Volatility Index

Published in August 2008 in the Journal of Business & Economics Research, the journal article titled "Measure of Investor Sentiment: A Comparative Analysis Put-Call Ratio Vs. Volatility Index" by authors Jones and Bandopadhyaya explores the ability of two indicators' to capture investor sentiment. The authors motivation of the study is to explore investor sentiment as a possible determinant of asset prices and how "investor sentiment may explain short-term movements in asset prices better than any other set of fundamental factors" (Jones and Bandopadhyaya, 2008). The horizon of the study spans from January 2, 2004 through April 11, 2006 and uses data compiled from the Chicago Board Options Exchange[16].

The first indicator the authors explore is the put-call ratio of daily options volume on the S&P 500, citing previous research that suggests as the put-call ratio of the S&P 500 rises (falls), the market is likely to sell-off (rally)[16]. The second indicator explored is the VIX, which is often considered to be "the world's premier barometer of investor sentiment and market sentiment" (Jones and Bandopadhyaya, 2008). The authors expect the relationship between the VIX and the market to be the same as the relationship between the put-call ratio and the market[16]. A rising VIX suggests that there is a heightened sense of fear and uncertainty among market participants, and as a result the market will fall as investors look to unwind their positions and reduce exposure[16].

The authors first use a random walk-model to see "what portion of the variability in the daily movement of the S&P 500 index is explained by past values of the index itself" (Jones and Bandopadhyaya, 2008). Whistling the same tune as the efficient market hypothesis of Fama-French[11], the authors argue that past values of the S&P 500 are expected to reflect all relevant and available economic information that could affect the index "especially if the data are high frequency" (Jones and Bandopadhyaya, 2008). The unexplained portion (i.e. the residuals) from regressing the S&P 500 on its previous day's value is then "a result of other non-economic related factors such as changes in market sentiment" (Jones and Bandopadhyaya, 2008). The authors report regression results including the variables, coefficients, test statistics and p-values from regressing both the put-call ratio and the VIX on the residuals from the initial random walk model. The authors report high statistical significance for both the put-call ratio and the VIX, concluding that the put-call ratio contains' greater explanatory power and statistical significance in comparison to the VIX[16]. As an additional exercise, our study will test the validity of this paper's results at the firm-specific level for FAANG+M stocks.

2.1.3 Measuring Investor Sentiment in Equity Markets

In the early 2000's, authors contributing to the behavioral finance literature seem to imply and suggest that in some instances, swings in investor sentiment appear to better explain fluctuations in asset prices than short term fundamentals such as price-to-earnings and book-to-equity ratios. In the paper titled "Measuring Investor Sentiment in Equity Markets" published in February 2006, authors Jones and Bandopadhyaya develop an equity stock market sentiment index to explore "how this measure can be used in a stock market setting by studying the price movements of a group of firms which represent a stock market index" (Jones and Bandopadhyaya, 2006) .

To construct an equity market sentiment, authors Bandopadhyaya and Jones build upon the previous work of Persaud (1996) to develop a measure of the general market's attitude or appetite towards risk[29]. Using daily return data from July 2003 to July 2004, returns are calculated for all securities contained in the Massachusetts Bloomberg Index[30] which is a "price-weighted index designed to measure the performance of the Massachusetts Economy" (Bloomberg, n.d.). Through use of the average standard deviation of the past 5 daily returns for each security contained in the MBI,"the daily rate of return and the historic volatility are ranked, and the Spearman rank correlation coefficient between the rank of the daily returns for each firm and the rank of the historic volatility of the returns for each firm is computed, and the result is multiplied by 100" (Jones and Bandopadhyaya, 2006). The authors find that fluctuations in the constructed equity market sentiment are significantly correlated with news flow concerning securities contained within the MBI[16].

Following construction of the equity market sentiment index, the authors regress the one-period lagged returns of the MBI and the current period return in the equity market sentiment index against the current period return of the MBI. A relevant finding is that "while the lagged value of the return in MBI has an insignificant impact on the dependent variable MBI, the coefficient on the equity market sentiment index is highly significant" (Jones and Bandopadhyaya, 2006). This finding suggests that daily returns in the MBI on any given day are primarily driven by investors' appetite for risk, not the previous day's returns in the MBI. The authors conclude the article very clearly by stating that "researchers and practitioners should pay close attention to investor sentiment as a determinant of changes in financial markets" (Jones and Bandopadhyaya, 2006).

2.1.4 Investor Sentiment in the Stock Market

In the paper titled "Investor Sentiment in the Stock Market" published in the spring of 2007, authors Baker and Wurgler define investor sentiment as "a belief about future cash flows and investment risks that is not justified by the facts at hand" (Baker and Wurlger, 2007). Prior to the periods of prolonged positive investor sentiment in recent years such as the lead up of the dotcom bubble crash in the late 1990's, it was not evident that investor sentiment had a material impact on asset prices[32]. The literature has shifted in recent years from asking whether investor sentiment has any material effect on asset prices to determining how to measure investor sentiment using a systematic, quantitative approach. Our study intends to provide a foundational framework to quantifying firm-specific investor sentiment and investor attention.

The paper explores which stocks are most likely to be affected by swings in investor sentiment[32]. Authors Baker and Wurgler initially hypothesize that "stocks of low capitalization, younger, unprofitable, high-volatility, non-dividend paying, growth companies or stocks of firms in financial distress are likely to be disproportionately sensitive to broad waves of investor sentiment" (Baker and Wurgler, 2007). Stocks of this nature tend to be more difficult to value and as a result more difficult to arbitrage. As arbitrageurs are less likely to actively trade these stocks back to prices more representative of their fundamental values, their prices are more likely to wander off due to positive or negative investor sentiment. More simply put, the underlying relationship is stocks that are speculative by nature and more difficult to value will have higher relative valuations during periods of heightened investor sentiment[32].

Prior to outlining their chosen proxies of investor sentiment, authors Baker and Wurgler write that"investor sentiment is not straightforward to measure, but there is no fundamental reason why one cannot find imperfect proxies that remain useful over time", suggesting that the more practical approach is to combine several imperfect measures into one aggregate investor sentiment index. A variety of sensible investor sentiment proxies are reviewed including investor surveys, investor mood, retail investor trades, mutual fund flows, trading volume, dividend premium, closed-end fund discount, option implied volatility, first-day IPO returns, IPO volume, equity issues over total new issues and insider trading[32]. Due to data availability, a sentiment index is constructed using the same six proxies used in Baker and Wurgler (2006)[12] (trading volume as measured by NYSE turnover; the dividend premium; the closed-end fund discount; the number and first-day returns in IPOs; and the equity share in new issues). As a result of the deregulation of brokerage commissions and persistent decline in trading costs leading to an upward trend in turnover, the log of turnover minus a five-year moving average is used[32]. The authors are also careful to isolate any influence of economic fundamentals by regressing each proxy on a set of macroeconomic indicators[32]. The residuals of the regressions are then used as the sentiment proxies.

Using monthly mutual fund flows data from The Investment Company Institute, principal component analysis is used to "detect general patterns across several time series while ironing out distracting idiosyncratic fluctuations" (Baker and Wurgler, 2007). Mutual funds tend to reveal aggregate decisions of a large set of investors who are generally less sophisticated and more likely to exhibit sentiment-driven investment behavior. The resulting correlation of 0.36 between the speculative demand contained in the second principal component of the mutual fund flows and the sentiment index is highly significant implying that Baker and Wurgler sentiment index based on the six proxies outlined earlier "to a large extent captures a prevailing "greed" versus "fear" or "bullish" versus "bearish" notion"[32]. Additional significant findings include that when sentiment is high, subsequent market returns are low as well as that the impact of market sentiment is stronger for smaller stocks with more volatile monthly returns (i.e., more difficult to arbitrage)[32].

2.1.5 Investor Sentiment Measures

In the paper titled "Investor Sentiment Measures" published July 2006 in the Social Science Research Network, authors Qiu and Welch attempt to validate two widely used proxies to empirically quantify and measure general investor sentiment. The two measures are the closed end fund discount and consumer confidence, two indicators often used as proxies in the behavioral finance literature[33].

Referring to previous literature[34] by authors Lee, et al (1991), Qiu and Welch interpret the closed-end fund discount measure as having a negative correlation with investor sentiment[32]. A closed-end fund (hereafter referred to as CEF) is a type of investment fund that raises a fixed amount of capital through an initial public offering. Contrary to open-end funds that issue and redeem shares based on the net asset value of the underlying assets, CEFs issue a fixed number of shares that are then listed and traded on an exchange[33].

The share price of a CEF is determined by traditional supply and demand dynamics allowing it to trade at a discount (premium) to its net asset value. The CEF discount refers to the difference in percentage terms between the market price of a CEF share and its net asset value per share. The second proxy for investor sentiment is one of the components in the UBS/Gallup's Survey of Investor Sentiment, which is believed "to be the best available empirical direct proxy for investor sentiment" (Qiu and Welch, 2006). Using CEF discount as a proxy for investor sentiment strongly relies on the widely accepted empirical observation that CEFs are primarily held by less informed retail investors[33].

The paper offers three significant findings. The first being that the CEF discount is not a good measure of investor sentiment and has no correlation with the results of the UBS/Gallup measure of investor sentiment. Secondly, the authors show that the changes in consumer confidence (Michigan Consumer Confidence Index provided by Michigan Consumer Research Center) "correlate strongly with changes in the UBS/Gallup proxy" (Qiu and Welch, 2006). The consumer confidence proxy and UBS/Gallup sample different respondents (UBS/Gallup samples investors with wealth in excess of $100,000 while consumer confidence samples investors with wealth under $100,000), suggesting that the shared common factor is investor sentiment. Lastly, the authors explore the relationship between investor sentiment and markets. Qiu and Welch provide evidence outlining that "there is no difference in how poor and wealthy investor sentiment changes month-to month" (Qiu and Welch, 2006), indicating that wealth is not a determinant of investor sentiment changes. Our study intends to create a single, composite proxy for investor sentiment (attention) that captures the aggregate emotion or moods (level of awareness) of informed and uninformed investors.

The underlying intuition behind the empirical analysis of the paper is that since small decile stocks are "disproportionately held by noise traders, sentiment changes should change the spread between small decline firms and large decile firms" (Qiu and Welch, 2006). Additionally, the authors propose that changes in investor sentiment "disproportionately influence stocks not held by institutional but retail investors, especially if these stocks have insufficient liquidity to allow arbitrageurs to impose rational pricing" (Qiu and Welch, 2006). The results show that changes in the consumer confidence and CEF discount series have strong explanatory power in the spread between the smallest and largest capitalization stocks held within the frequently used CRSP portfolios[33].

Towards the end of the paper, the authors humorously caution readers to interpret the presented evidence with extra care as there exists "considerable academic sentiment when it comes to interpretations of investor sentiment" (Qiu and Welch, 2006). Until the academic literature is presented with justifiable theories outlining explicitly and quantitatively how much investor sentiment should influence prices in financial markets, "no paper can fully confirm or reject one of the two perspectives in favor of the other" (Qiu and Welch, 2006). The referred perspectives are the classical rational investor theory and the behavioral irrational investor theory.

2.1.6 Firm-specific investor sentiment and daily stock returns

Published in the North American Journal of Economics and Finance in November of 2018, the journal article titled "Firm-specific investor sentiment and daily stock returns" by authors Seok et al. explores the relationship between investor sentiment and daily stock returns in the Korean stock market. As a result of Korea's high degree of collectivism, authors Seok et al. hypothesize that the influence of market sentiment in Korea is stronger[13]. Additionally, there is a strong level of participation among individual investors “who are usually uninformed, noisy, and sensitive to market sentiment” (Seok et al., 2019). Given these characteristics of the Korean stock market, a study pertaining to market sentiment has the potential to generate revealing results.

The authors note that much of the existing literature analyzing market sentiment makes use of relatively low frequency data (i.e. monthly or annually), and often conclude that periods of high investor sentiment are often followed by lower subsequent stock returns[13][12]. Our study uses daily, weekly and monthly time intervals to explore the relationship between data frequency and effectiveness of quantifying firm-specific investor sentiment and attention.

The authors cleverly point out that “low frequency measures cannot effectively capture the mispricing process and cannot correctly estimate the time frame of the mispricing” and that through the use of analyzing short term frequency data (i.e. daily or weekly), one can determine whether mispricings due to swings in investor sentiment are immediately corrected or develop an understanding of how long the mispricing lasts[13].

Firms are sorted on a variety of characteristics such as firm size, stock volatility, profitability and growth opportunities in order to determine if firms that are difficult to arbitrage and value are more sensitive to swings in investor sentiment. To quantify firm specific investor sentiment, authors Seok et al. use a similar approach[12] taken by authors Baker and Wurgler (2006).

Using principal component analysis, the authors combine the relative strength index (RSI), psychological line index (PLI), adjusted turnover rate (ATR) and logarithm of trading volume (LTV) into a firm-specific sentiment index[13]. It’s important to note that LTV and ATV are both constructed using trading volume which begs the question if it is better to include one or the other instead of both. Additionally, both the RSI and PLI are constructed using price data. As mentioned in section 1.3, these are indirect, second order measurements of investor sentiment. Our study includes sentiment proxies that do not use stock price data to ensure more direct, first order measurements of firm-specific sentiment.

The authors intend to “identify instances when the trading volume is high for no rational reason” (Seok et al., 2019). Again, using a similar approach[12] to Baker and Wurgler (2006), the raw proxies are individually orthogonalized to three factors (firm size, book to market ratio, earnings-price ratio) to control for the effect of a firm’s fundamental factors on returns[13]. The authors note that “the residuals from these regressions are cleaner proxies for investor sentiment” (Seok et al., 2019).

The authors report that the sign of the coefficient on sentiment is significantly positive, suggesting that at the firm level, periods of high investor sentiment are subsequently followed by positive returns in the short term[13]. The authors also report findings that indicate the influence of investor sentiment is significantly stronger for smaller firms. The explanation offered is that arbitrageurs in the Korean stock market are not able to effectively offset the mispricings occurring as a result of swings in investor sentiment in smaller, more volatile, less profitable firms[13]. The authors conclude by reiterating the findings of Baker and Wurgler (2006), confirming that publicly traded Korean firms more difficult to arbitrage and as a result more difficult to value are more strongly affected by investor sentiment[13].

2.2 Investor Attention

2.2.1 In Search of Attention

Published in the Journal of Finance in October of 2011, the article titled "In Search of Attention" written by authors Da et al. appears to be the first paper that explores the relationship between stock prices and Google's search frequency data platform titled Google Trends (`https://trends.google.com/`. Google Trends is a web-based platform offered by Google allowing users to observe the popularity of a specified search term throughout a specified period of time. Users are able to extract the relative search frequency of searches or phrases varied by regions, languages and time periods[51]. Using search frequency data, authors Da et al. construct a search volume index (hereafter referred to as "SVI") to analyze the correlation between the proposed proxy of investor attention and Russell 3000 stock prices[17].

The paper begins by outlining that investor attention is a limited and scarce resource[18] and attributes the initial idea of the paper stemming from recent studies that "provide a theoretical framework in which limited attention can affect asset pricing statics as well as dynamics" (Da et al., 2011). Traditional indirect proxies of investor attention such as headline news contained in a day's issue of the Wall Street Journal do not guarantee attention unless investors actually allocate their attention accordingly and read it. The authors of the paper define abnormal SVI as "the log of SVI during the current week minus the log median SVI during the previous eight weeks" (Da et al., 2011). They find that the "majority of the time-series and cross-sectional variation in ASVI remains unexplained by alternative measures of attention" (Da et al., 2011) and that a given stock's SVI has little to no correlation with a news-based measure of investor sentiment[17].

By studying the changes in equity turnover between trading venues that typically attract less sophisticated investors and venues that attract more sophisticated investors, the authors suggest that SVI likely captures the attention of less sophisticated investors[17]. The paper also reviews investor sentiment referencing prior sentiment focused literature, but the authors echo the feelings of many academics stating "it is not clear how investor attention and sentiment should be related to each other" (Da et al., 2011). Our study intends to provide clarity on this matter by explicitly differentiating investor sentiment from investor attention and quantifying them separately.

Through the use of vector auto-regressions across four different constructed variables, the authors are able to show that "SVI captures investor attention in a more timely fashion than extreme returns or news" (Da et al., 2011), suggesting that Google Trends is a better and more direct measure of investor attention. When regressing existing proxies of investor attention such as turnover or news coverage, a resulting R-squared of approximately 3.3% suggests that conventional proxies of investor attention only explain a small portion of the variation in ASVI[17]. The authors also use SVI to test the validity of the price pressure hypothesis of Barber and Odean (2008), and "find that an increase in SVI for Russell 3000 stock predicts higher stock prices in the next 2 weeks and an eventual price reversal within the next year" (Da et al., 2011).

Another interesting and relevant finding is the negative coefficient between two variables[17]. The two variables are "log market cap" and "ASVI", which intend to measure the relationship between the magnitude of changes in ASVI relative to the market cap of a firm included in the Russell 3000. The negative coefficient between these two variables suggests a larger price increase following an increase in ASVI among smaller Russell 3000 stocks[17].

Google Trends data generally captures the attention of less sophisticated retail investors as more sophisticated investors have access to better platforms such as a Bloomberg Terminal. Alongside the evidence that retail investors tended to gravitate towards smaller stocks in the paper's sample period, "the positive price pressure is only present among the smaller half of (their) Russell 3000 stock sample" (Da et al., 2011). It's interesting to note that between 2007 and the 1st quarter of 2023 (the time of writing), fractional share offerings on many US based investment platforms catering to retail investors have allowed less sophisticated and less wealthy investors to participate in trading larger stocks, specifically the FAANG+M stocks outlined in section 1.5.

2.2.2 Market Liquidity As A Sentiment Indicator

In the journal article titled "Market liquidity as a sentiment indicator" published in 2004 by authors Malcom Baker and Jeremy C. Stein, a theoretical hypothesis attempts to explain and analyze the relationship between market liquidity and expected returns. The fundamental idea behind the literature is that "in a world with short-selling constraints, market liquidity can be a sentiment indicator" (Baker and Stein, 2004). The paper attempts to understand why time-series variation in market liquidity, at either the firm-specific or market might forecast changes in future expected returns[14]. It is important to note that at the time of writing, with the rise of online trading platforms, substantially lower transaction costs, and reduced margin requirements, short-selling has become significantly more accessible to retail investors. As a result, our study assumes that market liquidity better serves as a proxy of investor attention instead of investor sentiment.

Two assumptions are explicitly made, one concerning market liquidity and the other individual investor behavior[14]. The former assumes that investors are subject to and limited by short selling constraints. The latter proposes the "existence of a class of irrationally overconfident investors" (Baker and Stein, 2004) who overweight the value of their own private trading signals. Due to the market friction assumption that refers to short selling constraints, individual irrational investors will only be active in the market when the valuations inferred from their private signals are higher than the valuation of rational investors[14], "i.e. when their sentiment is positive and when the market is, as a result, overvalued" (Baker and Stein, 2004) and "when the sentiment of irrational investors is negative, the short-selling constraints keeps them out of the market altogether" (Baker and Stein, 2004). The authors suggest that measures of market liquidity (such as trading volume or market turnover rate) can provide an "indicator of the relative presence or absence of these investors, and hence the level of prices relative to fundamentals" (Baker and Stein, 2004).

The positive relationship between market liquidity and investor sentiment is outlined and conveyed through visual representations of empirical results performed on US value-weighted and equally weighted equity portfolios contained in the CRSP database. These graphs display the development of price action alongside the participation of 'smart' and 'retail' investors and market liquidity[14]. In region 1, smart investors dominate the market and 'w' (a variable representing the price impact of an individual trade) is increasing and liquidity and trading volume are low[14]. In region 2, both retail and smart investors are participating in the market and added additional provision of liquidity from retail investors reduces price impact[14]. In region 3, retail investors dominate the market with "smart traders on the sidelines, w is low and hence liquidity and trading volume are high" (Baker and Stein, 2004). To summarize, as retail investors increasingly participate in region 2 and 3 as a result of increasing optimism about the potential outcome of a trade, "not only do liquidity and trading volume increase, but expected returns fall" (Baker and Stein, 2004). Despite retail investors subject to the short-selling constraints assumption outlined earlier in the paper, the authors indicate that as retail investors become increasingly optimistic and are doing all the buying, "smart investors continue to exert the same marginal influence on price by taking short positions" (Baker and Stein, 2004).

2.2.3 All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors

TO BE WRITTEN

2.2.4 The Effect of Social Media on Trading Behavior: Evidence From Twitter

Published in 2015, the article "The Effect of Social Media on Trading Behavior: Evidence from Twitter" written by Tanya Paul delves deep into the topic of Twitter's effect on trading patterns. The author hypothesizes that the use of twitter can cause asset mispricing as it allows companies to communicate in few characters, pushing out information to their investors. The author describes an environment where companies to a higher extent can communicate through their social media channels such as LinkedIN, Facebook and Twitter refers to a paper, Blankespoor, Miller, and White (2014) that concluded that increased twitter activity increased liquidity, which is a good indicator for an efficient market. . The author argues for choosing twitter as a study object due to the power companies have over what content that can be published, the amount of information that can be published and how far those publications can reach. The author presents results with comparisons between samples of data where companies did not have a twitter account and samples of data after they acquired an account. The author chooses two sets of data, monthly car sales announcements and product calls to explore the effects Twitter can have on abnormal returns and abnormal trading volume. The author presents results showing that tweets has an insignificant effect on abnormal returns for both sets of data but a significant effect on log abnormal trading volume for both sales announcements and product calls. The results show a negative relationship, with a coefficient of -1 implicating that an increase in tweets leads to a decrease in log abnormal trading volume. The results also show that twitter has a significant effect on abnormal returns but only for monthly sales announcements. The author concludes that the results indicate that twitter helps to reduce information assymmetry, decreasing the potential returns to be achieved from mismatched prices. The authors finally includes that twitter can lead to a higher market effieciency to comabt asset

mispricing.
Tanya Paul showcases some significant conclusions that affects attention and sentiment driven trading. As according to her paper, twitter can help to remove mispricings instead of letting uniformed investors act on quick posts and try to make a profit. This is interesting as it affects to what extent a computed ratio created from a combination of data from twitter and other variables that possibly have the opposite effect on asset mispricing can predict potential future returns.

3 Theoretical Framework

3.1 Individual & Institutional investors

In order to create an all-encompassing firm-specific measure of investor sentiment and investor attention for FAANG+M stocks, it is critical to understand the different actors participating in the market. Market participants can be divided into two groups[14]. The first group consists of less informed, retail investors which saw increasing participation during the US market boom of the 2010s[23]. The second group consists of more informed, professional and/or institutional investors. While both groups share the intention of maximizing future expected returns, they differ in their ability to effectively analyze market signals and trade accordingly[14].

Informed investors tend to complete more thorough, rigorous analysis of trading opportunities while retail investors tend to be more prone to speculation[12][14]. Informed investors have greater access to timely, relevant firm-specific or market moving financial news and data releases, rendering them less likely to trade speculatively. Institutional investors are more likely to spend time conducting thorough research because they are employed by financial institutions to maximize returns on behalf of their clients[12].

Retail investors on the other hand have limited access to timely, firm-specific or market moving information and as a result, tend to be less informed and more likely to make trading decisions based solely on financial news headlines [17] or social media posts[3]. Individual investors exhibit a tendency to be overconfident in their private trading signals[14] generated from asymmetric information[4]. Individual investors also tend to fall victim to sensation seeking behaviour since they are not limited by constraints stemming from a client's best interest[4].

The following section provides a theoretical framework by which composite firm-specific investor attention and firm-specific investor sentiment measures for FAANG+M stocks are constructed. Both measures intend to capture the aggregate firm-specific sentiment or attention of both groups of investors.

3.2 Principal Component Analysis

Principal Component Analysis (hereafter referred to as PCA) is a statistical technique used in the construction of an investor sentiment (attention) index[12][13][32]. Through leveraging the dimensionality reduction aspect of PCA, multiple individual proxies are aggregated into a single, composite index. Each individual proxy of investor sentiment (attention) is likely to "likely to include a sentiment component as well as idiosyncratic, non-sentiment-related components" (Baker and Wurlger, 2006).

PCA is used to identify the common variation in individual variables[7], or our case, proxies of investor sentiment. A similar method is used to aggregate individual investor attention proxies into a single composite attention index. In this study, PCA is used to reduce the dimensionality and capture the common sentiment or attention component[12] of a multivariate dataset consisting of individual proxies sentiment or attention. PCA is carried out by completing the following five steps[7]:

1. Standardizing the raw proxies to ensure unit variance and a mean of zero.
2. Calculating the covariance matrix of the standardized proxies.
3. Calculating the eigenvalues and eigenvectors of the covariance matrix.
4. Creating a feature vector to decide which principal components to keep.
5. Recasting the standardized proxies over the selected feature vector.

3.2.1 Standardization

Standardization of data is a useful technique when dealing with a multivariate dataset variances significantly differ across variables[7]. Variances can differ due to different units of measurement, different variances and/or characteristics[7]. Standardizing a dataset reduces the a single variable's ability to dominate an eigenvector. When standardizing a dataset, a z-score is calculated for each data point of a variable by subtracting the mean of the variable from each data point and then dividing by the variable's standard deviation. A Z-score is given by:

$$x_{i,standardized} = \frac{x_i - \mu_{variable}}{\sigma_{variable}} \tag{3.1}$$

Following the standardization, all variables are scaled to unit variance (i.e. a variance of 1) and possess a mean of 0. Without this crucial step, it would not be possible to produce reliable results using PCA[7].

3.2.2 Covariance Matrix

After standardizing the proxies, the second step requires calculating the covariance matrix. The covariance of two variables X and Y is given by:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1} \tag{3.2}$$

After computing the covariances of all variable pairs in the dataset, the covariances can be arranged into a matrix of m **x** n dimensions, where $m = n$ and n indicate the number of variables included in the covariance matrix[7]. The covariance matrix of the standardized proxies is given by:

$$C_{nXn} = \begin{pmatrix} COV(X_1,X_1) & COV(X_1,X_2) & \cdots & COV(X_1,X_n) \\ COV(X_2,X_1) & COV(X_2,X_2) & \cdots & COV(X_1,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ COV(X_n,X_1) & COV(X_n,X_2) & \cdots & COV(X_n,X_n) \end{pmatrix}$$

Following the calculation of C_{nXn} using m **x** n covariance values, a square covariance matrix of dimensions m **x** n is produced. Using a firm-specific sentiment or attention covariance matrix, the eigenvalues and eigenvectors are calculated[7][8].

3.2.3 Eigenvalues & Eigenvectors

An eigenvector's corresponding eigenvalue is denoted as λ_i and represents a value for which the following is true:

$$AX = (E \cdot \lambda)X \rightarrow (A - E\lambda)X = 0 \rightarrow A - E\lambda = 0 \tag{3.3}$$

where E is the unity matrix of n **x** n dimension for which $EA = AE = A$ is true. The unity matrix contains values of 1 throughout the diagonal:

$$E = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \rightarrow E \cdot \lambda = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{pmatrix} = \lambda$$

Equation 3.3 can then be displayed as:

$$\begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{1,2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{1,1} & \cdots & A_{n,n} \end{pmatrix} - \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{pmatrix} = \begin{pmatrix} A_{1,1} - \lambda & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} - \lambda & \cdots & A_{1,2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{1,1} & \cdots & A_{n,n} - \lambda \end{pmatrix}$$

The eigenvalues are found by calculating the determinant of 3.3 and setting equal to 0.

The corresponding eigenvectors for each respective eigenvalue are calculated by solving the system of equations given by each eigenvalue when inserted into 3.3[35]. There will always be an equal number of eigenvalues as number of variables within the covariance matrix.

3.2.4 Selecting The Feature Vector

After calculating the eigenvalues and eigenvectors of the covariance matrix, the number of principal components (hereafter referred to as PCs) to include is determined. PCs are the values within an eigenvector. The larger an eigenvector's eigenvalue, the greater the proportion of explained variance[7]. The explained variance of an eigenvalue is equivalent to the respective eigenvalue divided by the sum of all eigenvalues.

$$VAR_{PC_i} = \frac{\lambda_i}{\sum \lambda_i} \tag{3.4}$$

It is important to note that since the resulting eigenvalues are calculated using the covariance matrix of a standardized dataset where all variables contain unit variance, the sum of all eigenvalues is equal to the sum of the standard deviations of the variables. The sum of all eigenvalues is also equivalent to the total number of eigenvalues. This relationship is given by:

$$sum_{i=1}^{n} \lambda_i = n_\lambda = n_{variables} \tag{3.5}$$

The first PCs are the values within the eigenvector that correspond to the first eigenvalue. The second PCs are the values within the eigenvector that correspond to the second eigenvector, so on and so forth. In our study, the PCs corresponding to the eigenvector with the greatest explained variance (i.e. the PCs belonging to the eigenvector with the largest eigenvalue) are selected to capture the greatest common variation among the proxies of investor sentiment (attention). The feature vector refers to the PCs within the selected eigenvector.

3.2.5 Recasting The Standardized Dataset

After selecting the feature vector with the greatest explained variance, the original standardized dataset is multiplied with the selected feature vector. The values within the feature vector are hereafter referred to as the loadings.

3.3 Stationarity in Time-Series Analysis

In time-series analysis, stationarity must be taken into consideration. In this study, we regress our explanatory (independent) variables including the firm-specific sentiment and attention indices on our dependent variables including daily, weekly and monthly returns of FAANG+M stocks.

In the simplest of terms, stationarity is a measurement of the degree to which a time-series wanders off from its current path[36]. A time-series is considered stationary if it has a constant mean and variance over time. A non-stationary time-series contains a unit root. The presence of a unit root indicates current values are dependent on past values[36], such as stock prices [37]. Feeding non-stationary time-series data into a regression model can lead to spurious results and overstated statistical significance[36][37].

3.3.1 Checking For Stationarity In Time-Series Analysis

The stationarity of a time-series is determined by checking the following three conditions[36]:

$$
\begin{aligned}
&1)\; E(y_t) = \mu \rightarrow mean\, is\, finite\, and\, constant\, across\, t \\
&2)\; Var(y_t) = \sigma^2 \rightarrow variance\, is\, finite\, and\, constant\, across\, t \\
&3)\; Cov(y_t,\; y_{t-1})\; for\; s \neq 0\, is\, finite\, and\, a\, function\, of\; s\, but\, not\, of\; t
\end{aligned}
$$

Confirming stationarity is critical in time-series analysis as our models implicitly assume that current values are independent of past values. If previous values are a significant determinant of current values (i.e., the time-series contains a unit root[36]), regression models may lead to spurious results. This study intends to produce robust and reliable empirical results and thus the dangers of non-stationarity need to be taken into consideration and dealt with accordingly where necessary.

3.3.2 Overcoming Non-Stationary Time-Series Data

There exists multiple ways to overcome the dangers of non-stationarity in time-series analysis such as logarithmic transformation or differencing[5].

Logarithmic transformation consists of taking the natural logarithm of values in a time-series. Differencing transforms the data by taking the difference between the current and previous value given by:

$$\Delta x_{i,t} = x_{i,t} - x_{i,t-1} \tag{3.6}$$

When transforming a time-series via differencing, any inherent trend or seasonality present in the time-series is removed, resulting in a more stable mean and variance over time. If transformation via differencing does not suffice in rendering a time-series stationary, further transformation may be required. It is important to highlight that logarithmic transformation must always be followed by differencing[5].

3.3.3 Testing For Stationarity

To check for stationarity in a time-series, two tests are frequently used in statistical analysis[36][37]. The first test is the Augmented Dickey-Fuller test[36] (hereafter referred to as the ADF test), and the second is the Kwiatkowski Phillips Schmidt Shin test[36] (hereafter referred to as the KPSS test).

The ADF test is a parametric test which seeks to validate the existence of a unit root. The ADF test also takes into account the presence of trend and auto-correlation in a time-series[36]. The intuition behind the ADF test gives rise to the null hypothesis (H_0) that the time-series has a unit root and is non-stationary and an alternative hypothesis (H_A) that the time-series does not contain a unit root and is stationary.

The KPSS test is a non-parametric test that specifically determines whether a time-series has a constant mean and variance over time. The null and alternative hypothesis of the KPSS test are inverse to those of the ADF test, where the null hypothesis tests if a time-series is stationary and does not have a unit root while the alternative hypothesis tests if a time-series is non-stationary contains a unit root[6].

The optimal method of verifying the stationarity of a time-series is to use both the ADF and KPSS tests side by side in tandem as they are complementary to one another. If each tests confirms that a time-series is non-stationary (stationary), it can be concluded that the time-series is in fact non-stationary (stationary). If the KPSS tests indicates stationarity while the ADF test does not, the time-series is said to be trend-stationary and transformation via detrending is required. However, if the KPSS test indicates non-stationarity while the ADF test indicates stationarity, a time-series is said to be difference stationary and a differencing transformation may be be required[6].

3.4 Orthogonalization

The techniques to orthogonalize swings in firm-specific investor sentiment and investor attention to changes in US macroeconomic indicators, the Fama-French five factor model and a stock's previous price level are outlined in the following section.

Orthogonalization refers to the mathematical process of finding a set of orthogonal vectors that span a specific subspace[36]. Following the construction of the single composite firm-specific sentiment and attention indices (denoted $SENT$ and $ATTN$), it is important to reconsider that "principal component analysis cannot distinguish between a common sentiment component and a common business cycle component" (Baker and Wurgler 2006).

3.4.1 Macroeconomic Data

Fluctuations in macroeconomic data have an affect on investor sentiment which is well documented and covered extensively in the literature[32]. In order to isolate the common sentiment component from the common business cycle component[12], a similar approach to Baker and Wurgler (2006) and Beer and Zouaoui (2012) is followed. Each raw sentiment (attention) proxy used in the construction of a composite, firm-specific sentiment (attention) index is individually regressed on monthly US macroeconomic data[12].

The monthly macroeconomic dataset includes growth in industrial production index[42], growth in consumer durables[39], non-durables[41] and services[40] as well as a dummy variable for NBER recessions[38]. The NBER dummy variable is included to account for the Covid-19 pandemic that lead to a two-month technical recession in the US spanning from March until April of 2020. The residuals from these regressions (i.e., the variation not explained by fluctuations in the broader US economy) may then be a better proxy for firm-specific investor sentiment and attention[12]. The following linear regression model is used:

$$y_{i,j,t} = \alpha_{i,j} + \beta_{i,j} \cdot \sum_{t=1}^{i,j} MACRO_{k,t} + \epsilon_{i,j,t} \tag{3.7}$$

Where k indexes the macroeconomic indicator, i indexes the raw sentiment or attention proxy, and j indexes a FAANG+M stock. The orthogonalized proxy value represented by $y_{orth_{i,j,t}}$ is calculated by subtracting the predicted proxy value obtained using equation 3.7 from the actual value of proxy i at time t, which is given by:

$$y_{orth_{i,j,t}} = y_{actual_{i,j,t}} - y_{pred_{i,j,t}} \tag{3.8}$$

Due to U.S. macroeconomic data availability, the process of orthogonalizing the raw proxies is only applied to monthly investor sentiment and investor attention proxies. Alternative methods to orthogonalize the daily and weekly stock-specific datasets is outlined in following sections.

3.4.2 Fama-French: Five Factor Model

In this section, the process of orthogonalizing the raw investor sentiment (attention) proxies to the Fama-French five factor model is outlined. We follow a similar approach to authors Ding et al. (2019) and Yang and Zhou (2016). The underlying intuition is similar to section 3.4.1. The Fama-French five factor model[15] published in 2015 by Eugene Fama and Kenneth French is an expansion of the original three factor asset pricing model published in 1993[55]. The Fama-French five factor model attempts to explain the expected returns of a stock as a function of its sensitivity to five factors.

As outlined in the Introduction, the initial three factor model includes market risk, size and value factors[55]. The five factor model adds a profitability factor which is measured as the difference between return on assets and cost of equity, and an investment factor which is measured as the difference between total assets and book value of equity[15].

The Fama-French five factor model data is not available at the weekly time interval. The daily and monthly five factor model data is retrieved from the Kenneth R. French data library[56]. Individual investor sentiment (attention) proxies are orthogonalized to determine the portion of variation in investor sentiment (attention) proxies unexplained by the Fama-French five factor model. The following linear regression model is used:

$$\begin{aligned} y_{i,j,t} = \alpha_{i,j} &+ \beta_{1,i,j} \cdot RMRF_t + \beta_{2,i,j} \cdot SMB_t + \beta_{3,i,j} \cdot HML_t \\ &+ \beta_{3,i,j} \cdot PMU_t + \beta_{3,i,j} \cdot CMA_t + \epsilon_{i,j,t} \end{aligned} \tag{3.9}$$

where i indexes a raw proxy and j indexes the stock at time t. The proxy value orthogonal to the Fama-French five factor model (denoted $y_{FF_{i,j,t}}$) is calculated by subtracting the predicted proxy value obtained by equation 3.9 from the actual value of proxy i at time t, which is given by:

$$y_{FF_{i,j,t}} = y_{actual_{i,j,t}} - y_{pred,FF_{i,j,t}} \tag{3.10}$$

3.4.3 Random-Walk Model

An additional method used to determine what degree of variation in a stock price is not explained by fundamental or macroeconomic factor variation is a random-walk model[16]. A similar approach to Jones and Bandopadhyaya

(2008) is used[16]. A significant degree of the variation in current stock prices is explained by previous stock prices, hence why stock prices are non-stationary[36]. This is consistent with the efficient market hypothesis which suggests that markets are informationally efficient whereby all readily available information is continuously discounted into a stock's current market price[11].

Authors Jones and Bandopadhyaya point out that “past values of the (S&P500) index itself capture all relevant economic information that affects the contemporaneous index values” and that “any unexplained portion of the daily movement in the (S&P500) index must then result from changes in other non-economic factors” (Jones and Bandopadhyaya, 2008). The variation in the random-walk model residuals may be a result of fluctuations in firm-specific investor sentiment and/or attention, the precise variation our study intends to measure. In order to determine what portion of a stock price is not explained by past values, we use the following linear regression model:

$$P_{i,t} = \alpha_{i,0} + \beta_{i,1} P_{i,t-1} + Residual_{i,t} \tag{3.11}$$

In equation 3.11, $P_{i,t}$ is the current price and $P_{i,t-1}$ is the previous price of stock i. The variation not explained by the previous price is given by $Residual_{i,t}$. We can now regress $Residual_{i,t}$ on a sentiment or attention proxy to determine what degree of the variation in price not explained by previous values is explained by an individual or composite investor sentiment (attention) proxy. To determine the contemporaneous relationship, we use the current value of a proxy given by $Proxy_{i,t}$. To determine the predictive relationship, we use the previous value of a proxy given by $Proxy_{i,t-1}$. This leads to the following two linear regression models:

$$Residual_{i,t} = \alpha_{i,0} + \beta_{i,1} Proxy_{i,t} + \epsilon_{i,t} \tag{3.12}$$

$$Residual_{i,t} = \alpha_{i,0} + \beta_1 Proxy_{i,t-1} + \epsilon_{i,t} \tag{3.13}$$

where $Proxy_{i,t}$ or $Proxy_{i,t-1}$ is individual proxy or composite index i for that may explain $Residual_{i,t}$.

4 Methodology

4.1 Empirical Approach

To explore the relationship between firm-specific investor sentiment (attention) and the returns of FAANG+M stocks, multiple raw proxies of investor sentiment (attention) are aggregated using techniques outlined in section 3 to form a composite firm-specific investor sentiment (attention) index. As this is an exploratory study, multiple explanatory and predictive models are tested to analyze the relationships of interests. The following sections outline the methodology used in carrying out our study.

4.2 Data Collection

Data is collected from various databases at daily, weekly, and monthly time intervals. Individual datasets containing the raw, stock-specific sentiment (attention) proxies as well as stock-specific data are created for all FAANG+M stocks, generating a total of 18 stock and frequency specific datasets. Due to the BSV factors time-series data beginning January 1, 2015, the timeline of this study spans the seven year period of January 1, 2015 until December 31, 2021. All variables and database sources are outlined in the following sections.

4.2.1 Price, Trading Volume & Shares Outstanding

Daily, weekly and monthly prices ($P_{i,t}$) of stock i at time t are retrieved from the Bloomberg database. Dependent variables including simple returns, $R_{i,t} = \frac{P_{i,t}-P_{i,t-1}}{P_{i,t-1}}$, logarithmic returns, $R_{i,t_{log}} = ln(\frac{P_{i,t}}{P_{i,t-1}})$ and the random-walk model residuals as outlined in section 3.4.3 are calculated using stock prices.

Daily shares outstanding ($SHOUT_{i,t}$) and daily trading volume ($TV_{i,t}$) are retrieved from the North American Compustat database accessed through the Wharton Research Data Services platform. By summing the trading volume of each respective trading day in a specified week or month, weekly and monthly trading volumes are calculated. Weekly and monthly trading volume for stock i and time t is given by:

$$TV_{i,weekly} = TV_{i,t} + TV_{i,t+1} + ... + TV_{i,t+n_j} \tag{4.1}$$

$$TV_{i,monthly} = TV_{i,t} + TV_{i,t+1} + ... + TV_{i,t+n_k} \tag{4.2}$$

where n_j represents the number of trading days in trading week j and n_k represents the number of trading days in trading month k.

4.2.2 Market Turnover Rate

Market turnover rate for stock i at trading day t is calculated by $MTR_{i,t} = \frac{TV_{i,t}}{SHOUT_{i,t}}$, representing trading volume in a specified period as a portion of total shares outstanding. Authors Baker and Stein (2004) argue that market liquidity can serve as a sentiment indicator, arguing that "in the presence of short-sales constraints, high liquidity is a symptom of the fact that the market is dominated by these irrational investors, and hence is overvalued" (Baker and Stein, 2004).

Authors Yang and Zhou (2016) calculate an adjusted turnover rate as $ATR_{i,t} = \frac{R_{i,t}}{|R_{i,t}|} \cdot \frac{TV_{i,t}}{SHOUT_{i,t}}$, suggesting that "the turnover rate cannot judge whether the investor sentiment is optimistic or pessimistic" (Yang and Zhou, 2016). We subscribe to the reasoning of Yang and Zhou (2016) and consider market turnover rate a proxy for investor attention in this study.

Raw trading volume data is expected to be non-stationary, while market turnover rate is expected to be stationary possessing a relatively constant mean and variance over time. The market turnover rate at the weekly (monthly) time interval differs from the daily interval because the aggregate trading volume in a given week ($TV_{i,weekly}$) (month ($TV_{i,monthly}$)) is used in place of daily trading volume at time t.

4.2.3 Volatility Midpoint

The daily volatility midpoint ($VMP_{i,t}$) of stock i at trading day t is calculated as the daily midpoint of the implied volatility values of put and call options with 30 days until expiration. Implied volatility values corresponding to 30-day options are used in order to mirror the utility of the VIX as a market-wide sentiment indicator, albeit at a firm-specific level. The implied volatility data for is retrieved from the Ivy-DB US Options database accessed through the Wharton Research Data Services platform.

The daily volatility midpoint is calculated using equation 4.3 below, where $\sigma_{i,t,\,Put,\,30\,days}$ is the 30-day implied volatility of put-options for stock i at trading day t and $\sigma_{i,t,\,Call,\,30\,days}$ is the 30-day implied volatility of call options for stock i at trading day t.

This process is repeated for each trading day in this study's timeline to generate a firm-specific volatility midpoint time-series for FAANG+M stocks. In order to match the volatility midpoint with our weekly (monthly) data sets, the end of week (end of month) volatility midpoint is matched to the corresponding price data. For months ending on a Saturday or Sunday, the closest prior trading day's volatility midpoint is matched. The volatility midpoint for stock i at time t is given by:

$$VMP_{i,t} = \frac{\sigma_{i,t,\,put,\,30\,days} + \sigma_{i,t,\,call,\,30\,days}}{2} \tag{4.3}$$

4.2.4 Put-Call Ratio

The put-call ratio ($PCR_{i,t}$) of stock i at trading day t is calculated as the total put option volume divided by the total call option volume on stock i at trading period t. The daily put and call options trading volume is retrieved from the Ivy-DB US Option Metrics database accessed through the Wharton Research Data Services platform. Similarly to the volatility midpoint, we repeat this process for all stocks in our sample to create stock specific put-call ratio time-series. Daily put-call ratios are given by:

$$PCR_{i,t} = \frac{V_{put\ options,\,i,t}}{V_{call\ options,\,i,t}} \tag{4.4}$$

When calculating the put-call ratio on a weekly (monthly) time interval, total option volume for each trading day in week j (month k) is used. The weekly (monthly) put-call ratios are matched to the stock-specific weekly (monthly) datasets. Put-call ratios at the weekly and monthly time intervals are given by:

$$PCR_{i,weekly} = \frac{\sum_{t=1}^{n} V_{put\ options,\,i,j}}{\sum_{t=1}^{n} V_{call\ options,\,i,j}} \tag{4.5}$$

$$PCR_{i,monthly} = \frac{\sum_{t=1}^{n} V_{put\ options,\,i,k}}{\sum_{t=1}^{n} V_{call\ options,\,i,k}} \tag{4.6}$$

4.2.5 Search Volume Index

Daily search volume data is retrieved from Google Trends for FAANG+M stocks to construct our search volume index ($SVI_{i,t}$) which intends to serve as a proxy for investor attention. Google Trends does not provide raw search frequency data in search units, but instead represents search frequency values indexed on the maximum value within a specified time frame. The maximum value ($SVI_{max,i,t}$) in a specified time frame is set to 100 while the remaining values take on a value of any integer between 0 and 100 as a function of their relative size to $SVI_{max,i,t}$.

The literature[17] suggests using a stock's ticker as a search term since stock tickers are unambiguous and a unique identifier (Da et Al, 2011). However, as mentioned in the Theoretical Framework, $SVI_{i,t}$ is designed to capture the aggregate investor attention of informed and uninformed investors[14]. An informed investor will likely search for a stock ticker directly while an uninformed investor may be more likely to search for a firm's name followed by “stock” due to lack of experience and knowledge of stock tickers[17].

To capture informed and uninformed investor attention, both a stock’s ticker and the phrase “company name” + “stock” are entered as search terms. For example the search frequency data for “AMZN” and “Amazon Stock” is exported from Google Trends. To aggregate the two search terms into a single SVI value, the search frequency time-series for both search terms needs to be indexed to the same $SVI_{max,i,t}$. We search for both terms simultaneously by entering two search terms into the Google Trends platform[51].

In order to retrieve daily data from the Google Trends platform, data must be retrieved and aggregated from multiple smaller time frames. By design, as the specified time period entered into the Google Trends platform increases, the frequency of the time-series retrieved decreases from daily, to weekly, to monthly. To ensure the multiple smaller time frames can b combined, we ensure an overlap of exactly one day. The current time frame is denoted TF_t and the previous time frame is denoted TF_{t-1}. The $SVI_{i,t}$ value on the last day of previous time frame (TF_{t-1}) is equal to the $SVI_{i,t}$ value on the first day of current time frame t is denoted FDC_t. This equality is given by:

$$max(TF_{t-1}) = min(TF_t) \tag{4.7}$$

The ratio at time t between the search frequency value of the last day of the previous time frame and the first day of the current time frame is given by:

$$Comparative\ Ratio_t = \frac{max(TF_{t-1})}{min(TF_t)} \tag{4.8}$$

We multiply all search frequency values in the current time frame by the corresponding comparative ratio given by equation 4.8 . This is to ensure that all search frequency values in the current time frame are indexed according to the previous time frame. Since the raw, stock-specific SVI datasets each contain two time-series, a separated comparative ratio is calculated for each column.

This process of calculating a comparative ratio and multiplying the search frequency values contained in the subsequent time frame is repeated for all time frames. This process generates a resulting column for each date with all values indexed to the same SVI_{max}. In order to generate weekly and monthly search frequency data values,the search frequency values for the respective period are summed resulting in a total SVI for each week or month. The script used to carry out process of creating SVI time-series can be found in the python file titled ”GoogleTrends”[1]

Due to the repeated indexing throughout the process of converting the raw search frequency data into the SVI time-series, values can occasionally become very small. These values are manually set to 0 except if they belong to $t_{max,\,LDP}$ or $t_{min,\,FDC}$. If the value belongs to $t_{max,\,LDP}$ or $t_{min,\,FDC}$, they are manually set to < 0.01 and reflect the comparative ratio of the second column.

We acknowledge that there are some issues with the SVI data from Google Trends, primarily that values are repeatedly indexed and can only assume the value of an integer. We are forced to remodel our data following the process outlined above to overcome these issues. Despite the limitations of Google Trends data, we don't expect these issues or our workarounds to play a significant role in the overall purpose of Google Trends data in our study.

4.2.6 Bloomberg Social Velocity Factors

The remaining proxies for investor sentiment and investor attention are retrieved from the Bloomberg database. As mentioned in the literature review, the BSV factors include Twitter Positive Count (TPC), Twitter Negative Count (TNC), Twitter Publication Count (TC), News Positive Count (NPC), News Negative Count (NNC) and News Publication Count (NC).

When selecting the weekly (monthly) time interval in the Bloomberg database, the end-of-week (end-of-month) value is provided as opposed to the aggregate end-of-week (end-of-month) count. The weekly (monthly) BSV values for FAANG+M stocks are manually calculated and given by:

$$BSV_{i,s,weekly} = BSV_{i,s,t} + BSV_{i,s,t+1} + ... + BSV_{i,s,t+n_j} \tag{4.9}$$

$$BSV_{i,s,monthly} = BSV_{i,s,t} + BSV_{i,s,t+1} + ... + BSV_{i,s,t+n_k} \tag{4.10}$$

where i indexes the BSV factor (i.e., NC, NPC, NNC etc.), s indexes the stock, n_j represents the number of trading days in trading week j, and n_k represents the number of trading days in trading month k.

4.2.7 Raw Proxy Stationarity

To check the viability of our variables in regression models, both the ADF and KPSS tests are used to confirm a sentiment or attention proxy's stationarity. ADF and KPSS stationarity tests are conducted at the 5% significance level and are carried out for all firm-specific sentiment or attention proxies at daily, weekly and monthly time intervals.

Daily	AMZN	AAPL	META	GOOGL	MSFT	NFLX
Price - ADF	No	No	No	No	No	No
Price - KPSS	No	No	No	No	No	No
TPC - ADF	Yes	Yes	Yes	Yes	Yes	Yes
TPC - KPSS	No	No	No	No	No	No
TNC - ADF	Yes	Yes	Yes	Yes	Yes	Yes
TNC - KPSS	No	No	No	No	No	No
NPC - ADF	Yes	Yes	Yes	Yes	Yes	Yes
NPC - KPSS	No	No	No	No	No	No
NNC - ADF	Yes	Yes	Yes	Yes	Yes	Yes
NNC - KPSS	No	Yes	No	No	No	Yes
PCR - ADF	Yes	Yes	Yes	Yes	Yes	Yes
PCR - KPSS	No	No	No	No	No	No
VMP - ADF	Yes	Yes	Yes	Yes	Yes	Yes
VMP - KPSS	Yes	No	No	No	No	Yes
TV - ADF	Yes	No	Yes	Yes	Yes	Yes
TV - KPSS	Yes	No	No	No	Yes	No
TC - ADF	No	Yes	Yes	Yes	Yes	Yes
TC - KPSS	No	No	No	No	No	No
NC - ADF	Yes	Yes	Yes	Yes	Yes	Yes
NC - KPSS	No	No	No	Yes	No	No
SVI - ADF	No	Yes	Yes	No	No	Yes
SVI - KPSS	No	No	No	No	No	No
MTR - ADF	Yes	Yes	Yes	Yes	Yes	Yes
MTR - KPSS	Yes	Yes	No	No	Yes	No

Table 1: ADF & KPSS test results of raw sentiment and attention proxies at the daily interval.

For $Price_{Daily}$, both tests confirm non-stationarity across all FAANG+M stocks, in line with our expectation stemming from the literature[35][36]. For BSV_{Daily} variables with the exception of NC_{Daily}, ADF tests point to stationarity while the KPSS tests do not. For NC_{Daily}, only GOOGL tests positively for stationarity in the KPSS test. We conclude that NC_{Daily} generally tests negatively for stationarity in the KPSS tests. The stationarity test results hint towards the stock-specific NC_{Daily} time-series being difference stationary.

The test results for PCR_{Daily} are similar to most of the BSV_{Daily} variables, suggesting it is also difference stationary. The stationarity tests show similar results for VMP_{Daily} with the exception of two stocks ($AMZN$ and $NFLX$) where the KPSS test points to stationarity. Since the majority of FAANG+M stocks test negatively for stationarity using the KPSS test, we conclude that the VMP_{Daily} time-series are difference stationary.

The stationarity test results for TV_{Daily} are somewhat mixed. The majority of ADF tests point toward stationarity while the majority of KPSS tests point to non-stationarity, suggesting that TV_{Daily} is difference stationary. For SVI_{Daily}, no stocks test positively for stationarity using the KPSS tests while half test positively for stationarity using the ADF tests. We conclude that SVI_{Daily} is either non-stationary or difference stationary. The ADF tests on MTR_{Daily} prove stationarity while only half of the KPSS tests prove stationarity.

Weekly	**AMZN**	**AAPL**	**META**	**GOOGL**	**MSFT**	**NFLX**
Price - ADF	No	No	No	No	No	No
Price - KPSS	No	No	No	No	No	No
TPC - ADF	No	No	No	No	Yes	No
TPC - KPSS	No	No	No	No	No	No
TNC - ADF	Yes	No	Yes	Yes	Yes	No
TNC - KPSS	No	No	No	No	No	No
NPC - ADF	No	Yes	No	Yes	No	Yes
NPC - KPSS	No	No	No	No	No	No
NNC - ADF	Yes	Yes	Yes	No	Yes	Yes
NNC - KPSS	No	Yes	Yes	No	No	Yes
PCR - ADF	No	No	Yes	Yes	Yes	Yes
PCR - KPSS	No	No	No	No	No	No
VMP - ADF	Yes	Yes	No	Yes	Yes	Yes
VMP - KPSS	Yes	No	No	No	No	Yes
TV - ADF	Yes	No	Yes	Yes	Yes	No
TV - KPSS	Yes	No	No	No	Yes	No
TC - ADF	No	No	Yes	No	No	No
TC - KPSS	No	No	No	No	No	No
NC - ADF	No	Yes	Yes	Yes	Yes	No
NC - KPSS	No	No	No	Yes	No	No
SVI - ADF	No	No	Yes	No	No	No
SVI - KPSS	No	No	No	No	No	No
MTR - ADF	Yes	Yes	Yes	Yes	Yes	No
MTR - KPSS	Yes	Yes	No	No	Yes	No

Table 2: ADF & KPSS test results of raw sentiment and attention proxies at the weekly interval.

At the weekly interval, different stationarity test results are observed. Similar to the daily interval, both tests $Price_{Weekly}$ show non-stationarity. For TC_{Weekly}, NC_{Weekly} and TPC_{Weekly}, the majority of the ADF and KPSS tests prove non-stationarity. For NPC_{Weekly} and TNC_{Weekly}, a majority of ADF tests show stationarity suggesting both variables are difference stationary.

For NNC_{Weekly}, only half of the KPSS tests prove stationarity. We cannot conclude whether NNC_{Weekly} is difference stationary or stationary. The tests results for VMP_{Weekly} and PCR_{Weekly} are almost identical to the daily interval. We conclude that both variables remain difference stationary.

The tests for TV_{Weekly} are somewhat mixed. Majority of the ADF tests prove stationarity and a majority of the KPSS tests prove non-stationarity. We conclude that TV_{Weekly} remains difference stationary. For SVI_{Weekly}, no stocks test positively for stationarity in the KPSS tests and all but one test negatively for staionarity in the ADF tests. We conclude that SVI_{Weekly} is non-stationary. The MTR_{Weekly} test results are to those at the daily interval, showing that MTR_{Weekly} remains difference stationary.

Monthly	**AMZN**	**AAPL**	**META**	**GOOGL**	**MSFT**	**NFLX**
Price - ADF	No	No	No	No	No	No
Price - KPSS	No	No	No	No	No	No
TPC - ADF	No	No	No	No	Yes	No
TPC - KPSS	No	No	No	No	No	No
TNC - ADF	Yes	Yes	Yes	Yes	No	No
TNC - KPSS	Yes	No	Yes	No	No	No
NPC - ADF	No	Yes	Yes	No	Yes	Yes
NPC - KPSS	No	No	No	Yes	No	No
NNC - ADF	Yes	Yes	Yes	No	Yes	Yes
NNC - KPSS	No	Yes	Yes	No	No	Yes
PCR - ADF	No	No	Yes	Yes	Yes	Yes
PCR - KPSS	No	Yes	Yes	No	No	No
VMP - ADF	Yes	No	Yes	No	Yes	Yes
VMP - KPSS	Yes	No	No	Yes	No	Yes
TV - ADF	Yes	No	Yes	Yes	Yes	No
TV - KPSS	Yes	No	Yes	Yes	Yes	No
TC - ADF	No	No	No	No	No	No
TC - KPSS	No	No	No	No	No	No
NC - ADF	No	No	Yes	No	Yes	No
NC - KPSS	Yes	No	No	Yes	No	Yes
SVI - ADF	No	No	No	No	No	No
SVI - KPSS	No	No	No	No	No	No
MTR - ADF	Yes	Yes	Yes	Yes	Yes	No
MTR - KPSS	Yes	Yes	No	Yes	Yes	No

Table 3: ADF & KPSS test results of raw sentiment and attention proxies at the weekly interval.

At the monthly interval, different stationarity test results are observed. Similar to the daily and weekly intervals, both tests for $Price_{Monthly}$ point toward non-stationarity. For $TC_{Nonthly}$ and $TPC_{Monthly}$, the majority of both the ADF and KPSS tests prove non-stationarity. However, half of the KPSS tests for $NC_{Nonthly}$ point toward stationarity.

For $NPC_{Monthly}$ and $TNC_{Monthly}$, the majority of ADF tests point toward stationarity indicating both time-series are difference stationary for all FAANG+M stocks. Half of the KPSS tests on $NNC_{Monthly}$ prove stationarity and we cannot decisively conclude that $NNC_{Monthly}$ is difference stationary.

The majority of ADF tests on $PCR_{Monthly}$ prove stationarity while a majority of KPSS tests prove non-stationarity, suggesting that $PCR_{Monthly}$ is difference stationary. For $VMP_{monthly}$, the majority of the ADF tests reveal stationarity while only half of the KPSS tests reveal stationarity, suggesting that $VMP_{monthly}$ is either stationary or difference stationary.

The results for $TV_{monthly}$ are again mixed at the monthly interval. A majority of the ADF and KPSS tests prove stationarity, suggesting that $TV_{monthly}$ is stationary. For $SVI_{monthly}$, no stocks test positively for stationarity in either the KPSS or ADF tests. We can conclude that $SVI_{monthly}$ is non-stationary, suggesting that there is a trend present in Google Trends search frequency data for this study's stock-specific search terms. MTR_{Montly} results are different from the daily and weekly interval as the majority of the KPSS test results now prove stationarity. We conclude that MTR_{Montly} is stationary at a monthly interval.

The mixed results of the stationarity tests suggest that a majority of our sentiment and attention proxies appear to be difference stationary or non-stationary. Transformation via differencing must be performed on the raw sentiment (attention) proxies prior to constructing the frequency and firm-specific sentiment (attention) indices.

4.3 Correlation Coefficients

An important concept to understand when looking at linear relationships between variables is correlation, a measurement of the strength of the linear relationship between two variables. Correlation, denoted ρ can either be positive or negative, where a positive correlation indicates that as one variable increases, so does the other. A negative correlation indicates the inverse.

A correlation coefficient can assume any value between -1 and 1 where -1 indicates a perfect negative linear relationship and 1 indicates a perfect positive linear relationship while a correlation of 0 indicates the absence of a linear relationship[53]. Correlation is calculated using the following equation[53]:

$$\rho = \frac{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sigma_X \cdot \sigma_Y} \tag{4.11}$$

To determine the statistical significance of a correlation coefficient, a hypothesis test is used to determine if ρ is significantly different from 0. The null and alternative hypotheses are $H_0 : \rho = 0$ and $H_a : \rho \neq 0$ respectively. The unknown correlation is denoted ρ and the sample coefficient is denoted r[53]. The significance is measured by its t-stat which is calculated using the following;

$$t_c = \frac{r - \rho}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} \tag{4.12}$$

4.4 Granger Causality

In order to measure the causality between two variables it possible to conduct a Granger test. The Granger test determines whether past values of an explanatory variable(X) can be good in predicting a response variable(Y). The null hypothesis of the test is that past values or lagged values of X cannot predict values of Y, which is also called Granger cause. Rejecting the null hypothesis indicates that lags of X can predict variations in Y. The test can also be switched so that the test checks if lagged values of Y can predict variations in X. Performing the test both ways can give a better understand of the causality between two variables[58].

4.5 Sentiment & Attention Index Construction

In the following the section, the process of aggregating the raw sentiment (attention) proxies into a single, composite sentiment (attention) index is outlined. A composite investor sentiment (attention) index is denoted $SENT$ ($ATTN$). $SENT$ is constructed by aggregating TPC, TNC, NPC, NNC, VMP and PCR proxies, while $ATTN$ is constructed by aggregating TC, NC, SVI and MTR proxies.

4.5.1 Detrending & Standardizing

The initial step to constructing the $SENT$ and $ATTN$ indices is to remove the presence of trend from the raw proxies to render them stationary. This is done through differencing as outlined in section 3.3.2. Following the removal of trend, raw proxies are standardized using the method outlined in section 3.1.

4.5.2 First Principal Components

After standardizing, the first PCs of the 6 sentiment (4 attention) proxies and their respective lags are calculated. Following the approach first laid out by authors Baker and Wurgler (2006), lags of the proxies are included to take into the fact that "some variables take longer to reveal the same sentiment" or attention[12]. After including the lagged proxies, an attention dataset resembles the following table below:

Date	TC	TC_{lag1}	NC	NC_{lag1}	SVI	SVI_{lag1}	MTR	MTR_{lag1}
t	TC_t	TC_{t-1}	NC_t	NC_{t-1}	SVI_t	SVI_{t-1}	MTR_t	MTR_{t-1}
t+1	TC_{t+1}	TC_t	NC_{t+1}	NC_t	SVI_{t+1}	SVI_t	MTR_{t+1}	MTR_t
t+2	TC_{t+2}	TC_{t+1}	NC_{t+2}	NC_{t+1}	SVI_{t+2}	SVI_{t+1}	MTR_{t+2}	MTR_{t+1}
t+3	TC_{t+3}	TC_{t+2}	NC_{t+3}	NC_{t+2}	SVI_{t+3}	SVI_{t+2}	MTR_{t+3}	MTR_{t+2}

The sum-product of the proxies and their lags with their respective first principal components (hereafter referred to as loadings) generates the first stage index denoted $FSS_{i,t}$ for sentiment and $FSA_{i,t}$ for attention. The correlation coefficient of each proxy and lagged proxy with the first stage index is calculated. $SENT$ ($ATTN$) is defined as the first principal component of the covariance matrix of each respective proxy's current or lagged value, whichever has a higher correlation with the first-stage index[12]. $FSS_{i,t}$ and $FSA_{i,t}$ are given by:

$$\begin{aligned} FSS_{i,t} = \beta_{i,1} \cdot TPC_{i,t} + \beta_{i,2} \cdot TPC_{i,t-1} + \beta_{i,3} \cdot TNC_{i,t} + \beta_{i,4} \cdot TNC_{i,t-1} \\ + \beta_{i,5} NPC_{i,t} + \beta_{i,6} \cdot NPC_{i,t-1} + \beta_{i,7} \cdot NNC_{i,t} + \beta_{i,8} \cdot NNC_{i,t-1} \\ + \beta_{i,9} \cdot VMP_{i,t} + \beta_{i,10} \cdot VMP_{i,t-1} + \beta_{i,11} \cdot PCR_{i,t} + \beta_{i,12} \cdot PCR_{i,t-1} \end{aligned} \tag{4.13}$$

$$\begin{aligned} FSA_{i,t} = \beta_{i,1} \cdot TC_{i,t} + \beta_{i,2} \cdot TC_{i,t-1} + \beta_{i,3} \cdot NC_{i,t} + \beta_{i,4} \cdot NC_{i,t-1} \\ + \beta_{i,5} \cdot SVI_{i,t} + \beta_{i,6} \cdot SVI_{i,t-1} + \beta_{i,7} \cdot MTR_{i,t} + \beta_{i,8} \cdot MTR_{i,t-1} \end{aligned} \tag{4.14}$$

where i indexes a FAANG+M stock's sentiment (attention) proxy value at time t. The time-series values of $SENT$ ($ATTN$) indices are calculated using matrix multiplication between the sentiment (attention) proxies and a transpose of the loadings.

Following the calculation of FSS (FSA), the proxy's current or lagged value is kept depending on whichever has the higher absolute correlation coefficient with FSS (FSA). This is given by:

$$Max_{Corr} = Max(|Corr(FSS,\ X)|,\ |Corr(FSS,\ X_{lag1})|) \tag{4.15}$$

$$Max_{Corr} = Max(|Corr(FSA,\ X)|,\ |Corr(FSA,\ X_{lag1})|) \tag{4.16}$$

where X represents a sentiment or attention proxy. The above process is repeated and a new covariance matrix, set of eigenvalues and eigenvectors are produced to calculate loadings to construct the final stage sentiment (attention) index denoted $SENT$ ($ATTN$). The sentiment (attention) proxy loadings are divided by the standard deviation of $SENT$ ($ATTN$) to ensure unit variance in the final stage index[12].

4.5.3 Final Stage Sentiment Index

The final stage sentiment index ($SENT$) is constructed using 6 individual proxies of investor sentiment. The included proxies (i.e. current or lagged) vary across FAANG+M stocks. An example of $SENT$ construction is given by:

$$\begin{aligned} SENT_{i,t} = \beta_{i,1} \cdot TPC_{i,t} + \beta_{i,2} \cdot TNC_{i,t-1} + \beta_{i,3} \cdot NPC_{i,t} \\ + \beta_{i,4} \cdot NNC_{i,t} + \beta_{i,5} \cdot PCR_{i,t} + \beta_{i,6} \cdot VMP_{i,t-1} \end{aligned} \tag{4.17}$$

4.5.4 Final Stage Attention Index

The final stage attention index ($ATTN$) is constructed using 4 individual proxies of investor attention. The included proxies (i.e. current or lagged) vary for FAANG+M stocks. An example of $ATTN$ construction is given by:

$$ATTN_{i,t} = \beta_{i,1} \cdot TC_{i,t} + \beta_{i,2} \cdot NC_{i,t} + \beta_{i,3} \cdot SVI_{i,t-1} + \beta_{i,4} \cdot MTR_{i,t-1} \tag{4.18}$$

4.6 Regression Models

In this section, the regression models used to explore the relationship between the composite investor sentiment (attention) indices and the returns of FAANG+M stocks are outlined. The regression models are used to determine to what extent (if any at all) does investor sentiment (attention) impact the returns of FAANG+M stocks.

4.6.1 Null & Alternative Hypotheses

The linear and exponential regression models presented in the following sections seek to test if the coefficient of the composite investor sentiment or investor attention index is equal to zero. The outlined research questions and subsequent regression models give rise to the following null and alternative hypotheses:

$$\begin{aligned} H_{0,1} &: \beta_{SENT_{i,t-1}} = 0 \\ H_{A,1} &: \beta_{SENT_{i,t-1}} \neq 0 \end{aligned} \tag{4.19}$$

$$\begin{aligned} H_{0,2} &: \beta_{ATTN_{i,t-1}} = 0 \\ H_{A,2} &: \beta_{ATTN_{i,t-1}} \neq 0 \end{aligned} \tag{4.20}$$

where i indexes the stock (indicating a $SENT$ or $ATTN$ index is stock specific) and t indexes the time (indicating that we are testing the significance of previous period $SENT$ or $ATTN$). The null and alternative hypothesis outlined in equation 4.19 tests whether the coefficient of a composite, firm-specific investor sentiment index is statistically different from zero. The null and alternative hypothesis outlined in equation 4.20 tests whether the coefficient of a composite, firm-specific investor attention index is statistically different from zero.

4.6.2 Linear Regression Models

As this is an exploratory study, the most reasonable starting point is a linear regression model. A linear regression model is used to determine if a linear relationship exists between a FAANG+M stock's composite sentiment (attention) index and its return. Prior to running regressions, the stationarity of independent and dependent variables must be confirmed. The first proposed linear regression model uses the change in price at time t ($\Delta P_t = P_t - P_{t-1}$) as the dependent variable and the values of $SENT$ and $ATTN$ at time $t-1$ as the independent variables. As stock prices tend to be non-stationary, we expect this model to be unsuitable for our analysis.

$$\Delta P_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot ATTN_{i,t-1} + \beta_{i,s} \cdot SENT_{i,t-1} + \epsilon_{i,t} \tag{4.21}$$

where i indexes the stock, a indexes the firm-specific composite attention index, s indexes the firm-specific composite sentiment index, and t indexes time.

The second proposed linear model intends to explore the linear relationship between a FAANG+M stock's return at time t and the level of its firm-specific sentiment and attention index at time $t-1$. Converting stock prices to stock returns will generally render a non-stationary time-series stationary[36]. The second linear regression model is given by:

$$R_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot ATTN_{i,t-1} + \beta_{i,s} \cdot SENT_{i,t-1} \tag{4.22}$$

The third proposed linear model is similar to equation 4.22, except sentiment and attention at time $t-1$ are replaced with the unit change in sentiment and attention at time $t-1$ given by $\Delta SENT_{i,t-1} = SENT_{i,t-1} - SENT_{i,t-2}$ and $\Delta ATTN_{i,t-1} = ATTN_{i,t-1} - ATTN_{i,t-2}$ respectively. We expect both attention and sentiment indices to be stationary after differencing.

$$R_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \Delta SENT_{i,t-1} \tag{4.23}$$

The fourth proposed linear considers the percentage change in $SENT$ and $ATTN$ given by $\%\Delta SENT_{i,t-1} = \frac{SENT_{i,t-1} - SENT_{i,t-2}}{SENT_{i,t-2}}$ and $\%\Delta ATTN_{i,t-1} = \frac{ATTN_{i,t-1} - ATTN_{i,t-2}}{ATTN_{i,t-2}}$ respectively. We expect any remaining non-stationarity after differencing to be removed.

$$R_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \%\Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \%\Delta SENT_{i,t-1} \tag{4.24}$$

The final proposed linear model explores the relationship between change in price at time t and change in sentiment and attention at time $t-1$. We expect differencing a FAANG+M stock's price level time-series to make it more suitable for use in a linear regression model.

$$\Delta P_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \Delta SENT_{i,t-1} \tag{4.25}$$

4.6.3 Exponential Regression Models

The following section considers whether the relationship between investor sentiment and attention and FAANG+M stock returns is non-linear. Should the results of the proposed linear regression models outlined in section 4.6.2 produce insignificant results, the relationships of interest may be non-linear. The mathematical derivations used to transform an exponential model into a linear model are outlined in the following section. The basic logarithmic rules[54] that are used in such a transformation are the following:

$$\begin{aligned}
&1)\ Product\,rule:\ ln(x \cdot y) = ln(x) + ln(y)\\
&2)\ Power\,rule:\ ln(b^k) = k \cdot ln(b)\\
&3)\ Logarithm\,of\,e:\ ln(e) = 1\\
&4)\ Undefined\,logarithms:\ ln(x)\,is\,undefined\,for\,x \leq 0
\end{aligned}$$

Using the above four rules, we transform an exponential equation in the form of $y = a \cdot e^{b \cdot x}$ to a linear equation in the form of $y = bx + a$ given by:

$$\begin{aligned}
&y = a \cdot e^{b \cdot x} \\
&ln(y) = ln(a \cdot e^{b \cdot x}) \\
&ln(y) = ln(a) + ln(e^{b \cdot x}) \\
&ln(y) = ln(a) + b \cdot x \cdot ln(e) \\
&ln(y) = ln(a) + b \cdot x \cdot 1
\end{aligned}$$

The transformation has two requirements since a logarithm in the form of $ln(x)$ is undefined for $x \leq 0$. The two requirements are that both a and y are positive numbers ($y, a > 0$). Through the outlined transformation, linear regressions are performed on exponential models[44]. As our proposed models contain stock returns at time t as the dependent variable, $R_{i,t}$ must be positive. Knowing $R_{i,t} > -100\%$ holds true for stock returns, 1 is added to $R_{i,t}$ and $R_{i,t} + 1$ then becomes our dependent variable. As $R_{i,t} + 1 > -1 + 1 \rightarrow R_{i,t} + 1 > 0$, the logarithmic transformation of $R_t + 1$ is given by:

Theorem 1. *Log Returns*

$$ln(R_t + 1) = ln\left(\frac{P_t}{P_{t-1}}\right)$$

Proof.

$$\begin{aligned}
ln(R_t + 1) &= ln\left(\frac{P_t - P_{t-1}}{P_{t-1}} + 1\right) \\
ln(R_t + 1) &= ln\left(\frac{P_t - P_{t-1}}{P_{t-1}} + \frac{P_{t-1}}{P_{t-1}}\right) \\
ln(R_t + 1) &= ln\left(\frac{P_t - P_{t-1} + P_{t-1}}{P_{t-1}}\right) \\
ln(R_t + 1) &= ln\left(\frac{P_t}{P_{t-1}}\right)
\end{aligned}$$

□

The exponential models are on the form $y_{i,t} = \alpha \cdot x_{1,i,t}^{\beta_1} \cdot e^{\beta_2 \cdot x_{2,i,t}}$ where $x_{1,i,t}$ and $x_{2,i,t}$ are two different explanatory variables and β_1 and β_2 are their respective coefficients. The equation's constant (or y-intercept) is represented by α.

$$
\begin{aligned}
R_{i,t} + 1 &= \alpha \cdot x_{1,i,t}^{\beta_{1,i}} \cdot e^{\beta_{2,i} \cdot x_{2,i,t}} \\
ln(R_{i,t} + 1) &= ln(\alpha \cdot x_{1,i,t}^{\beta_{1,i}} \cdot e^{\beta_{2,i} \cdot x_{2,i,t}}) \\
R_{log,i,t} &= ln(\alpha) + ln(x_{1,i,t}^{\beta_{1,i}}) + ln(e^{\beta_{2,i} \cdot x_{2,i,t}}) \\
R_{log,i,t} &= ln(\alpha) + \beta_1 \cdot ln(x_{1,i,t}) + \beta_{2,i} \cdot x_{2,i,t} \cdot ln(e) \\
R_{log,i,t} &= ln(\alpha) + \beta_{1,i} \cdot ln(x_{1,i,t}) + \beta_{2,i} \cdot x_{2,i,t}
\end{aligned}
$$

Since logarithms are undefined for values less than or equal to 0 and assuming the variable $x_{1,i} \neq 0$, our independent variables must be transformed. This is accomplished by adding the minimum value of a time-series $x_{i,min}$ to each $x_{1,i}$ and uniformly scaling all values. To make sure that the difference between $x_{1,i,min}$ and $x_{1,i,t}$ does not equal 0 when $x_{1,i,t} = x_{1,i,min}$, an ϵ equal to 1 is added to all exponential models. The exponential model becomes:

$$
R_{i,t} = \alpha \cdot x_{1,i,t}^{\beta_{1,i}} \cdot e^{\beta_{2,i} \cdot x_{2,i,t}} \rightarrow R_{i,t} = \alpha \cdot (x_{1,i,t} - x_{1,i,min} + 1)^{\beta_{1,i}} \cdot e^{\beta_{2,i} \cdot x_{2,i,t}}
$$

and the linear transformation becomes:

$$
\begin{aligned}
&R_{log,i,t} = ln(\alpha) + \beta_{1,i} \cdot ln(x_{1,i,t}) + \beta_{2,i} \cdot x_{2,i,t} \\
&\downarrow \\
&R_{log,i,t} = ln(\alpha) + \beta_{1,i} \cdot ln(x_{1,i,t} - x_{1,i,min} + 1) + \beta_{2,i} \cdot x_{2,i,t}
\end{aligned}
$$

Each exponential model is first presented in its exponential form followed by its linear form. In order to ensure robustness of our exponential model results, stationarity of variables must be determined. The transformation of the exponential models will create new variables whose stationarity are confirmed. As the linear transformations of the exponential models will use logarithmic FAANG+M stock returns, the log returns' stationarity are confirmed. As with simple returns, we expect logarithmic returns to be stationary.

The first exponential model uses $ATTN_{i,t-1}$ as $x_{1,i,t}$ and $SENT_{i,t-1}$ as $x_{2,i,t}$ to explore the linear relationship between $R_t + 1$ and $(ATTN_{i,t-1} - ATTN_{i,min}+1)$ and the exponential relationship between $R_{i,t}+1$ and $SENT_{i,t-1}$. When transformed, the relationships of interest are between $R_{log,i,t}$ and $(ATTN_{i,t-1} - ATTN_{i,min} + 1)$ as well as $R_{log,i,t}$ and $SENT_{i,t-1}$.

$$
\begin{aligned}
R_{i,t} + 1 &= \alpha_i \cdot (ATTN_{i,t-1} - ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot SENT_{i,t-1}} \\
R_{log,i,t} &= ln(\alpha_i) + \beta_{A,i} \cdot ln(ATTN_{i,t-1} - ATTN_{i,min} + 1) \\
&\quad + \beta_{S,i} \cdot SENT_{i,t-1}
\end{aligned}
\tag{4.26}
$$

The second exponential model builds on 4.26 where the raw index levels are replaced with unit changes ($\Delta ATTN_{i,t-1}$ and $\Delta SENT_{i,t-1}$). $\Delta ATTN_{i,t-1}$ is the variable $x_{1,i,t}$ and $\Delta SENT_{i,t-1}$ is the variable $x_{2,i,t}$. The second model explores the linear relationship between $R_{i,t}+1$ and $(\Delta ATTN_{i,t-1} - \Delta ATTN_{i,min}+1)$ and the exponential relationship between $R_{i,t}+1$ and $\Delta SENTi, t-1$. We explore the linear relationship between $R_{log,i,t}$ and $(\Delta ATTN_{i,t-1} - \Delta ATTN_{i,min}+1)$ as well as $\Delta SENT_{i,t-1}$.

$$\begin{aligned} R_{i,t}+1 =& \alpha_i \cdot (\Delta ATTN_{i,t-1} - \Delta ATTN_{i,min}+1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot \Delta SENT_{i,t-1}} \\ R_{log,i,t} =& ln(\alpha_i) + \beta_{A,i} \cdot ln(\Delta ATTN_{i,t-1} - \Delta ATTN_{i,min}+1) \\ &+ \beta_{S,i} \cdot \Delta SENT_{i,t-1} \end{aligned} \tag{4.27}$$

The third exponential model builds on 4.26 where the raw index levels are replaced with percentage change ($\%\Delta ATTN_{i,t-1}$ and $\%\Delta SENT_{i,t-1}$). $\%\Delta ATTN_{i,t-1}$ is the variable $x_{1,i,t}$ and $\%\Delta SENT_{i,t-1}$ is the variable $x_{2,i,t}$. The third model explores the linear relationship between $R_{i,t}+1$ and $(\%\Delta ATTN_{i,t-1} - \%\Delta ATTN_{i,min}+1)$ and the exponential relationship between $R_{i,t}+1$ and $\%\Delta SENTi, t-1$. We explore the linear relationship between $R_{log,i,t}$ and $(\%\Delta ATTN_{i,t-1} - \%\Delta ATTN_{i,min}+1)$ as well as $\%\Delta SENT_{i,t-1}$.

$$\begin{aligned} R_{i,t}+1 =& \alpha_i \cdot (\%\Delta ATTN_{i,t-1} - \%\Delta ATTN_{i,min}+1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot \%\Delta SENT_{i,t-1}} \\ R_{log,i,t} =& ln(\alpha_i) + \beta_{A,i} \cdot ln(\%\Delta ATTN_{i,t-1} - \%\Delta ATTN_{i,min}+1) \\ &+ \beta_{S,i} \cdot \%\Delta SENT_{i,t-1} \end{aligned} \tag{4.28}$$

5 Empirical Results

5.1 Summary Statistics

5.1.1 Raw Data

In this section, summary statistics of the 18 raw datasets are briefly mentioned. For the sake of brevity, six stock-specific tables displaying the daily, weekly and monthly summary statistics can be found in the Appendix. FAANG+M stocks exhibit positive mean returns across the daily, weekly, and monthly time intervals in line with our expectations stemming from reasons given in the US Market Environment section.

5.1.2 SENT & ATTN Indices

In table 4, the summary statistics of the firm-specific attention indices are displayed. The standard deviations are exactly equal to 1 as outlined in section 4.5. The skewness tends to be positive across all FAANG+M stocks and time intervals except for GOOGL:Daily, GOOGL:Weekly and AMZN:Daily.

SENT	Ticker	Obs.	Mean	Std	Max	Min	Skew.	Kurt.
Daily	AAPL	1741	-0.0601	1.0	9.2159	-8.5537	0.6935	28.366
	AMZN	1741	0.0099	1.0	10.1999	-7.9618	1.569	30.4282
	GOOGL	1741	0.0112	1.0	9.4584	-9.1035	-0.1838	23.3151
	META	1741	-0.0219	1.0	10.4176	-7.4335	1.1257	33.1439
	MSFT	1741	0.0248	1.0	16.5718	-11.8673	2.5587	64.8582
	NFLX	1741	-0.0244	1.0	8.6923	-7.7973	0.1632	26.4659
Weekly	AAPL	359	0.0294	1.0	4.5004	-3.8991	0.7248	4.9073
	AMZN	359	-0.0273	1.0	4.8436	-4.1849	0.797	6.2813
	GOOGL	359	0.073	1.0	4.1268	-4.1132	-0.2428	3.4626
	META	359	0.0741	1.0	6.4165	-4.2015	1.1016	9.419
	MSFT	359	0.1346	1.0	6.8553	-6.462	0.5946	11.965
	NFLX	359	-0.0097	1.0	4.3108	-3.3158	0.8474	4.8562
Monthly	AAPL	81	0.7784	1.0	2.9049	-2.4824	0.3461	0.522
	AMZN	81	-0.1005	1.0	2.3838	-2.3073	-0.0347	-0.1078
	GOOGL	81	-0.6354	1.0	3.1812	-2.7218	0.1434	1.4281
	META	81	0.0107	1.0	3.4898	-2.9704	0.2478	1.3567
	MSFT	81	-0.4894	1.0	3.5643	-3.4059	0.0322	3.1979
	NFLX	81	-0.025	1.0	1.9274	-1.9242	0.0583	-0.6362

Table 4: Summary statistics of $SENT$ indices for FAANG+M stocks across daily, weekly and monthly time intervals. Mean values are scaled by 10^2.

The value ranges of stock-specific sentiment indices decreases as the time interval widens across all FAANG+M stocks. This is in line with expectations as the BSV factors are more likely to exhibit relatively larger jumps at shorter (i.e. daily) intervals. This is due to the intermittent burst-like nature of activity on Twitter.

At the weekly and monthly intervals, NPC, NNC, NC, TPC, TNC and TC time-series become relatively smoother as daily data is summed to construct weekly and monthly values as outlined in section 4.2.6. Since BSV factors represent four of the six the proxies used to construct the firm-specific sentiment indices, we expect swings in investor sentiment to be largest at the daily interval.

In table 5, the summary statistics of the firm-specific attention indices are displayed. The standard deviations are exactly equal to 1 as outlined in section 4.5. The skewness of $ATTN$ across all time intervals except for AAPL:Weekly is positive, which suggests the distribution is asymmetric.

ATTN	**Ticker**	**Obs.**	**Mean**	**Std**	**Max**	**Min**	**Skew.**	**Kurt.**
Daily	AAPL	1741	0.0012	1.0	8.6223	-7.2228	0.4991	11.0904
	AMZN	1741	-0.0272	1.0	7.4121	-5.4191	1.1899	9.6578
	GOOGL	1741	0.0134	1.0	7.9581	-5.1939	1.0624	8.425
	META	1741	-0.0053	1.0	11.6925	-6.2924	2.2055	29.9789
	MSFT	1741	0.0215	1.0	10.5405	-6.2171	1.0207	13.2542
	NFLX	1741	0.0231	1.0	10.1581	-8.9926	0.9544	23.118
Weekly	AAPL	359	-0.0724	1.0	3.6712	-3.5252	-0.366	1.7739
	AMZN	359	-0.0846	1.0	3.6223	-2.8439	0.5721	0.9681
	GOOGL	359	-0.0128	1.0	3.1968	-2.8973	0.3623	0.5636
	META	359	0.1279	1.0	6.7491	-5.3162	1.4119	15.2032
	MSFT	359	0.1266	1.0	6.081	-4.9637	0.6337	7.768
	NFLX	359	-0.0101	1.0	4.6625	-2.9592	1.2672	3.4181
Monthly	AAPL	81	-0.089	1.0	2.848	-2.3184	0.3636	0.3241
	AMZN	81	0.0382	1.0	2.5628	-2.2084	0.4664	0.3494
	GOOGL	81	-1.1243	1.0	2.672	-2.1997	0.0164	-0.086
	META	81	1.3889	1.0	4.6406	-3.1842	1.5121	8.1588
	MSFT	81	0.03	1.0	3.8983	-3.1805	0.572	3.0827
	NFLX	81	0.6494	1.0	2.5274	-1.9193	0.4987	-0.3776

Table 5: Summary statistics of $ATTN$ indices for FAANG+M stocks across daily, weekly and monthly time intervals. Mean values are scaled by 10^2.

The value ranges of the attention indices appear to be narrower than the sentiment indices across all time intervals. BSV factors represent two of the four proxies used to construct firm-specific attention indices. The remaining proxies (MTR and SVI) have far lower standard deviations and thus result in less extreme swings in the investor attention relative to investor sentiment.

5.2 Sentiment & Attention Loadings

In the following section, the loadings of each sentiment (attention) proxy used in the construction of a firm-specific, composite investor sentiment (attention) index are displayed.

5.2.1 Daily Loadings

Daily $SENT$ Loadings						
Ticker	TPC	TNC	NPC	NNC	PCR	VMP
AAPL	0.42	-0.34	0.38	0.13*	0.02*	0.18*
AMZN	-0.31	0.28	-0.32	0.26	0.07*	-0.21*
GOOGL	-0.37	0.34	-0.38	0.31	0.04	-0.26*
META	0.42	-0.30	0.41	0.09*	-0.03	0.23*
MSFT	-0.39	0.29	-0.42	-0.32*	-0.10*	-0.25*
NFLX	0.36	-0.31	0.36	0.16*	0.03*	0.24*

Table 6: Proxy loadings for the construction of $SENT$ at the daily interval. * indicates lagged proxy selection.

The proxy loadings for $SENT$ are generally mixed at the daily interval. The put-call ratio tends to consistently have the weakest loadings across all FAANG+M stocks, which may indicate that FAANG+M stock options traders (who are generally viewed as more informed than the common equity investor) represent a small portion of aggregate investor sentiment of a given FAANG+M stock. TPC and NPC have stronger loadings than TNC and NNC across all stocks, suggesting financial news and tweets containing positive sentiment are a better gauge of firm-specific sentiment than financial news and tweets containing negative sentiment at the daily interval.

At the daily interval, TPC and TNC have opposite loading signs across all FAANG+M stocks, which is slightly misleading at first glance. Bloomberg assigns negative values to indicate the count of tweets containing negative sentiment. As a result, the TNC loading sign is negative. The final sentiment index is unaffected by this, as the resulting polarity of the value remains the same. This is due to both the value and PCA-loading switching polarity and as the index is a product of the PCA-loading and the value, the index's polarity remains unchanged.

TPC and TNC are strongly positively correlated at the daily interval across all FAANG+M stocks. The total number of positive and negative tweets in a trading day tends to rise and fall together. A proposed explanation is the general speculative discourse that takes place on Twitter and other social media platforms where remarks containing positive speculation are often immediately met by remarks containing negative speculation.

Daily $ATTN$ Loadings				
Ticker	TC	NC	SVI	MTR
AAPL	-0.29	0.29*	-0.45	-0.44
AMZN	0.38*	0.38*	-0.33	-0.27
GOOGL	0.43*	0.45*	-0.30	-0.27
META	0.48*	0.49*	-0.23	-0.19
MSFT	-0.34	0.33*	-0.46	-0.42
NFLX	-0.34	0.25*	-0.41	-0.37

Table 7: Proxy loadings for the construction of $ATTN$ at the daily interval. * indicates lagged proxy selection.

At the daily interval, the lag of NC is selected across all FAANG+M stocks when constructing $ATTN$ suggesting that the number of financial news articles that mention a specific stock tends take longer to reveal the same increase in investor attention as the other proxies. This is in line with expectations as it takes longer to publish financial news than it does to publish a tweet.

NC and TC have the same loading signs when the lag of TC is selected and consistently have opposite loading signs when non-lagged TC is selected. NC and TC are strongly positively correlated for all FAANG+M stocks at the daily interval. One proposed explanation for this observation is the pace of change of Twitter activity. A news article about a stock published today is subsequently followed by an increase in Twitter activity the same day. By tomorrow, the news of today is no longer top of mind as investors on Twitter mention the stock less and less.

5.2.2 Weekly Loadings

Weekly $SENT$ Loadings						
Ticker	TPC	TNC	NPC	NNC	PCR	VMP
AAPL	0.33*	0.21	0.32*	-0.26*	-0.24	-0.16*
AMZN	-0.36	0.3	-0.36	-0.16*	0.11*	-0.19*
GOOGL	-0.22*	-0.41	0.32	-0.41	-0.20*	0.17*
META	-0.35	-0.31*	-0.37	-0.28*	0.06*	-0.16*
MSFT	0.37*	0.29	0.40*	0.41	0.09	-0.19*
NFLX	-0.28	0.27	-0.29	0.25	0.10*	-0.20*

Table 8: Proxy loadings for the construction of $SENT$ at the weekly interval. * indicates lagged proxy selection.

At the weekly interval, the relationship between TPC:TNC and NPC:NNC is similar with the exception of GOOGL. News and social media exhibiting positive sentiment appears to be a better gauge of investor sentiment than news and social media exhibiting negative sentiment. The lag of VMP is selected for all stocks when constructing $SENT$, suggesting that VMP takes longer to reveal the same swing in investor sentiment as other proxies.

BSV proxies (i.e. TPC, TNC, NPC and NNC) appear to have stronger loadings than PCR and VMP across all stocks at the weekly interval. PCR and VMP are seen as gauges of more informed investor sentiment as they are both functions of a FAANG+M stock's options market. It appears that informed and uninformed investor sentiment do not vary contemporaneously at the weekly interval and perhaps one leads the other.

Weekly $ATTN$ Loadings				
Ticker	TC	NC	SVI	MTR
AAPL	0.31	0.34	-0.33*	-0.34*
AMZN	-0.32	-0.34	0.34*	0.34*
GOOGL	-0.34	0.43*	0.39*	-0.31
META	-0.38	-0.40	0.29*	0.29*
MSFT	-0.33	-0.38	0.37*	0.35*
NFLX	0.34*	-0.33	0.34*	-0.30

Table 9: Proxy loadings for the construction of $ATTN$ at the weekly interval. * indicates lagged proxy selection.

The lag of SVI is consistently selected in the construction of $ATTN$ at the weekly interval, suggesting Google Trends data takes longer to reflect the same swing in investor sentiment as other proxies at the weekly interval. The absolute values of the proxy loadings assume a relatively tighter range when compared to the sentiment loadings at the daily interval. A similar relationship between NC and TC observed at the daily interval is also observed at the weekly interval.

At the weekly interval, lagged proxies have the same loading signs across all FAANG+M stocks. The non-lagged proxies tend to have opposite signs, revealing that all proxies of investor attention are positively correlated with one another ultimately reinforcing the previously outlined notion in section 1.2 that investor attention is a two-dimensional concept.

5.2.3 Monthly Loadings

Monthly $SENT$ Loadings						
Ticker	TPC	TNC	NPC	NNC	PCR	VMP
AAPL	0.28*	0.31	-0.24	0.28	-0.18*	-0.32*
AMZN	0.29	-0.21	0.27	-0.23	-0.07	0.27*
GOOGL	-0.37*	-0.28	-0.32*	-0.25	0.25*	0.18*
META	-0.31	0.24	-0.31	0.26	-0.06*	-0.28*
MSFT	0.18	0.32*	0.31	0.31*	0.27	0.27*
NFLX	-0.24	0.25	-0.24	0.21	0.04*	-0.26*

Table 10: Proxy loadings for the construction of $SENT$ at the monthly interval. * indicates lagged proxy selection.

The previously noted relationship between TPC:TNC and NPC:NNC at the daily and weekly interval does not hold at the monthly interval. PCR tends to have the weakest loadings at the monthly interval. VMP tends to have stronger loadings at the monthly interval relative to the weekly interval. Proxy lag selection is generally mixed across FAANG+M stocks and no obviously discernible relationship exists.

Monthly $ATTN$ Loadings				
Ticker	TC	NC	SVI	MTR
AAPL	0.36*	0.40*	-0.31	-0.33
AMZN	-0.43	-0.39	0.32*	-0.33
GOOGL	-0.41*	-0.38*	0.37	0.36
META	0.42*	0.45*	0.24*	-0.25
MSFT	-0.16	0.34*	0.41*	0.40*
NFLX	-0.31	0.34*	0.31*	-0.31

Table 11: Proxy loadings for the construction of $ATTN$ at the monthly interval. * indicates lagged proxy selection.

At the monthly interval, a lag selection of NC occurs for all but one stock suggesting that news publication count takes longer to reflect the same swing in firm-specific investor attention as other proxies. The absolute values of the proxy loadings assume a relatively tighter range when compared to the sentiment loadings at the daily and weekly intervals. Lagged proxy loadings consistently have the same loading signs at the monthly interval, again revealing that our proxies of FAANG+M stock investor attention are positively correlated with one another at the monthly interval.

5.3 Correlation Coefficients

The correlation matrices of the firm-specific sentiment indices are displayed below at the daily, weekly, and monthly time interval. The vast majority of correlation coefficients among FAANG+M stock sentiment indices are statistically significant at the 1% level. There appears to be no obviously discernible relationship between the varying pairwise correlations across different intervals. However, the absolute values of the correlation coefficients tend to steadily increase as the time interval widens, suggesting that the level of differentiation among firm-specific sentiment is greatest at shorter intervals.

SENT	**AAPL**	**AMZN**	**GOOGL**	**META**	**MSFT**	**NFLX**
			Daily			
AAPL	1.0					
AMZN	-0.07***	1.0				
GOOGL	-0.13***	0.33***	1.0			
META	-0.0	-0.06***	-0.11***	1.0		
MSFT	-0.1***	0.23***	0.27***	-0.1***	1.0	
NFLX	0.11***	-0.09***	-0.11***	0.09***	-0.08***	1.0
			Weekly			
AAPL	1.0					
AMZN	0.31***	1.0				
GOOGL	-0.08	-0.29***	1.0			
META	0.14***	0.45***	-0.17***	1.0		
MSFT	0.02	0.08	-0.02	-0.03	1.0	
NFLX	-0.01	-0.3***	-0.05	-0.2***	0.02	1.0
			Monthly			
AAPL	1.0					
AMZN	-0.38***	1.0				
GOOGL	-0.21*	0.32***	1.0			
META	0.37***	-0.56***	-0.23**	1.0		
MSFT	-0.35***	0.47***	0.4***	-0.31***	1.0	
NFLX	0.57***	-0.77***	-0.43***	0.62***	-0.5***	1.0

Table 12: FAANG+M *SENT* correlation matrices at the daily, weekly, monthly time interval. *, **, *** indicates a statistical significance at the 10%, 5% and 1% level respectively.

The empirical results in table 12 suggest that even among similar stocks such as the FAANG+M group, firm-specific investor sentiment does not fluctuate uniformly. These empirical results validate our study's intent of differentiating firm-specific sentiment indices from traditional market-wide sentiment indices[12][32]. A peculiar pattern exists between GOOGL:AMZN and GOOGL:MSFT, where a positive correlation exists at the daily and monthly interval and a negative correlation exists at the weekly time.

The correlation matrices of the firm-specific attention indices are displayed below at the daily, weekly, and monthly interval. At the daily interval, no negative correlations are observed, suggesting that firm-specific investor attention across FAANG+M stocks tends to move together at shorter intervals. At the daily interval, all pairwise correlations are significant at the 1% level. Negative correlations begin to emerge at the weekly and monthly intervals.

ATTN	**AAPL**	**AMZN**	**GOOGL**	**META**	**MSFT**	**NFLX**
			Daily			
AAPL	1.0					
AMZN	0.24***	1.0				
GOOGL	0.22***	0.34***	1.0			
META	0.16***	0.21***	0.22***	1.0		
MSFT	0.29***	0.34***	0.31***	0.15***	1.0	
NFLX	0.13***	0.2***	0.14***	0.13***	0.14***	1.0
			Weekly			
AAPL	1.0					
AMZN	-0.28***	1.0				
GOOGL	-0.2***	0.45***	1.0			
META	-0.16***	0.2***	0.18***	1.0		
MSFT	-0.3***	0.38***	0.25***	0.2***	1.0	
NFLX	0.03	-0.15***	0.07	-0.11***	0.01	1.0
			Monthly			
AAPL	1.0					
AMZN	-0.14	1.0				
GOOGL	-0.1	-0.44***	1.0			
META	0.05	0.09	-0.25**	1.0		
MSFT	0.01	0.2*	-0.06	0.19*	1.0	
NFLX	0.01	0.52***	-0.34***	0.35***	0.28***	1.0

Table 13: FAANG+M *ATTN* correlation matrices at the daily, weekly, monthly time interval. *, **, *** indicates a statistical significance at the 10%, 5% and 1% level respectively.

The empirical results in table 13 suggest that at the daily interval, investor attention for FAANG+M stocks appears to exhibit a cascading effect. An increase in investor attention for one stock is met by a contemporaneous increases in others. A proposed explanation is that many of the FAANG+M firms are direct competitors with one another[49]. Mention of a FAANG+M stock in a financial news article or tweet typically involves mention of another by means of comparison. This relationship does not appear to hold as firmly at the weekly or monthly time interval.

5.3.1 Sentiment, Attention & Returns

The correlation coefficients between $SENT_{t-1}$ and $ATTN_{t-1}$ and current period FAANG+M stock returns are displayed below at the daily, weekly, and monthly intervals. The simple correlation coefficients for $SENT_{t-1}$ reveal little statistical significance for next period FAANG+M stock returns. There appears to be inconsistent incremental changes across all intervals for FAANG+M stocks. Current period returns tend to exhibit stronger absolute correlation coefficients as the interval widens.

	$\boldsymbol{SENT_{t-1}}$			$\boldsymbol{ATTN_{t-1}}$		
Frequency	**Daily**	**Weekly**	**Monthly**	**Daily**	**Weekly**	**Monthly**
AAPL	-0.00	-0.02	0.08	-0.03	0.03	0.17
AMZN	0.00	0.10*	-0.10	-0.01	0.05	-0.05
GOOGL	0.03	-0.09	0.21*	-0.06**	0.08	0.02
META	-0.03	-0.15***	0.02	-0.04	-0.03	0.11
MSFT	0.03	0.1	0.3***	-0.06**	0.06	-0.3***
NFLX	-0.00	0.05	-0.04	-0.02	0.01	-0.09

Table 14: Correlation coefficients of $SENT_{t-1}$ & $ATTN_{t-1}$ with FAANG+M stock returns at the daily, weekly and monthly time intervals. *, **, *** indicates a statistical significance at the 10%, 5% and 1% level respectively.

The correlation coefficients of $ATTN_{t-1}$ with current period returns are consistently negative at the daily time interval, suggesting that increased investor attention tends to imply lower subsequent returns the following trading day. Correlations between $ATTN_{t-1}$ and current period returns tend to be strongest at the monthly interval.

5.3.2 Granger Causality

Performing Granger tests on Returns, $ATTN$ and $SENT$ across all stocks on a daily interval and with a lag of one, reveals some interesting results. First it can be seen that across all stocks the null hypothesis is rejected at all times when $ATTN$ and $SENT$ are picked as either as the explanatory variable(X) or the response variable(Y). This indicates that lagged values of $ATTN$ can explain variations in $SENT$ as well as lagged values of $SENT$

can explain variations in $ATTN$ indicating a causal relationship that goes in both directions. This can be due to the ratios being computed from proxies dependent of each other like TC with TPC and TNC.

Another interesting find is that for $GOOGL$ and $MSFT$ the test results show that lagged Returns can explain variations in both $ATTN$ and $SENT$ and that lagged values of $ATTN$ can explain variations in Returns. This indicates that the stocks that our models will get the best results for will probably be $GOOGL$ and $MSFT$.

	Return(X)	$\boldsymbol{ATTN(X)}$	$\boldsymbol{SENT(X)}$
		AAPL	
Return(Y)	0.0	0.2395	0.9644
$\boldsymbol{ATTN(Y)}$	0.6268	0.0	0.0
$\boldsymbol{SENT(Y)}$	0.0006	0.0	0.0
		AMZN	
Return(Y)	0.1939	0.7528	0.9717
$\boldsymbol{ATTN(Y)}$	0.4978	0.0	0.0
$\boldsymbol{SENT(Y)}$	0.217	0.0	0.0
		GOOGL	
Return(Y)	0.0005	0.0112	0.589
$\boldsymbol{ATTN(Y)}$	0.0	0.0	0.0
$\boldsymbol{SENT(Y)}$	0.0	0.0	0.0
		META	
Return(Y)	0.0091	0.1087	0.2581
$\boldsymbol{ATTN(Y)}$	0.2608	0.0	0.0
$\boldsymbol{SENT(Y)}$	0.0	0.0	0.0
		MSFT	
Return(Y)	0.0	0.0103	0.6518
$\boldsymbol{ATTN(Y)}$	0.0008	0.0	0.0
$\boldsymbol{SENT(Y)}$	0.0	0.0	0.0
		NFLX	
Return(Y)	0.0861	0.3188	0.983
$\boldsymbol{ATTN(Y)}$	0.6749	0.0	0.0
$\boldsymbol{SENT(Y)}$	0.8391	0.0	0.0

Table 15: Granger Causality tests for Returns, $ATTN$ and $SENT$ in the daily interval. P-Values are displayed in the graph where the H_0 is that the explanatory variable(X) does not Granger Cause the response variable(Y)

5.4 Empirical Results: Linear Models

In the following section, empirical results of the exponential regression models outlined in section 4.6.2 are displayed. Each table presents the firm-specific $ATTN$ or $SENT$ index coefficients across all intervals. The test-statistics of coefficients are displayed in circular brackets below.

5.4.1 Stationarity Test Results

The ADF and KPSS tests confirm that $SENT$, $\Delta SENT$ and $\%\Delta SENT$ are stationary across all FAANG+M stocks and time intervals. $\Delta ATTN$ and $\%\Delta ATTN$ are stationary across all FAANG+M stocks and time intervals. The ADF and KPSS tests generally indicate stationarity for $ATTN$ with the exception of two time-series, one showing signs of being trend stationary and other showing signs of being difference stationary.

FAANG+M stock returns (R_t) perform similarly in the ADF and KPSS tests with only one time-series showing signs of being difference stationary. Since $ATTN$ and R_t have very few cases of non-stationarity, we conclude $ATTN$ and R_t are generally stationary. The results of the ADF and KPSS tests for ΔP_t are mixed with many instances being difference stationary and few being non-stationary. We err on the side of caution and conclude ΔP_t is non-stationary. The ADF and KPSS test results for all variables used in our linear regression models can be found in the Appendix.

The results of the stationarity tests indicate that ΔP_t is unfit for use in our linear regressions models. Two of the proposed linear models (equation 4.21 and 4.25) are unfit and dropped from our analysis. Three linear models remain including equation 4.22 which will be referred to as model 1.1, equation 4.23 referred to as model 1.2 and lastly, equation 4.24 referred to as model 1.3. The three linear models are as follows:

Model 1.1: $R_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot ATTN_{i,t-1} + \beta_{i,s} \cdot SENT_{i,t-1} + \epsilon_{i,t}$
Model 1.2: $R_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \Delta SENT_{i,t-1} + \epsilon_{i,t}$
Model 1.3: $R_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \%\Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \%\Delta SENT_{i,t-1} + \epsilon_{i,t}$

5.4.2 Model 1.1

Frequency	Daily		Weekly		Monthly	
Proxy	***ATTN***	***SENT***	***ATTN***	***SENT***	***ATTN***	***SENT***
AAPL	-0.86 (-1.69)*	-0.48 (-0.95)	0.95 (0.38)	-0.27 (-0.11)	13.42 (1.39)	3.41 (0.36)
AMZN	-0.19 (-0.36)	0.14 (0.27)	-0.85 (-0.33)	4.31 (1.67)*	-16.13 (-1.31)	-18.09 (-1.46)
GOOGL	-1.32 (-3.09)***	0.9 (2.11)**	1.38 (0.6)	-2.17 (-0.95)	-5.76 (-0.72)	16.31 (2.03)**
META	-1.05 (-2.11)**	-0.98 (-1.96)**	4.31 (1.61)	-8.57 (-3.2)***	9.08 (0.96)	-1.84 (-0.2)
MSFT	-2.01 (-4.03)***	1.71 (3.42)***	0.28 (0.15)	2.98 (1.54)	-13.67 (-2.26)**	14.11 (2.32)**
NFLX	-1.27 (-1.39)	-1.02 (-1.11)	-3.33 (-0.72)	5.24 (1.13)	-10.95 (-0.66)	4.93 (0.3)

Table 16: Model 1.1 results given by $R_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot ATTN_{i,t-1} + \beta_{i,s} \cdot SENT_{i,t-1} + \epsilon_{i,t}$. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Coefficients are scaled by 10^3. Test statistics are displayed in circular brackets below the respective coefficient.

The results of model 1.1 reveal varying levels of statistical significance across intervals. A waning number of statistically significant $SENT_{t-1}$ and $ATTN_{t-1}$ coefficients are observed as the time interval widens from daily to monthly. The empirical results provide some evidence supporting the hypothesis that investor attention and investor sentiment are significant determinants of short-term FAANG+M stock returns.

Authors Seok et al. note that much of the existing literature analyzing market-wide sentiment makes use of low frequency data such as monthly or annual intervals[13]. The empirical results of model 1.1 suggest that a predictive study of firm-specific investor sentiment and investor attention of FAANG+M stocks is optimal at the shortest interval.

Authors Baker and Wurgler (2006) find that stocks most difficult to value and arbitrage are more sensitive to swings in market-wide investor sentiment. Two significant determinants of a stock's sensitivity to swings in sentiment are volatility of returns and a firm's age[12]. At the daily interval, MSFT's investor sentiment is the most statistically significant across all FAANG+M stocks. However, MSFT's lower standard deviation of daily returns (1.69%) compared to META (1.99%), AAPL (1.82%) and AMZN (1.82%) provides countering evidence to the findings of Baker and Wurgler (2006). The relationships presented in Baker and Wurgler (2006) regarding market-wide sentiment may not be applicable to firm-specific sentiment.

META has the highest standard deviation of daily returns during our study's sample period and is also the youngest of the FAANG+M stocks. The coefficient of META stock's previous day investor sentiment index is -0.98 and is statistically significant at the 5% level. In Baker and Wurgler (2006), the authors find that periods of high investor sentiment are subsequently followed by statistically significant lower returns of stock's most sensitive to investor sentiment[12]. These findings appear to hold true for META's firm-specific sentiment index at the daily interval.

Using Google Trends[51] search frequency data as a proxy for investor attention, authors Da et al. conclude that "an increase in SVI predicts higher stock prices in the next 2 weeks" (Da et al., 2011) among a sample of Russell 3000 stocks from 2004 to 2008[17]. Although comparison of a two week and daily interval is somewhat arbitrary, the empirical results of the composite investor attention indices generally provide countering evidence to the findings of authors Da et al. All coefficients (statistically significant and insignificant) of $ATTN_{t-1}$ at the daily time interval are negative for FAANG+M stocks, suggesting that previous trading days exhibiting high investor attention are subsequently followed by lower returns the next trading day.

5.4.3 Model 1.2

Frequency	**Daily**		**Weekly**		**Monthly**	
Proxy	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$
AAPL	-0.25 (-0.77)	0.03 (0.09)	-0.87 (-0.57)	-1.52 (-1.03)	3.38 (0.58)	3.18 (0.55)
AMZN	-0.1 (-0.31)	0.01 (0.04)	0.58 (0.36)	1.61 (0.98)	-12.0 (-1.59)	-5.87 (-0.79)
GOOGL	-0.71 (-2.61)***	0.45 (1.66)*	0.93 (0.66)	-1.63 (-1.16)	-1.53 (-0.33)	11.69 (2.58)**
META	-0.81 (-2.46)**	-0.58 (-1.77)*	0.83 (0.49)	-2.76 (-1.65)*	-1.91 (-0.26)	1.98 (0.33)
MSFT	-1.36 (-4.32)***	0.94 (3.19)***	1.02 (0.87)	0.66 (0.57)	-9.87 (-2.47)**	9.23 (2.71)***
NFLX	-1.37 (-2.37)**	-1.06 (-1.94)*	-5.22 (-1.78)*	4.94 (1.61)*	-7.7 (-0.81)	1.15 (0.12)

Table 17: Model 1.2 results given by $R_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \Delta SENT_{i,t-1} + \epsilon_{i,t}$. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Coefficients are scaled by 10^3. Test statistics are displayed in circular brackets below the respective coefficient.

Model 1.2 tests the predictive power of the unit change in $SENT_{t-1}$ and $ATTN_{t-1}$ on current period FAANG+M stock returns. $\Delta SENT_{t-1}$ and

$\Delta ATTN_{t-1}$ are a statistically significant determinant of current period returns at the daily interval for GOOGL, META, MSFT and NFLX. Negligible statistical significance of $\Delta SENT_{t-1}$ and $\Delta ATTNt-1$ is observed at the weekly and monthly interval with the exception of MSFT where both $\Delta SENT_{t-1}$ and $\Delta ATTN_{t-1}$ are statistically significant at the 1% and 5% level respectively. However, it is interesting to note that $SENT_{t-1}$ and $ATTN_{t-1}$ for MSFT are highly significant at the daily and monthly interval, yet insignificant at the weekly interval. Compared to model 1.1, the change in predictability is somewhat mixed as model 1.2 appears to perform better for some stocks and worse for others.

5.4.4 Model 1.3

Frequency	**Daily**		**Weekly**		**Monthly**	
Proxy	**%Δ*AT.***	**%Δ*SE.***	**%Δ*AT.***	**%Δ*SE.***	**%Δ*AT.***	**%Δ*SE.***
AAPL	0.0	-0.0	-0.07	-0.18	0.6	-0.23
	(0.2)	(-0.1)	(-0.69)	(-0.91)	(0.99)	(-1.1)
AMZN	-0.0	0.01	-0.12	-0.0	-0.04	-0.36
	(-0.63)	(0.61)	(-0.91)	(-0.01)	(-0.34)	(-1.32)
GOOGL	0.0	0.0	0.0	-0.13	-0.53	0.04
	(0.42)	(0.31)	(0.05)	(-0.96)	(-0.43)	(0.18)
META	0.02	0.01	-0.06	0.03	0.09	0.08
	(1.17)	(1.41)	(-0.94)	(0.2)	(0.4)	(0.14)
MSFT	0.01	-0.01	0.02	-0.0	-1.05	-0.25
	(0.32)	(-1.08)	(0.88)	(-0.02)	(-1.74)*	(-0.17)
NFLX	-0.0	0.01	-0.3	0.12	-0.0	0.29
	(-0.21)	(0.58)	(-0.75)	(0.43)	(-0.0)	(0.96)

Table 18: Model 1.3 results given by $R_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \%\Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \%\Delta SENT_{i,t-1} + \epsilon_{i,t}$. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Coefficients are scaled by 10^3. Test statistics are displayed in circular brackets below the respective coefficient.

Model 1.3 tests the predictive power of the percentage change of both $SENT_{t-1}$ and $ATTN_{t-1}$ on current period FAANG+M stock returns. Virtually no statistical significance is observed. Comparing model 1.3 to models 1.1 and 1.2, we conclude that model 1.3 is the worst fit for the relationships of interests.

5.5 Orthogonal Results: Linear Models

In the following section, the empirical results of the orthogonalized $SENT$ and $ATTN$ indices are presented. As outlined in section 3.4, raw proxies are orthogonalized to US Macroeconomic data and the Fama-French five factor model. We also present the empirical results of the random-walk model approach outlined by Jones and Bandopadhyaya (2008).

5.5.1 US Macroeconomic Data: Linear Results

As outlined in section 3.4, the raw investor sentiment (attention) proxies are orthogonalized to US macroeconomic data in order to isolate the common sentiment (attention) component from the common business cycle component[12]. The orthogonal proxies then undergo the same process outlined in section 4.5 to construct monthly firm-specific investor sentiment (attention) indices orthogonal to US macroeconomic data denoted $SENT_{\perp}$ ($ATTN_{\perp}$).

For the sake of brevity, the ADF and KPSS test results of all regression variables constructed using $SENT_{\perp}$ and $ATTN_{\perp}$ can be found in the Appendix. None of the ADF and KPSS tests suggest the orthogonalized indices are unsuitable for regression analysis. The empirical results of model $1.1_{\perp}$, $1.2_{\perp}$, and $1.3_{\perp}$ using $SENT_{\perp}$ and $ATTN_{\perp}$ at the monthly interval are displayed below.

	Model $1.1_{\perp}$		**Model $1.2_{\perp}$**		**Model $1.3_{\perp}$**	
Proxy	$\boldsymbol{ATTN}$	$\boldsymbol{SENT}$	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$	$\boldsymbol{\%\Delta ATTN}$	$\boldsymbol{\%\Delta SENT}$
AAPL	-1.76 (-0.18)	-9.48 (-0.98)	3.33 (0.58)	-3.78 (-0.67)	0.01 (0.01)	0.12 (0.11)
AMZN	-16.88 (-1.72)*	-3.47 (-0.35)	-7.35 (-1.18)	-3.85 (-0.64)	-0.7 (-1.8)*	0.09 (0.39)
GOOGL	1.57 (0.2)	16.08 (2.03)**	3.98 (0.89)	11.41 (2.45)**	1.13 (0.98)	-0.01 (-0.01)
META	6.4 (0.49)	-8.06 (-0.73)	7.95 (1.02)	1.19 (0.18)	-0.3 (-0.36)	0.07 (0.52)
MSFT	-14.77 (-2.34)**	11.79 (1.88)*	-8.9 (-2.44)**	7.25 (2.05)**	-3.71 (-3.22)***	-0.12 (-0.72)
NFLX	13.28 (0.86)	-1.31 (-0.09)	5.14 (0.57)	2.72 (0.3)	-0.42 (-0.91)	-0.19 (-0.45)

Table 19: Linear regression results of Model $1.1_{\perp}$, $1.2_{\perp}$, and $1.3_{\perp}$ using $SENT_{\perp}$ and $ATTN_{\perp}$ at the monthly interval. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Coefficients are scaled by 10^3. Test statistics are displayed in circular brackets below the respective coefficient.

Isolating the common business cycle component from the individual investor sentiment and attentions proxies produces some interesting findings. Comparing the results of model $1.1_{\perp}$ at the monthly interval, the statistical significance of $ATTN$ for MSFT actually marginally increases, while the statistical significance of $SENT$ decreases. The statistical significance of the results of model $1.2_{\perp}$ remain unchanged after orthogonalizing the raw attention and sentiment proxies to US macroeconomic data.

The empirical results are aligned with the remarks of Baker and Wurgler (2006) where "orthogonalizing to macro variables is a second-order issue" (Baker and Wurgler, 2006), suggesting that our composite investor sentiment and investor attention indices are able to effectively quantify irrational swings in sentiment and attention unexplained by fluctuations in US macroeconomic data. This finding is especially relevant as the FAANG+M stocks accounted for over 23% of S%P 500 market capitalization at the market peak of our sample period[47].

The test statistics of $SENT$:MSFT and $SENT$:GOOGL actually increase following orthogonalization to US macroeconomic data. Comparing the results of model $1.3_{\perp}$ before and after orthogonalizing, the statistical significance increases for a number of stocks, MSFT in particular. $SENT$ remains statistically insignificant after orthogonalizing in model $1.3_{\perp}$.

5.5.2 Fama-French Five Factor Model Results

In the following section, the empirical results of the Fama-French five factor model are displayed at daily and monthly time interval. Please note the factors are not available at the weekly interval. As outlined in section 3.4.2, the raw investor sentiment (attention) proxies are orthogonalized to isolate the variation in FAANG+M stock returns unexplained by the five Fama-French factors (market risk, size, value, profitability and investment[15]).

The orthogonalized proxies then undergo the same process outlined in section 4.5 to construct the daily and monthly firm-specific sentiment (attention) indices denoted $SENT_{FF_5}$ ($ATTN_{FF_5}$). The Fama-French five factor models are similar to models 1.1, 1.2, and 1.3 except the independent variable $R_{i,t}$ is replaced with residuals denoted $y_{FF_{i,j,t}}$. The linear regression models 1.1, 1.2 and 1.3 using $y_{FF_{i,j,t}}$ are denoted 1.1_{FF_5}, 1.2_{FF_5} and 1.3_{FF_5} respectively. The ADF and KPSS tests conclude that all FAANG+M stock residuals denoted $y_{FF_{i,j,t}}$ are stationary across all intervals.

The empirical results of model 1.1_{FF_5} are displayed below at the daily and monthly interval. The observed statistical significance of the orthogonal indices $SENT_{FF_5}$ and $ATTN_{FF_5}$ is reduced when compared to the results of model 1.1. These findings suggest that a significant portion of the variation in stock returns explained by $SENT$ and $ATTN$ can instead be explained by the Fama-French five factor model[15]. These empirical results indicate the Fama-French factors capture some degree of investor sentiment and attention.

Frequency	Daily		Monthly	
Proxy	$\boldsymbol{ATTN}$	$\boldsymbol{SENT}$	$\boldsymbol{ATTN}$	$\boldsymbol{SENT}$
AAPL	-0.16 (-0.32)	-0.17 (-0.34)	10.63 (1.02)	-2.8 (-0.27)
AMZN	0.23 (0.44)	-0.23 (-0.44)	3.93 (0.39)	-10.83 (-1.07)
GOOGL	-1.0 (-2.32)**	0.7 (1.62)	1.52 (0.18)	11.56 (1.39)
META	-0.97 (-1.94)*	-0.88 (-1.76)*	2.73 (0.24)	0.24 (0.02)
MSFT	-0.47 (-1.03)	0.69 (1.5)	-1.77 (-0.28)	17.23 (2.77) ***
NFLX	-0.52 (-0.57)	-0.17 (-0.19)	-0.75 (-0.05)	-7.51 (-0.48)

Table 20: Model 1.1_{FF_5} results using $ATTN_{FF_5}$ and $SENT_{FF_5}$ indices. Coefficients are scaled by 10^3. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Test statistics are displayed in circular brackets below the respective coefficient.

Frequency	Daily		Monthly	
Proxy	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$
AAPL	0.04 (0.12)	0.06 (0.19)	-2.44 (-0.38)	-5.71 (-0.89)
AMZN	0.16 (0.49)	-0.2 (-0.63)	5.69 (0.89)	-3.14 (-0.52)
GOOGL	-0.57 (-2.08) **	0.43 (1.58)	1.73 (0.35)	9.38 (1.97) *
META	-0.78 (-2.36)**	-0.56 (-1.68)*	-3.58 (-0.46)	3.78 (0.52)
MSFT	-0.65 (-2.22) **	0.6 (2.12) **	-4.62 (-1.13)	9.22 (2.65) **
NFLX	-0.89 (-1.57)	-0.57 (-1.05)	-4.73 (-0.52)	-3.5 (-0.38)

Table 21: Model 1.2_{FF_5} results using $\Delta ATTN_{FF_5}$ and $\Delta SENT_{FF_5}$ indices. Coefficients are scaled by 10^3. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Test statistics are displayed in circular brackets below the respective coefficient.

The empirical results of model 1.2_{FF_5} are displayed above at the daily and monthly interval. The statistical significance of $\Delta SENT_{FF_5}$ and $\Delta ATTN_{FF_5}$

is reduced across all FAANG+M stocks following the process of orthogonalizing raw proxies to the Fama-French factors, albeit to a much lesser extent than observed in the results of model 1.1_{FF_5}.

$\Delta ATTN_{FF_5}$ remains statistically significant at the 5% level for GOOGL, META and MSFT. In the case of NFLX, $\Delta ATTN_{FF_5}$ is no longer statistically significant after orthogonalizing to the Fama-French five factor model. This suggests much of the variation in firm-specific investor attention is already captured by the Fama-French five factor model.

Frequency	**Daily**		**Monthly**	
Proxy	**%Δ*ATTN***	**%Δ*SENT***	**%Δ*ATTN***	**%Δ*SENT***
AAPL	0.0 (0.13)	0.0 (0.29)	-2.61 (-1.67)*	0.42 (0.45)
AMZN	-0.01 (-0.61)	-0.01 (-0.51)	0.01 (0.11)	-0.13 (-0.09)
GOOGL	0.01 (0.7)	-0.0 (-0.51)	-0.8 (-1.13)	1.81 (1.11)
META	-0.02 (-1.16)	0.0 (0.62)	0.01 (0.01)	0.85 (1.8)
MSFT	-0.01 (-1.09)	-0.01 (-1.03)	-0.03 (-1.05)	0.15 (0.22)
NFLX	0.0 (0.48)	-0.02 (-1.36)	0.59 (1.11)	-0.46 (-0.21)

Table 22: Model 1.3_{FF_5} results using $\%\Delta ATTN_{FF_5}$ and $\%\Delta SENT_{FF_5}$ indices. Coefficients are scaled by 10^3. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Test statistics are displayed in circular brackets below the respective coefficient.

The empirical results of model 1.3_{FF_5} are displayed above at the daily and monthly interval. Virtually no statistical significance is observed. This is in line with expectations as we conclude that model 1.3 (which investigates the relationship between percentage change in $SENT$ and $ATTN$ and FAANG+M stock returns) is a poor fit. It is interesting to note that $\%\Delta ATTN_{FF_5}$ for AAPL becomes statistically significant at the 10% level after orthogonalizing to the Fama-French five factor model. However, we are careful not to delve to deep into this result due to the poor prior performance of model 1.3.

5.5.3 Random-Walk Model Results

In the following section, the random-walk model results are displayed at the daily, weekly and monthly interval. ADF and KPSS tests prove random-walk model residuals are stationary across all FAANG+M stocks and intervals.

The random-walk models are similar to model 1.1, 1.2 and 1.3 except the independent variable $R_{i,t}$ is replaced with $Residual_{i,t}$. The linear regression models 1.1, 1.2 and 1.3 using $Residual_{i,t}$ are denoted 1.1_{RW}, 1.2_{RW} and 1.3_{RW} respectively.

Varying levels of statistical significance are observed for model 1.1_{RW} and 1.2_{RW} at the daily interval. No statistically significant results are observed in the model 1.3_{RW} panel in line with expectations. These results are similar to those of section 5.4 suggesting that the firm-specific investor sentiment and attention indices are robust across varying models. Authors Jones and Bandopadhyaya (2008) report the VIX and put-call ratio are statistically significant determinants of contemporaneous random-walk model residuals of the S&P 500 index[16]. Our results reinforce the findings of Jones and Bandopadhyaya (2008) at a firm-specific level.

	Model 1.1_{RW}		**Model 1.2_{RW}**		**Model 1.3_{RW}**	
Proxy	***ATTN***	***SENT***	**Δ*ATTN***	**Δ*SENT***	**%Δ*ATTN***	**%Δ*SENT***
AAPL	-107.89	-61.58	-35.72	-9.69	0.28	-1.94
	(-2.77)***	(-1.58)	(-1.44)	(-0.41)	(0.28)	(-0.91)
AMZN	14.54	35.72	1.55	6.28	-0.25	1.52
	(0.29)	(0.71)	(0.05)	(0.2)	(-0.54)	(1.26)
GOOGL	-75.98	55.1	-41.35	32.0	0.24	0.1
	(-2.68)***	(1.94)*	(-2.31)	(1.79)*	(0.33)	(0.44)
META	-241.7	-202.68	-173.36	-98.39	2.41	2.66
	(-2.41)**	(-2.02)	(-2.63)***	(-1.5)	(0.78)	(1.49)
MSFT	-256.75	249.5	-204.38	146.58	2.9	-1.42
	(-3.45)***	(3.35)***	(-4.36)***	(3.36)***	(0.84)	(-1.38)
NFLX	-119.63	-108.46	-245.2	-196.35	-0.01	1.66
	(-0.42)	(-0.38)	(-1.36)	(-1.15)	(-0.11)	(0.22)

Table 23: Random-walk model results at the daily time interval. Model 1.1_{RW} results are given by $Residual_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot ATTN_{i,t-1} + \beta_{i,s} \cdot SENT_{i,t-1} + \epsilon_{i,t}$, model 1.2_{RW} results are given by $Residual_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \Delta SENT_{i,t-1} + \epsilon_{i,t}$, and model 1.3_{RW} results are given by $Residual_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \%\Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \%\Delta SENT_{i,t-1} + \epsilon_{i,t}$. Coefficients are scaled by 10^3. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Test statistics are displayed in circular brackets below the respective coefficient.

The empirical results take the findings of authors Jones and Bandopadhyaya (2008) multiple steps further by using previous period investor sentiment at a firm-specific level instead of current period investor sentiment at a market-wide level. All statistically significant coefficients of $ATTN$ are negative as expected indicating that higher investor attention is subsequently followed by lower FAANG+M stock returns. The statistically significant coefficients of $SENT$ for GOOGL and MSFT provide countering evidence to the findings of authors Jones and Bandopadhyaya (2008), suggesting that at the daily interval, high investor sentiment at a firm-specific level is followed by subsequently higher returns. The empirical results of the random-walk model reinforce the underlying purpose of this study that differentiating firm-specific sentiment from market-wide sentiment is a worthwhile exercise.

The empirical results suggest that the predictive nature of investor sentiment and investor attention go hand-in-hand. For the vast majority of FAANG+M stock $SENT$ and $ATTN$ indices, more often than not both indices are statistically significant. These findings reinforce the rhetoric found in much of the previous literature that investor sentiment and attention are significant determinants of the variation in stock prices unexplained by past values. The weekly and monthly random-walk model results can be found in the Appendix.

5.6 Empirical Results: Exponential Models

In the following section, results of the exponential regression models outlined in section 4.6.3 are displayed. Each table presents the firm and frequency specific $ATTN$ or $SENT$ index coefficients with the respective test-statistics displayed in circular brackets below. The purpose of this section is to achieve a better understanding of whether the relationship between firm-specific investor sentiment (attention) is better modeled in a non-linear form.

5.6.1 Stationarity Test Results

The ADF and KPSS test results for the proposed exponential models generally indicate stationarity for $ln(ATTN - ATTN_{min} + 1)$, $ln(\Delta ATTN - \Delta ATTN_{min} + 1)$, $ln(\%\Delta ATTN - \%_{min} + 1)$ and R_{log} across all FAANG+M stocks and intervals. Very few stationarity tests confirm instances of variables being non-stationary, trend stationary or difference stationary. We conclude all three proposed transformed exponential models are suitable for linear regression analysis. The three models are as follows:

Model 2.1: $R_{i,t} + 1 = \alpha_i \cdot (ATTN_{i,t-1} - ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot SENT_{i,t-1}}$

Model 2.2: $R_{i,t} + 1 = \alpha_i \cdot (\Delta ATTN_{i,t-1} - \Delta ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot$
$\mathrm{e}\hat{\ }\beta_{S,i} \cdot \Delta SENT_{i,t-1}$

Model 2.3: $R_{i,t} + 1 = \alpha_i \cdot (\%\Delta ATTN_{i,t-1} - \%\Delta ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot$
$\mathrm{e}\hat{\ }\beta_{S,i} \cdot \%\Delta SENT_{i,t-1}$

5.6.2 Model 2.1

The empirical results of the linear transformation of equation 4.26 reveal varying levels of statistical significance at the daily interval for a majority of FAANG+M stocks, reconfirming the idea that the relationship between firm-specific investor sentiment and attention is optimally explored at shorter intervals. The firm-specific $ATTN$ indices tend to be more significant than the firm-specific $SENT$ indices for FAANG+M stocks.

Frequency	Daily		Weekly		Monthly	
Proxy	***ATTN***	***SENT***	***ATTN***	***SENT***	***ATTN***	***SENT***
AAPL	-9.33 (-2.57)**	-0.64 (-1.31)	5.05 (0.53)	-0.03 (-0.01)	32.28 (1.1)	4.66 (0.49)
AMZN	0.2 (0.07)	0.05 (0.1)	-0.13 (-0.01)	3.93 (1.61)	-43.54 (-1.23)	-17.0 (-1.43)
GOOGL	-6.94 (-2.8)***	0.81 (1.93)*	6.54 (0.82)	-2.04 (-0.94)	-20.19 (-0.93)	16.42 (2.11)**
META	-5.77 (-1.74)*	-0.88 (-1.79) *	20.51 (1.5)	-7.56 (-3.08)**	50.58 (1.41)	-2.28 (-0.26)
MSFT	-12.01 (-3.8)***	1.45 (3.08) ***	3.35 (0.36)	2.86 (1.57)	-55.97 (-2.56)**	13.3 (2.28)**
NFLX	-1.46 (-0.2)	-0.21 (-0.25)	-2.1 (-0.13)	3.19 (0.77)	-16.57 (-0.39)	0.89 (0.06)

Table 24: Model 2.1 results given by $R_{i,t}+1 = \alpha_i \cdot (ATTN_{i,t-1} - ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot SENT_{i,t-1}}$. Coefficients are scaled by 10^3. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Test statistics are displayed in circular brackets below the respective coefficient.

The coefficients of $ATTN$ at the daily interval are consistently negative across all FAANG+M stocks with the exception of AMZN. Given the form of model 2.1, the negative coefficients of $ATTN$ indicate investor attention is acting as an attenuator of investor sentiment, effectively reducing the magnitude of $SENT$. A proposed explanation is that in periods exhibiting high firm-specific investor attention, the effect of firm-specific investor attention is weakened. Increased investor attention may lead to more rational decision making by FAANG+M investors instead of irrational, sentiment driven decision making. Increased investor attention generally leads to increased market efficiency which may explain why $ATTN$ appears to attenuate the impact of firm-specific investor sentiment. As investors pay closer attention to a FAANG+M stocks, they are more likely to incorporate all readily available information into a trading decision. As a result, markets align more closely to the efficient market hypothesis[11]. In periods of heightened investor attention, FAANG+M stock prices are less likely to exhibit large swings driven by sentiment alone.

The interpretation of $SENT$ coefficients is less clear as signs tend to vary across FAANG+M stocks and intervals. MSFT again reveals high levels of statistical significance at the daily and monthly interval and no statistical significance at the weekly interval. GOOGL appears to exhibit the same tendency as MSFT across intervals, albeit less significant. For the vast majority of FAANG+M stocks, statistical significance of $SENT$ and $ATTN$ appears to go hand in hand. Every occurrence of a statistically significant attention index is paired with a statistically significant sentiment index for all but two observations in table 23.

5.6.3 Model 2.2

The empirical results of the linear transformation of equation 4.27 are displayed below. Less statistical significance is observed in the results of model 2.2, similar observations are found as in the results of model 2.1. The coefficients of $\Delta ATTN$ prove to be significant for half of the FAANG+M stocks and are consistently negative for all statistically significant observations. This further confirms the conclusion that periods of higher firm-specific investor attention are subsequently followed by lower period returns in FAANG+M stocks. Despite observing a few statistically significant $SENT$ and $ATTN$ coefficients, the empirical results of model 1.2 are generally not very good at showcasing the hypothesized non-linear relationship between $R_{i,t}$ and $\Delta ATTN$ or $\Delta SENT$.

Frequency	**Daily**		**Weekly**		**Monthly**	
Proxy	$\Delta ATTN$	$\Delta SENT$	$\Delta ATTN$	$\Delta SENT$	$\Delta ATTN$	$\Delta SENT$
AAPL	-3.99 (-1.36)	-0.03 (-0.11)	-3.22 (-0.36)	-1.3 (-0.94)	10.68 (0.41)	4.13 (0.74)
AMZN	1.04 (0.36)	-0.1 (-0.32)	4.74 (0.56)	1.53 (1.01)	-24.7 (-0.83)	-2.08 (-0.31)
GOOGL	-5.56 (-2.07)**	0.32 (1.25)	4.76 (0.71)	-1.67 (-1.26)	-5.17 (-0.28)	11.15 (2.59)**
META	-7.99 (-1.94)*	-0.44 (-1.39)	2.59 (0.23)	-2.37 (-1.46)	12.53 (0.4)	0.52 (0.09)
MSFT	-14.89 (-3.94)***	0.86 (2.93)***	7.06 (0.82)	0.72 (0.64)	-48.85 (-2.45)**	9.64 (2.97) ***
NFLX	-6.07 (-1.16)	-0.49 (-0.99)	-20.24 (-1.15)	2.63 (1.04)	-21.21 (-0.59)	-0.13 (-0.01)

Table 25: Model 2.2, $R_{i,t}+1 = \alpha_i \cdot (\Delta ATTN_{i,t-1} - \Delta ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot \Delta SENT_{i,t-1}}$. Coefficients are scaled by 10^3. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Test statistics are displayed in circular brackets below the respective coefficient.

5.6.4 Model 2.3

The empirical results of the linear transformation of equation 4.28 are displayed below. Virtually no statistical significance significance is observed across all FAANG+M stocks and intervals with the exception of $ATTN$:GOOGL at the daily interval. In model 2.3, investor sentiment intends to serve as an amplifier of investor sentiment whereby periods of high investor attention lead to amplified swings in firm-specific investor sentiment. This model tests the inverse of the attenuation theory outlined previously. The results reveal that no statistically significant non-linear relationship exists between R_t and $\%\Delta SENT$ where $\%\Delta ATTN$ acts as a possible amplifier. We therefore discard this model as a source of any strong conclusions about the relationship of firm-specific investor sentiment and attention and FAANG+M stock returns.

Frequency	**Daily**		**Weekly**		**Monthly**	
Proxy	$\%\Delta ATTN$	$\%\Delta SENT$	$\%\Delta ATTN$	$\%\Delta SENT$	$\%\Delta ATTN$	$\%\Delta SENT$
AAPL	1.18 (0.45)	-0.0 (-0.09)	-1.2 (-0.18)	-0.18 (-0.9)	12.46 (0.74)	-0.22 (-1.08)
AMZN	-2.42 (-0.99)	0.01 (0.62)	-9.2 (-1.22)	0.0 (0.03)	-10.51 (-0.64)	-0.35 (-1.32)
GOOGL	7.56 (2.82)***	0.0 (0.32)	-5.37 (-0.93)	-0.12 (-0.97)	-2.08 (-0.12)	0.04 (0.16)
META	3.04 (1.09)	0.01 (1.39)	-5.75 (-0.91)	0.02 (0.15)	-3.76 (-0.23)	0.03 (0.06)
MSFT	-0.26 (-0.1)	-0.01 (-1.09)	4.41 (0.86)	-0.0 (-0.01)	-19.09 (-1.51)	0.3 (0.22)
NFLX	0.08 (0.03)	0.01 (0.61)	-8.16 (-0.66)	0.12 (0.44)	8.78 (0.29)	0.27 (0.95)

Table 26: Model 2.3, $R_{i,t} + 1 = \alpha_i \cdot (\%\Delta ATTN_{i,t-1} - \%\Delta ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot \%\Delta SENT_{i,t-1}}$. Coefficients are scaled by 10^3. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Test statistics are displayed in circular brackets below the respective coefficient.

5.7 Orthogonal Results: Exponential Models

The following section briefly presents the empirical results of the exponential models using orthogonalized $SENT$ and $ATTN$ indices. Similar to section 5.5, the $SENT$ and $ATTN$ indices orthogonal to US macroeconomic data (denoted $SENT_{\perp}$ and $ATTN_{\perp}$) become the independent variables. The ADF and KPSS test results confirm stationarity in all variables used in models $2.1_{\perp}$, $2.2_{\perp}$ and $2.3_{\perp}$ and the linear transformations are deemed suitable for regression analysis.

5.7.1 US Macroeconomic Data: Exponential Results

The empirical results of model $2.1_{\perp}$, model $2.2_{\perp}$ and model $2.3_{\perp}$ at the monthly interval are displayed in the table below. Comparing the results of model 2.1 and model model $2.1_{\perp}$, the statistical significance decreases for the majority of $SENT$ and $ATTN$ indices. This is in line to observations made in section 5.5, where the significance of $SENT$ and $ATTN$ indices was reduced following orthogonalization in linear models. This suggests that isolating the common business cycle component lowers the accuracy and validity of our firm-specific investor sentiment and investor attention indices in predicting next month stock returns. The same change in significance as model $2.1_\perp$ can be observed in model $2.2_\perp$ before and after orthogonalizing proxies to US macroeconomic data.

All statistically significant investor attention coefficients are negative in the panels of model $2.1_\perp$ and model $2.2_\perp$, further confirming the idea that investor attention serves to attenuate the magnitude of impact of investor sentiment of FAANG+M stock returns. The empirical results of the orthogonalized exponential models do not provide any more clarity on the nature of $SENT$, $\Delta SENT$ and $\%\Delta SENT$ coefficients. Given our small sample size of stocks, it is difficult to conclude what factors may be driving the changing signs of statistically significant sentiment indices across FAANG+M stocks.

	Model $2.1_\perp$		**Model $2.2_\perp$**		**Model $2.3_\perp$**	
Proxy	$\boldsymbol{ATTN}$	$\boldsymbol{SENT}$	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$	$\boldsymbol{\%\Delta ATTN}$	$\boldsymbol{\%\Delta SENT}$
AAPL	-3.92	-10.1	14.06	-3.68	-4.03	0.17
	(-0.14)	(-1.06)	(0.52)	(-0.66)	(-0.2)	(0.17)
AMZN	-45.09	-1.5	-31.88	-2.56	-31.49	0.07
	(-1.67)*	(-0.16)	(-1.15)	(-0.44)	(-1.66)	(0.31)
GOOGL	-5.94	14.23	4.01	9.79	27.5	0.01
	(-0.28)	(1.84)*	(0.23)	(2.16)**	(1.59)	(0.03)
META	16.99	-7.93	25.41	1.84	-1.98	0.07
	(0.47)	(-0.74)	(0.79)	(0.29)	(-0.1)	(0.56)
MSFT	-36.93	11.07	-25.66	7.4	-35.48	-0.11
	(-1.96)**	(1.78)*	(-1.67)*	(2.11)**	(-2.56)**	(-0.64)
NFLX	14.99	0.9	-7.57	4.99	-15.45	-0.18
	(0.39)	(0.06)	(-0.22)	(0.61)	(-0.72)	(-0.44)

Table 27: Empirical results of Model $2.1_\perp$, Model $2.2_\perp$, Model $2.3_\perp$ at the monthly interval. Coefficients are scaled by 10^3. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Test statistics are displayed in circular brackets below the respective coefficient.

CONTINUE FROM HERE

5.8 Exponential Fama French 5

Once again the three exponential models use ratios produced from orthogonalized data but instead of using proxies regressed over macroeconomical data they make use of proxies regressed over the Fama French 5 Factor Model. This allows for the addition of daily data to the models and another interval compared results in. The versions of the three exponential models will be denoted with a FF_5. Computing the logarithmic values from the ratios computed from the residual values of Fama French 5 Factor Model results in the same prognosis of stationarity as when orthogonalizing over macroeconomical data. The staionarity tests conclude that all three exponential models are fit for linear regressions once transformed logarithmically.

5.8.1 Model 2.1_{FF_5}, Model 2.2_{FF_5} & Model 2.3_{FF_5}

Frequency	**Daily**		**Monthly**	
Proxy	$\boldsymbol{ATTN}$	$\boldsymbol{SENT}$	$\boldsymbol{ATTN}$	$\boldsymbol{SENT}$
AAPL	-4.39 (-1.27)	-0.4 (-0.81)	32.76 (1.04)	-3.03 (-0.3)
AMZN	2.36 (0.77)	-0.25 (-0.51)	5.49 (0.19)	-9.7 (-1.01)
GOOGL	-5.24 (-2.06)**	0.63 (1.5)	7.0 (0.32)	11.7 (1.47)
META	-5.04 (-1.55)	-0.78 (-1.59)	6.19 (0.17)	0.61 (0.06)
MSFT	-2.64 (-0.57)	0.56 (1.3)	-6.68 (-0.35)	15.87 (2.61)**
NFLX	3.39 (0.46)	0.45 (0.54)	-4.79 (-0.12)	-7.28 (-0.5)

Table 28: Regression Results for Model 2.1_{FF_5}, $R_{i,t}+1 = \alpha_i \cdot (ATTN_{i,t-1} - ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot SENT_{i,t-1}}$. Coefficients are scaled by 10^3.

Frequency	**Daily**		**Monthly**	
Proxy	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$
AAPL	-1.53 (-0.5)	-0.03 (-0.1)	-15.29 (-0.48)	-6.29 (-1.01)
AMZN	2.96 (1.03)	-0.28 (-0.9)	25.64 (1.08)	-3.78 (-0.66)
GOOGL	-4.53 (-1.68)*	0.34 (1.29)	10.6 (0.53)	9.57 (2.06)**
META	-7.75 (-1.87)*	-0.42 (-1.32)	3.67 (0.1)	1.23 (0.18)
MSFT	-5.95 (-1.26)	0.43 (1.57)	-15.58 (-1.07)	8.69 (2.56)**
NFLX	-2.92 (-0.55)	-0.13 (-0.27)	-19.28 (-0.53)	-3.36 (-0.42)

Table 29: Regression Results for Model 2.2_{FF_5}, $R_{i,t}+1 = \alpha_i \cdot (\Delta ATTN_{i,t-1} - \Delta ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot \Delta SENT_{i,t-1}}$. Coefficients are scaled by 10^3.

Frequency	**Daily**		**Monthly**	
Proxy	$\boldsymbol{\%\Delta ATTN}$	$\boldsymbol{\%\Delta SENT}$	$\boldsymbol{\%\Delta ATTN}$	$\boldsymbol{\%\Delta SENT}$
AAPL	-2.85 (-1.17)	0.04 (2.74)***	-29.92 (-1.18)	0.28 (1.74)*
AMZN	1.39 (0.54)	-0.0 (-0.29)	-9.94 (-0.76)	0.78 (1.01)
GOOGL	3.96 (1.52)	0.05 (2.63)***	-7.27 (-0.42)	-0.54 (-0.2)
META	-2.64 (-1.0)	0.0 (0.44)	-30.0 (-1.44)	-2.33 (-2.18)**
MSFT	-0.07 (-0.04)	-0.0 (-0.23)	-1.8 (-0.2)	-2.34 (-2.6)**
NFLX	-17.31 (-5.12)***	-0.0 (-0.39)	15.67 (0.81)	0.09 (0.19)

Table 30: Regression Results for Model 2.3_{FF_5}, $R_{i,t}+1 = \alpha_i \cdot (\%\Delta ATTN_{i,t-1} - \%\Delta ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot \%\Delta SENT_{i,t-1}}$. Coefficients are scaled by 10^3.

None of the exponential models under the Fama French 5 Factor model produce any convincing results with very few significant coefficients and consistent stocks. Both *GOOGL* and *MSFT* remains the stocks that produce the most significant results but vary alot across intervals and which variable's coefficient is significant when comparing the different models to each. An interesting change to note when applying the models through Fama French 5 Factor Model is that the sign of the coefficients for $ATTN$, $\Delta ATTN$ and $\%\Delta ATTN$ changes more frequently. The sign become more varied in both the daily and monthly interval switching between being negative or positive across all stocks which is different compared to previous results for the exponential models. However it is hard to confirm and conclusions due to the low frequency of significant results. Hence all three models, Model 2.1_{FF_5}, Model 2.2_{FF_5} & Model 2.3_{FF_5} are discarded from being possible to explain any predictability of future returns by looking at sentiment or attention.

5.9 Put-Call Ratio & Volatility Midpoint: Proxies of Firm-Specific Investor Sentiment

In this section, we verify the findings of authors Jones and Bandopadhyaya (2008) at a firm-specific level for FAANG+M stocks. Jones and Bandopadhyaya (2008) report that both the put-call ratio and the VIX are significantly related to the S&P 500 residuals produced by a random-walk model as outlined in section 3.4.3[16].

In the period spanning January 2, 2004 until April 11, 2006, authors Jones and Bandopadhyaya (2008) report a test statistic of -8.3735 and -5.6157 for the put-call ratio and VIX respectively in relation to the S&P 500 index[16]. Based on their empirical results, the authors conclude that “the put-call ratio is a better measure of such factors than is the VIX and thus the PCR is a better choice as a measure of market sentiment” (Jones and Bandopadhyaya, 2008).

Daily	AAPL	AMZN	GOOGL	META	MSFT	NFLX
TC	0.0	-0.01	0.01	0.01	0.01	0.0
	(0.02)	(-0.28)	(1.58)	(0.23)	(0.37)	(0.03)
TPC	0.04	0.19	0.27	0.51	0.06	1.53
	(0.57)	(1.75)*	(3.37)***	(2.24)**	(0.72)	(2.69)***
TNC	0.07	0.22	0.05	0.83	0.1	2.67
	(1.21)	(1.73)*	(0.56)	(2.97)***	(0.56)	(3.25)***
NC	0.01	-0.01	0.06	0.14	0.15	0.01
	(0.6)	(-0.12)	(1.48)	(0.88)	(1.66)*	(0.03)
NPC	0.39	1.37	4.75	9.32	2.02	15.86
	(0.77)	(1.46)	(5.79)***	(3.09)***	(1.29)	(2.89)***
NNC	1.0	4.48	-0.03	0.91	5.29	33.49
	(1.82)*	(2.53)**	(-0.08)	(0.87)	(1.58)	(4.63)***
VMP	-2.8	-2.6	-2.92	-7.6	-7.12	-5.59
	(-22.16)***	(-15.18)***	(-28.51)***	(-23.07)***	(-35.58)***	(-9.7)***
PCR	-1.75	-2.65	-1.28	-5.58	-1.24	-12.07
	(-8.7)***	(-9.49)***	(-10.48)***	(-9.14)***	(-5.21)***	(-12.68)***
SVI	4.89	3.11	5.18	-6.63	4.14	11.78
	(2.63)***	(3.37)***	(0.84)	(-2.75)***	(3.04)***	(2.29)*
MTR	-4.55	0.08	-0.59	-4.23	-14.71	3.71
	(-3.57)***	(0.06)	(-0.61)	(-2.07)**	(-3.7)***	(2.67)***

Table 31: Empirical results at the daily interval given by $Residual_{i,t} = \alpha_i + \beta_i \cdot proxy_{i,t} + \epsilon_{i,t}$. VMP and MTR coefficients are scaled by 10^{-2}, PCR coefficients are scaled by 1 and the remaining proxy coefficients are scaled by 10^3. *, **, *** indicates statistical significance at the 10%, 5% and 1% level respectively. Test statistics are displayed in circular brackets below the respective coefficient.

The random-walk model results of all individual proxies of firm-specific investor sentiment and attention included in this study are displayed in the table above. The results are generated using an identical approach as authors Jones and Bandopadhyaya (2008) using a random-walk model in the form $Residual_{i,t} = \alpha_i + \beta_i \cdot proxy_{i,t} + \epsilon_{i,t}$. We present similar findings as authors Jones and Bandopadhyaya (2008) at a firm-specific level. As outlined previously, the VMP intends to mirror the utility of the VIX as a gauge of investor sentiment, albeit at a firm-specific level. Both the signs of the coefficients of VMP and PCR come in as expected across all FAANG+M stocks. However, our results contradict the findings of authors Jones and Bandopadhyaya (2008). The implied volatility midpoint of 30-day put and call options appears to be a better gauge of firm-specific sentiment compared to the put-call ratio based on test statistic values. These findings are consistent across all FAANG+M stocks except NFLX.

6 Discussion

6.1 Firm-specific Investor Sentiment

CONTINUE FROM HERE

The empirical results presented throughout section 5 reinforce many of the conclusions found in previous literature. Our firm-specific investor sentiment measures appear to be significant determinants of next period FAANG+M stock returns, particularly at the daily interval. Our firm-specific sentiment indices aggregate multiple proxies of uninformed and informed investor sentiment and prove to possess predictive power in some instances for FAANG+M stocks. This study enriches the literature by exploring the relationship between firm-specific sentiment and a basket of US stock returns.

Authors Seok et al. (2019) find that Korean stock returns increase following periods of high firm-specific investor sentiment "which contrast the findings of the US market" (Seok et al., 2019). Despite the efficiency of US markets and the strong presence of arbitrageurs[12], our results reveal both positive and negative statistically significant coefficients of investor sentiment indices at the daily interval for FAANG+M stocks indicating that the relationship between firm-specific investor sentiment and stock returns isn't as black and white as determining whether or not a strong presence of arbitrageurs exists within a specified market. It is important to mention that our study is limited to a select group of US stocks. In order to determine the factors influencing the signs of our firm-specific sentiment indices coefficients, a more comprehensive analysis of all listed stocks must be undertaken.

As highlighted in the literature review section, stocks that are most difficult to value and most difficult to arbitrage are documented as being the most sensitive to swings in market-wide investor sentiment[12][15]. Authors Baker and Wurgler (2006) identify characteristics such as no earnings, no dividends and young age as indications of how difficult a stock is to arbitrage[12][32]. Within our study's sample period of January 1, 2015 until December 31, 2021, virtually all (with the exception of Facebook where an IPO took place in May of 2012) of the FAANG+M stocks possess decade-long earnings and divided history and relatively nonvolatile returns.

6.2 Firm-Specific Investor Attention

6.3 Our Models

Our models vary in result but show the most significant results in the daily interval suggesting that $SENT$ and $ATTN$ can be effective in predicting future returns on a daily or narrower time interval. These findings agree with the findings of the Bloomberg White paper first mentioned in the introduction[26], that found that intraday trading could use sentiment and attention as stock predictors. At the same time their results showed .... LIAM LOOK AT THIS —¿¿¿

The models that are performing the best, resulting in the most significant results are Model 1.1 and Model 2.1 and their respective versions when using ratios computed from orthogonalized proxies. These models have in common that they look at the raw ratios, $SENT$ and $ATTN$ or in the case of Model 2.1, $ln(ATTN)$. Model 1.2 performs quite well while its exponential opposite, Model 2.2 performs not as well showing quite few significant results. Models 1.3 and 2.3 that uses the variables $\%\Delta ATTN$ and $\%\Delta SENT$ perform very poorly across most stocks and time intervals, indicating that these variables are very poor at helping to predict future stock returns.

When orthogonalizing the ratios, both under macroeconomical data and the Fama French 5 Factor Model, retesting the models generates fewer and less significant results across a majority of intervals. The lack of data for some orthogonalized data makes it difficult to draw conclusions although the daily interval exists for both orthogonalizations which is where our models generates the most significant results. As such it is our most relevant time interval and we can draw some significant conclusions from comparing our different models in the daily interval. The results indicate that when taking into consideration market cycles and market factors, our models does not perform as well. This conclusion can indicate that some of the effect that $ATTN$ and $SENT$ has on future returns is captured within market cycles and factors. Once orthogonalized, a part of the predictability within $ATTN$ and $SENT$ is removed, lessening our models' ability to predict future returns.

6.4 Attention & Sentiment as Predictors

The empirical results for our PCA-loadings show that for $ATTN$ our computed from proxies where the proxies at time t are positively correlated with each other and the lags of our proxies all be positively correlated with each other but the proxies at time t being negatively correlated with the lagged proxies. This For $SENT$ we see something similar as for $ATTN$ but solely for the BSV proxies, keeping in mind that NNC and NTC are input as negative values. However we see that both VMP and PCR show opposite polarity compared to the BSV proxies, portraying a negative correlation. Including proxies negatively correlated with each other into one index might create issues .

Looking at correlation between the computed ratios across all stocks at each time interval we see very significant correlations for $ATTN$ between all stocks and for $SENT$ between all stocks but between $AAPL$ and $META$ at the daily interval. This shows that $ATTN$ and $SENT$ for most stocks move and behave similarly, sort of herding which can be expected as they all operate within the same industry. If something is written about the industry in general it would have an effect on all stocks or if something is published about one company it could also concern the other stocks, affecting them in a similar way. That the ratios across all stocks is significant correlated is expected to be seen as we study companies within the same sector, a sector that is both very new but also has grown a lot in the past years creating a lot of hype for itself and its companies.

Comparing the correlation between our predictors $ATTN$ and $SENT$ and returns by looking at both simple correlations and Granger causality we can draw some interesting conclusions. The results for correlations between next day's returns and $ATTN$ and $SENT$ show a significant correlation at the daily interval only for $ATTN$ when looking at data for $GOOGL$ and $MSFT$. Comparing this to the results of the Granger test which is done solely on daily data, we see that $ATTN$ is only Granger causal with returns for the same two stocks, $GOOGL$ and $AAPL$ strengthening the idea that for $GOOGL$ and $MSFT$ it is possible to use variations of $SENT$ and $ATTN$ to predict future returns. The Granger test also shows that for these two stocks, returns is Granger causal with both $SENT$ and $ATTN$ indicating an opposite relationship where returns can predict future values of $SENT$ and $ATTN$, now looking at returns as the explanatory variable.

Building upon the results of the correlation tests $GOOGL$ and $MSFT$ perform by far the most significant results when performing regression analysis in the daily interval with a variety of models. The very significant results for both $ATTN$ and $SENT$ indicates that there is some predictability for future returns that can be found within $SENT$ and $ATTN$. The difference in results across all stocks also indicates that this predictability could be firm-specific and not be applicable on a general level although that is far from certain as the stocks we have studied are very particular stocks which have all grown tremendously in the past years.

7 Conclusion

7.1 Our Final Results

Our results show that for some stocks $ATTN$ and $SENT$ is significant in models that attempt to predict future returns. These stocks also show strong significant results in correlation tests, showing a predictive power when looking at the variables' ability to explain variations in returns. Out final results show that although there does not seem to be a significant relationship between future returns and $SENT$ or $ATTN$ for a general level, there seems to be such a relationship existing on an individual stock level. We conclude that $ATTN$ and $SENT$ can be used to predict future returns for individual stocks but we cannot conclude that it is possible on a wide market scale.

Our results points toward the daily interval being the most optimal and significant time period, agreeing with the Bloomberg White paper. The Bloomberg White finds that sentiment and attention are most efficient and significant in shorter timespans, which goes in line with our findings that the daily interval produces the most significant results. ——¿ LIAM LOOK AT THIS

The decrease in significance when applying ratios computed from orthogonalized data suggests that the predictability that can be found within $ATTN$ and $SENT$ is absorbed by both market factors and market indicators. This does not necessarily discredit the ability for these variables to predict future returns, but rather indicates that $SENT$ and $ATTN$ alone are not the most effective predictor variables. It additionally lends evidence to the efficient market hypothesis ——¿ LIAM LOOK AT THIS

7.2 Limitations

While conducting our research there were a few limitations we came across, the first being the sample size of our data. As we lacked data for some variables we had to adjust our timeframe of our analysis and hence have a shorter timeframe and smaller sample size than we would want. The data on Twitter(TC, TPC TNC) started from the start of February in 2015 and the data for volatility midpoint(VMP) ended in the end of December 2021. As such our sample timeframe was between the start of February 2015 and the end of December 2021 providing a quite short timeframe to compute our ratios from. Once we had differenced, standardized and applied a PCA to our proxies to create our ratios we were left with even fewer data points. In the end we were left with 81 values for our monthly ratios, 359 for our weekly ratios and 1741 values for our daily ratios which was a smaller sample size then previously wished for.

Another limitation was that we lacked data for some particular timeframes when orthogonalizing our proxies. We could only retrieve macroeconomical data on monthly interval and the Fama French 5 Factor Model could only be retrieved on daily and monthly intervals. As such it was harder to compare the results between the models using ratios not computed from orthogonalized proxies and the models making use of such proxies. This limitation affected our ability to draw conclusions from our results as we could not compare our models across certain timeframes.

Time and space was a limitation as we would have liked to explore more companies, indexes and the market in general. We explored the $FAANG+M$ stocks but would have liked to compute an $FAANG + M$ index to see how our results would have differed. Would the same conclusions be drawn or would we see completely different results as we bundle several companies together? We would have liked to explore the differences between an equally weighted, price weighted and value weighted index to see if the results would differ. Besides creating an index we would have liked to explore more known market indexes like the NYSE or SP500 to see if Returns can be predicted by $SENT$ or $ATTN$ when applied on a wider market scale. Had there been more time and space it would have been possible to explore some or all of these ideas

7.3 Further Research

Our paper only touches a tiny potential of this topic and there are quite a few ways to draw it further building on the analysis and conclusions of this paper. As mentioned in the previous part this paper was limited by time and there is much more within this topic that can be researched and explored. Besides exploring $SENT$ and $ATTN$ for indexes and the market it is possible to explore more individual companies within different sectors than the $FAANG + M$ stocks. Two issues with the stocks we chose are that they have all grown tremendously during the years of our sample size and they are all large very well-known companies. This can skew the results giving incorrect conclusions. Therefore it would be interesting to explore smaller lesser known companies to see if the conclusions are the same or if they change when exploring companies where perhaps the informed institutional investor is less present and more private investors roam.

Another path worthy exploring is narrowing down the time interval for the samples. As our results and conclusions show that the daily interval is most significant it would interesting to explore shorter intervals. For example looking at hourly or minute intervals to see if $ATTN$ and $SENT$ becomes more significant as time becomes shorter. Some papers such as the Bloomberg White paper suggest that $ATTN$ and $SENT$ can be useful to predict future returns within intraday trading but not beyond.

References

[1] Murray, L Lindberg,K. *GoogleTrends*(2023) Copenhagen Business School
https://github.com/karlolofvincentlindberg/MasterThesis/blob/cf0447d141d4d69e6ebb00392473a517e757d1cb/GoogleTrends.py

[2] Murray, L Lindberg,K. *MainDataBase*(2023) Copenhagen Business School
https://github.com/karlolofvincentlindberg/MasterThesis/blob/cf0447d141d4d69e6ebb00392473a517e757d1cb/MainDataBase.py

[3] Durbin, Michael. *All About Frequency Trading* (2010). US: McGraw-Hill.

[4] Barber, Brad M Odean, Terrance *The Behavior of Individual Investors* (2013).Handbook of the Economics of Finance, in: G.M. Constantinides & M. Harris & R. M. Stulz (ed.), Handbook of the Economics of Finance, volume 2, chapter 0, pages 1533-1570, Elsevier

[5] Rasheed, Rayhaan *Why Does Stationarity Matter in Time Series Analysis?* (2020). https://towardsdatascience.com/why-does-stationarity-matter-in-time-series-analysis-e2fb7be74454#:~:text=Stationarity%20is%20an%20important%20concept,is%20independent%20of%20one%20another.

[6] Statsmodels *Why Does Stationarity Matter in Time Series Analysis?* (2020). https://www.statsmodels.org/dev/examples/notebooks/generated/stationarity_detrending_adf_kpss.html

[7] Jaadi Zakaria *A Step-by-Step Explanation of Principal Component Analysis* (2020). https://builtin.com/data-science/step-step-explanation-principal-component-analysis

[8] Linda Smith *A tutorial on Principal Components Analysis* (2002). http://www.iro.umontreal.ca/~pift6080/H09/documents/papers/pca_tutorial.pdf

[9] Holland, S m *PRINCIPAL COMPONENTS ANALYSIS (PCA)* (2019). http://strata.uga.edu/8370/handouts/pcaTutorial.pdf

[10] Jaadi Zakaria *When and Why to Standardize Your Data* (2020). https://builtin.com/data-science/when-and-why-standardize-your-data

[11] Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), 383–417. https://doi.org/10.2307/2325486

[12] Baker, M & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance, Vol. LXI, no. 4, August 2006.

[13] Sang Ik Seok, Hoon Cho, Doojin Ryu. 2019. Firm-specific investor sentiment and daily stock returns. The North American Journal of Economics and Finance, Volume 50. DOI: https://doi.org/10.1016/j.najef.2018.10.005

[14] Baker, M., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. Science Direct, 7(3), Pages 271-299. https://doi.org/10.1016/j.finmar.2003.11.005.

[15] Fama, E. F., French, K. R. (2015). A five factor asset pricing model. Journal of Financial Economics, 116(1), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

[16] Bandopadhyaya, A., & Jones, A. L. (2008). Measures Of Investor Sentiment: A Comparative Analysis Put-Call Ratio Vs. Volatility Index. Journal of Business & Economics Research (JBER), 6(8). https://doi.org/10.19030/jber.v6i8.2458

[17] DA, Z., ENGELBERG, J. and GAO, P. (2011), In Search of Attention. The Journal of Finance, 66: 1461-1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

[18] Kahneman, Daniel., 1973. Attention and Effort (Prentice-Hall, Englewood Cliffs, NJ).

[19] World Federation of Exchanges (2022, May 1). Number of Listed Companies. World Exchanges. Retrieved February 14, 2023, from https://focus.world-exchanges.org/articles/number-listed-companies

[20] Chicago Board Options Exchange (n.d.). VIX Volatility Products. CBOE. Retrieved March 17, 2023, from https://www.cboe.com/tradable_products/vix/

[21] Zhang, Q.T., Li, B., Xie, D. (2022). Sentiment Factors in Finance. In: Alternative Data and Artificial Intelligence Techniques. Palgrave Studies in Risk and Insurance. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-11612-4_9

[22] Hull, John, 1946-. (2012). Options, futures, and other derivatives. Boston :Prentice Hall,

[23] Sharma, R. (2020, May 1). The Comeback Nation. Foreign Affairs. Retrieved February 13, 2023, from https://www.foreignaffairs.com/articles/united-states/2020-03-31/comeback-nation

[24] Board of Governors of the Federal Reserve System (US) (2023, March 1). Federal Funds Effective Rate. FRED. Retrieved February 21, 2023, from https://fred.stlouisfed.org/series/FEDFUNDS

[25] (2019, December 31). As 2010s conclude, investors have enjoyed bull market for the ages — But many Americans have been left out. The Washington Post. Retrieved April 20, 2023, from https://www.washingtonpost.com/business/economy/as-2010s-conclude-investors-have-enjoyed-bull-market-for-the-ages--but-many
2019/12/31/da76a8a0-282e-11ea-ad73-2fd294520e97_story.html

[26] Bloomberg LP (2014, February 20). Embedded Value in Bloomberg News and Social Sentiment Data, Sentiment Analysis White Paper. https://www.Bloomberg.com/. Retrieved January 5, 2023, from https://www.Bloomberg.com/professional/sentiment-analysis-white-papers/

[27] Twitter (2011, March 14). #numbers. Twitter Blog. Retrieved April 2, 2023, from https://blog.twitter.com/official/en_us/a/2011/numbers.html

[28] Wikipedia contributors. (2023, April 8). Quantitative easing. In Wikipedia, The Free Encyclopedia. Retrieved 11:18, April 23, 2023, from https://en.wikipedia.org/w/index.php?title=Quantitative_easing&oldid=1148882756

[29] Persaud, A. (1996) Investors' Changing Appetite for Risk, J.P. Morgan Securities Ltd., Global FX Research.

[30] Bloomberg LP (n.d.). BBMX:IND Bloomberg Massachusetts Index. Bloomberg. Retrieved February 7, 2023, from https://www.bloomberg.com/quote/BBMX:IND#xj4y7vzkg

[31] Bandopadhyaya, A., Jones, A. Measuring investor sentiment in equity markets. J Asset Manag 7, 208–215 (2006). https://doi.org/10.1057/palgrave.jam.2240214

[32] Baker, M & Wurgler, J. (2007). “Investor Sentiment in the Stock Market”. The Journal of Economic Perspectives , Spring, 2007, Vol. 21, No. 2 (Spring,2007), pp. 129-151. American Economic Association https://www.jstor.org/stable/30033721

[33] Qiu, L. X., & Welch, I. (2006). Investor Sentiment Measures. Social Science Research Network. https://doi.org/http://dx.doi.org/10.2139/ssrn.589641.

[34] LEE, C.M.C., SHLEIFER, A. and THALER, R.H. (1991), Investor Sentiment and the Closed-End Fund Puzzle. The Journal of Finance, 46: 75-109. https://doi.org/10.1111/j.1540-6261.1991.tb03746.x

[35] Andersson, K. G. (2000). Lineär Algebra (2nd ed.). Studentlitteratur AB.

[36] Campbell, J. Y., Lo, A. W., & MacKinlay, C. A. (1997). The Econometrics of Financial Markets. Princeton University Press.

[37] Cuthbertson, K., & Nitzsche, D. (2005). Quantitative financial economics: stocks, bonds and foreign exchange. John Wiley & Sons.

[38] Federal Reserve Bank of St. Louis (n.d.). NBER based Recession Indicators for the United States from the Period following the Peak through the Trough. FRED. Retrieved March 31, 2023, from https://fred.stlouisfed.org/series/USREC

[39] Board of Governors of the Federal Reserve System (US) (n.d.). Industrial Production: Durable Consumer Goods. FRED. Retrieved March 31, 2023, from https://fred.stlouisfed.org/series/IPDCONGD

[40] U.S. Bureau of Economic Analysis (n.d.). Personal Consumption Expenditures: Services. FRED. Retrieved March 31, 2023, from https://fred.stlouisfed.org/series/PCES

[41] U.S. Bureau of Economic Analysis (n.d.). Industrial Production: Non-Durable Consumer Goods. FRED. Retrieved March 31, 2023, from https://fred.stlouisfed.org/series/IPNCONGD

[42] Board of Governors of the Federal Reserve System U.S. (n.d.). Industrial Production and Capacity Utilization - G.17. Federal Reserve. Retrieved March 31, 2023, from https://www.federalreserve.gov/releases/g17/current/default.htm

[43] Beer, F., Zouaoui, M. (2012). Measuring Stock Market Investor Sentiment. Journal of Applied Business Research (JABR), 29(1), 51–68. https://doi.org/10.19030/jabr.v29i1.7555

[44] University of Waterloo (n.d.). Topic 6.3: Transformations to Linear Regression. Retrieved March 31, 2023, from https://ece.uwaterloo.ca/~dwharder/NumericalAnalysis/06LeastSquares/transformations/complete.html

[45] Ding, W., Mazouz, K. & Wang, Q. Investor sentiment and the cross-section of stock returns: new theory and evidence. Rev Quant Finan Acc 53, 493–525 (2019). https://doi.org/10.1007/s11156-018-0756-z

[46] Yang, C., & Zhou, L. (2016). Individual stock crowded trades, individual stock investor sentiment and excess returns. The North American Journal of Economics and Finance, 38, 39-53. https://doi.org/10.1016/j.najef.2016.06.001

[47] Bhargava, K. (2021, October 11). Significance of FAANGM: One-fifth of entire US stock market capitalization stored in these 6 shares. Financial Express. Retrieved January 15, 2023, from https://www.financialexpress.com/business/investing-abroad-significance-of-faangm-one-fifth-of-entire-us-stock-market-

[48] Yardeni, E. (2023, April 23). Stock Market Briefing: FAANGMs. Retrieved January 15, 2023, from https://www.yardeni.com/pub/faangms.pdf

[49] Pisal, Shivraj, "RISE OF FACEBOOK, AMAZON, APPLE, NETFLIX, GOOGLE DURING COVID-19 PANDEMIC" (2021). Electronic Theses, Projects, and Dissertations. 1311. https://scholarworks.lib.csusb.edu/etd/1311

[50] Paul, Tanya, "The Effect of Social Media on Trading Behavior: Evidence From Twitter" (2015). Wharton Research Scholars. 131. https://repository.upenn.edu/wharton_research_scholars/131

[51] Google (n.d.). Google Trends. Google Trends. Retrieved December 12, 2022, fromhttps://trends.google.com/home

[52] Barber, Brad Odean, Terrance. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. Review of Financial Studies. 21. 785-818. 10.2139/ssrn.460660.

[53] Holmes, A., Illowsky, B., Dean, S. (2017). Preface. In Introductory Business Statistics. OpenStax. https://openstax.org/books/introductory-business-statistics/pages/preface

[54] Månsson, J., Nordbeck, P. (2011). Endimensionell analys. Studentlitteratur AB.

[55] Fama, E. F. & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3-56. https://doi.org/10.1016/0304-405X(93)90023-5

[56] French, K. R. (n.d.). Current Research Returns. Kenneth R. French - Data Library. Retrieved April 22, 2023, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[57] Paul, Tanya, "The Effect of Social Media on Trading Behavior: Evidence From Twitter" (2015). Wharton Research Scholars. 131. https://repository.upenn.edu/wharton_research_scholars/131

[58] Stephanie Glen. "Granger Causality: Definition, Running the Test" From StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/granger-causality/

A Summary Statistics

AAPL	Proxy	Obs.	Mean	StdDev.	Min	Max	Skewness	Kurtosis
Daily	Return	1744	0.0012	0.0182	-0.1286	0.1198	-0.0868	6.3515
	VMP	1744	0.2692	0.0859	0.1351	0.8752	1.9144	6.4907
	PCR	1744	1.068	0.1256	0.6876	1.3676	-0.359	-0.3149
	MTR	1744	0.0069	0.0032	0.0022	0.0284	1.937	5.4037
	SVI	1744	67.9697	42.2448	22.5806	581.3392	3.3959	21.9075
	TC	1744	4394.621	5834.909	0.0	142267.0	10.1416	195.4196
	TPC	1744	234.676	506.663	0.0	7780.0	6.9968	67.6646
	TNC	1744	255.7924	604.7332	0.0	12621.0	9.7866	147.2361
	NC	1744	2605.996	1323.2332	0	15849	3.973	26.9579
	NPC	1744	36.5831	70.1706	0	900	5.1852	41.6873
	NNC	1744	32.4627	57.456	0	1226	9.4101	147.2947
Weekly	Return	361	0.0057	0.038	-0.1753	0.1473	-0.1714	2.5688
	VMP	361	0.2665	0.0843	0.1389	0.7706	1.7401	5.2031
	PCR	361	1.0599	0.1248	0.7203	1.3356	-0.2941	-0.4142
	MTR	361	0.0333	0.0133	0.0133	0.0921	1.452	2.896
	SVI	361	328.184	177.4889	131.1603	1769.0029	2.8843	14.217
	TC	361	21213.4349	20392.0752	2946	206854	3.3418	20.6436
	TPC	361	1131.3269	1686.141	24	11260	3.091	11.3744
	TNC	361	1234.8283	2071.206	49	23778	5.6562	47.6192
	NC	361	12582.5983	4210.5029	5346	33744	1.8529	5.4051
	NPC	361	176.3019	258.3003	0	1704	2.7641	9.1871
	NNC	361	156.6981	178.1693	0	2114	5.5412	48.6115
Monthly	Return	83	0.0252	0.0815	-0.184	0.2144	-0.1766	-0.3248
	VMP	83	0.2725	0.0811	0.156	0.5457	1.3255	1.802
	PCR	83	1.0618	0.1219	0.7291	1.2866	-0.3239	-0.3276
	MTR	83	0.1432	0.0454	0.0691	0.3571	1.6616	5.4883
	SVI	83	1427.091	684.5412	713.9745	4531.5231	2.3664	6.831
	TC	83	92265.6627	71651.0214	23191	392180	1.5441	3.0374
	TPC	83	4920.5904	5272.7718	229	19275	1.2154	0.5152
	TNC	83	5370.759	6216.1352	386	33989	2.2763	6.8178
	NC	83	54726.7229	11258.0573	32577	84615	0.7564	0.2093
	NPC	83	766.8072	905.72	0	4364	2.1367	4.1037
	NNC	83	681.5422	475.7299	0	3593	3.2339	16.7784

Table 32: Summary statistics - AAPL.

AAPL	Proxy	Obs.	Mean	StdDev.	Min	Max	Skewness	Kurtosis
Daily	Return	1744	0.0015	0.019	-0.0792	0.1413	0.701	7.2301
	VMP	1744	0.3003	0.0897	0.1475	0.6724	0.8353	0.3058
	PCR	1744	0.7848	0.1192	0.5607	1.2087	0.7265	-0.0508
	MTR	1744	0.0086	0.0044	0.0018	0.0514	2.6491	13.0338
	SVI	1744	173.1209	117.6464	16.9436	813.9376	1.2916	2.3271
	TC	1744	2765.988	2423.4841	0.0	23116.0	2.3785	8.0605
	TPC	1744	145.6537	293.1171	0.0	5637.0	10.8685	156.4762
	TNC	1744	127.4719	257.7008	0.0	4899.0	9.7825	130.4275
	NC	1744	736.004	437.1935	0	4459	2.1098	7.9125
	NPC	1744	19.4467	35.5468	0	433	6.3066	50.0226
	NNC	1744	13.0023	20.242	0	200	4.5032	27.2256
Weekly	Return	361	0.007	0.0383	-0.1446	0.1852	0.1925	2.7478
	VMP	361	0.2942	0.0866	0.1478	0.5571	0.7739	0.0161
	PCR	361	0.7899	0.1226	0.5695	1.2087	0.805	0.1388
	MTR	361	0.0412	0.0165	0.0174	0.1076	1.3078	1.7812
	SVI	361	836.0136	488.1212	118.8513	2623.9173	0.6236	0.0807
	TC	361	13349.9612	9550.0784	1835	60933	1.6901	3.1546
	TPC	361	701.1773	723.1914	89	6017	3.7223	18.6366
	TNC	361	613.4155	685.1987	48	5313	3.7688	17.5563
	NC	361	3551.8698	1526.4935	1124	9773	1.0846	1.1263
	NPC	361	93.1994	96.7033	0	584	2.308	5.9829
	NNC	361	62.7645	61.8275	0	475	2.722	10.284
Monthly	Return	83	0.0306	0.082	-0.2022	0.2689	0.4181	1.0823
	VMP	83	0.2943	0.0854	0.1668	0.5147	0.633	-0.4591
	PCR	83	0.7834	0.1162	0.5881	1.1083	0.6717	-0.2081
	MTR	83	0.1793	0.0536	0.1043	0.372	1.3182	1.7214
	SVI	83	3635.7457	1985.7357	571.6205	8667.4603	0.3162	-0.6098
	TC	83	58064.2892	37526.0187	10450	177456	1.4503	1.5822
	TPC	83	3049.6988	1897.1182	835	9229	1.251	1.1863
	TNC	83	2667.988	2164.4365	358	14237	2.8038	10.8296
	NC	83	15448.494	5292.2118	6000	29145	0.5881	-0.5524
	NPC	83	405.3614	282.8439	0	1463	1.3325	2.2542
	NNC	83	272.988	168.45	0	838	1.2768	1.571

Table 33: Summary statistics - AMZN.

GOOGL	Proxy	Obs.	Mean	StdDev.	Min	Max	Skewness	Kurtosis
Daily	Return	1744	0.0011	0.0167	-0.1163	0.1626	0.4351	9.9369
	VMP	1744	0.2464	0.0728	0.1265	0.7304	1.4137	4.3081
	PCR	1744	0.9525	0.0978	0.6936	1.206	-0.1331	-0.5905
	MTR	1744	0.0061	0.0031	0.0015	0.0446	3.081	19.2994
	SVI	1744	72.9649	44.0074	29.1514	328.8889	2.2088	5.3244
	TC	1744	3597.1778	3135.9737	0.0	40774.0	4.4111	33.6768
	TPC	1744	104.7856	265.4023	0.0	6979.0	14.2917	301.0697
	TNC	1744	147.9444	257.8471	0.0	4489.0	7.665	85.7088
	NC	1744	1185.6921	593.4613	0	5010	2.2538	7.9893
	NPC	1744	11.4415	31.4611	0	435	6.0922	49.0371
	NNC	1744	24.5224	53.4752	0	1284	10.4626	193.194
Weekly	Return	361	0.0053	0.035	-0.1203	0.258	1.0286	8.1009
	VMP	361	0.2428	0.0714	0.1265	0.6015	1.1816	2.5431
	PCR	361	0.9478	0.0977	0.7183	1.1714	-0.0467	-0.6268
	MTR	361	0.0296	0.0117	0.0117	0.1009	1.8817	6.126
	SVI	361	352.22	105.4913	203.0563	597.6111	0.6051	-0.7447
	TC	361	17365.3158	12073.5861	2707	88182	2.3347	7.6467
	TPC	361	503.964	733.4414	72	9240	6.3729	60.9911
	TNC	361	712.7673	804.9055	65	7288	4.2653	24.0496
	NC	361	5721.3075	1884.8823	1907	12582	1.0611	1.6027
	NPC	361	55.0249	111.6916	0	839	3.5728	15.1114
	NNC	361	118.3463	164.1079	0	1769	4.5095	32.7939
Monthly	Return	83	0.0225	0.0651	-0.1324	0.2175	0.2757	0.2668
	VMP	83	0.2459	0.0714	0.1265	0.4723	0.7501	0.2255
	PCR	83	0.9538	0.0993	0.7396	1.1417	-0.1901	-0.5643
	MTR	83	5.8938	1.7639	3.3917	12.2573	1.3452	1.8302
	SVI	83	1530.9849	450.7151	957.0761	2579.3889	0.634	-0.776
	TC	83	75528.6627	42914.6372	22805	221857	1.386	1.4051
	TPC	83	2191.9398	1779.2565	578	10929	2.2602	6.7855
	TNC	83	3100.1084	1847.1523	1059	9208	1.4471	1.568
	NC	83	24884.241	4717.7396	16541	36911	0.5247	-0.1706
	NPC	83	239.3253	374.2783	0	1876	2.8662	8.1909
	NNC	83	514.7349	440.13	0	2790	2.0363	7.9428

Table 34: Summary statistics - GOOGL.

META	Proxy	Obs.	Mean	StdDev.	Min	Max	Skewness	Kurtosis
Daily	Return	1744	0.001	0.0199	-0.1896	0.1552	-0.2898	10.8466
	VMP	1744	0.3025	0.0868	0.151	0.8302	1.0976	2.6699
	PCR	1744	0.7197	0.0853	0.5084	0.9667	0.4119	-0.3085
	MTR	1744	0.0093	0.0055	0.0025	0.0705	3.2606	19.0808
	SVI	1744	155.1548	96.1282	70.4394	1425.6342	4.2658	37.5793
	TC	1744	3493.5533	3932.7514	0.0	46434.0	4.2383	27.7072
	TPC	1744	123.8062	364.2583	0.0	7501.0	12.395	195.1717
	TNC	1744	169.8509	371.45	0.0	5631.0	7.9421	83.567
	NC	1744	738.9232	664.6028	0	9187	5.1697	44.0949
	NPC	1744	12.5579	29.1669	0	379	6.3089	52.4659
	NNC	1744	46.8108	109.3346	0	1535	7.997	80.4979
Weekly	Return	361	0.005	0.0398	-0.159	0.2014	-0.3415	3.7606
	VMP	361	0.2983	0.0858	0.151	0.7023	0.9458	1.4572
	PCR	361	0.7205	0.0807	0.5384	0.952	0.443	-0.228
	MTR	361	0.0448	0.0212	0.0159	0.1878	2.0853	7.7113
	SVI	361	749.294	219.0729	496.8071	2829.9739	5.4952	42.4976
	TC	361	16866.3878	15519.501	2523	111435	2.813	10.8035
	TPC	361	597.2715	1058.9199	61	10375	5.7966	40.4392
	TNC	361	820.1357	1389.0086	83	18783	7.727	83.9197
	NC	361	3564.482	2512.2803	778	24516	4.0785	25.2968
	NPC	361	60.3407	93.6114	0	657	3.0528	11.5713
	NNC	361	226.072	418.9864	0	4768	6.9952	61.8504
Monthly	Return	83	0.0208	0.0757	-0.1334	0.2716	0.5514	0.855
	VMP	83	0.2996	0.0769	0.1724	0.5421	0.7607	0.3604
	PCR	83	0.7188	0.0834	0.5605	0.9325	0.4451	-0.0659
	MTR	83	0.1949	0.0649	0.094	0.4131	1.0809	1.225
	SVI	83	3258.1151	776.5282	2454.4	7445.7516	3.2329	13.8921
	TC	83	73358.6265	53627.888	20105	260300	1.4606	1.8581
	TPC	83	2597.7711	2504.5654	415	11954	2.1373	4.6554
	TNC	83	3567.0964	4057.6748	792	30649	4.5524	26.2487
	NC	83	15503.3494	8129.4429	7063	54099	2.4818	8.4281
	NPC	83	262.4458	248.3125	0	1222	1.6121	2.5711
	NNC	83	983.2771	1283.6708	0	8224	4.1047	19.9946

Table 35: Summary statistics - META.

MSFT	**Proxy**	**Obs.**	**Mean**	**StdDev.**	**Min**	**Max**	**Skewness**	**Kurtosis**
Daily	Return	1744	0.0013	0.0169	-0.1474	0.1422	0.2038	10.5318
	VMP	1744	0.2359	0.077	0.1263	0.8579	2.2242	9.9342
	PCR	1744	0.8583	0.1205	0.5237	1.3058	0.3701	0.4368
	MTR	1744	0.0038	0.0018	0.001	0.017	2.4406	9.7308
	SVI	1744	193.3886	155.776	30.0	1486.7777	1.9272	6.3773
	TC	1744	2050.1909	2418.5894	0.0	59136.0	9.4672	187.0733
	TPC	1744	102.2345	589.2366	0.0	22332.0	31.9392	1172.2095
	TNC	1744	80.3842	290.7095	0.0	6962.0	14.1348	255.4007
	NC	1744	1540.5195	593.575	0	7539	2.1371	12.223
	NPC	1744	18.012	38.1547	0	371	4.0733	22.3208
	NNC	1744	7.672	14.7431	0	206	6.4963	62.9225
Weekly	Return	361	0.0064	0.0309	-0.1352	0.1503	-0.0132	3.0774
	VMP	361	0.2319	0.0753	0.1264	0.6714	1.8232	5.62
	PCR	361	0.8604	0.1219	0.5438	1.2778	0.4068	0.455
	MTR	361	0.0185	0.007	0.007	0.0552	1.7294	4.7926
	SVI	361	934.0534	698.5697	208.8571	3888.6867	1.4501	2.3161
	TC	361	9895.2825	8924.8615	938	102692	4.0289	32.95
	TPC	361	493.2964	1685.204	50	30980	16.6381	299.6931
	TNC	361	388.1939	822.6045	25	10178	7.6263	73.0034
	NC	361	7436.6205	2113.3596	2446	23301	1.6505	9.1556
	NPC	361	86.8199	138.1915	0	776	2.1481	4.3122
	NNC	361	37.0305	43.5641	0	298	2.6471	9.6887
Monthly	Return	83	0.0276	0.0601	-0.097	0.1964	0.3643	0.5066
	VMP	83	0.234	0.0744	0.1264	0.5281	1.4956	3.4236
	PCR	83	0.8537	0.1204	0.5621	1.2322	0.4227	0.5835
	MTR	83	0.0804	0.0233	0.0486	0.2113	2.4221	11.0256
	SVI	83	4061.9334	2917.2892	1041.7143	14582.2278	1.26	1.4183
	TC	83	43038.5181	30907.8159	11023	167846	1.3325	1.9951
	TPC	83	2145.5422	3612.8533	506	32915	7.7559	65.9929
	TNC	83	1688.4096	2073.1041	354	13111	3.7708	16.7969
	NC	83	32344.8193	6434.9085	20015	61372	1.4024	4.4926
	NPC	83	377.6145	521.1162	0	2498	2.2027	4.4072
	NNC	83	161.0602	128.0323	0	648	1.368	1.7254

Table 36: Summary statistics - MSFT.

NFLX	Proxy	Obs.	Mean	StdDev.	Min	Max	Skewness	Kurtosis
Daily	Return	1744	0.0016	0.0256	-0.1313	0.1903	0.6548	7.7286
	VMP	1744	0.403	0.1215	0.2135	0.9527	0.9961	1.0497
	PCR	1744	1.1496	0.1316	0.7653	1.529	0.1359	-0.1125
	MTR	1744	0.0222	0.0185	0.0026	0.2463	3.5016	22.9777
	SVI	1744	88.0784	65.5666	11.3333	796.0454	3.2924	18.8721
	TC	1744	1320.9966	1541.0717	0.0	15700.0	4.0284	23.5526
	TPC	1744	104.3108	283.6734	0.0	4934.0	9.8088	119.3384
	TNC	1744	85.9358	206.8672	0.0	3938.0	10.249	145.4874
	NC	1744	558.8205	435.5247	0	3534	2.2611	7.8549
	NPC	1744	12.4065	32.8205	0	495	6.903	59.8348
	NNC	1744	10.172	25.5974	0	343	7.3612	67.0234
Weekly	Return	1758	0.0016	0.0261	-0.1602	0.2573	2.288	26.9451
	VMP	361	0.3987	0.1185	0.2135	0.8431	0.8454	0.4118
	PCR	361	1.137	0.1292	0.7653	1.51	0.171	-0.0683
	MTR	361	0.1066	0.0758	0.02	0.651	2.287	9.0459
	SVI	361	425.4261	265.5097	89.8333	2044.4608	2.304	8.2749
	TC	361	6379.41	5458.2064	763.0	32248.0	1.8962	4.3607
	TPC	361	503.8172	831.0295	39.0	6917.0	4.6447	24.9974
	TNC	361	415.0859	628.4111	29.0	6174.0	5.0169	33.5053
	NC	361	2698.7756	1637.8113	631.0	9298.0	1.1098	1.3594
	NPC	361	59.9003	104.9298	0.0	826.0	3.8698	16.987
	NNC	361	49.1385	83.19	0.0	731.0	4.8589	29.5512
Monthly	Return	83	0.0328	0.1065	-0.1971	0.4081	0.7082	1.7413
	VMP	83	0.3963	0.1178	0.2259	0.8089	0.8317	0.69
	PCR	83	1.1422	0.1327	0.7807	1.51	0.2991	0.259
	MTR	83	0.4609	0.2639	0.1233	1.1926	1.1197	0.7296
	SVI	83	1850.2748	934.2201	477.5	4717.7485	1.2489	1.528
	TC	83	27746.5904	17829.4832	4688	94247	0.9283	1.1551
	TPC	83	2191.3012	2063.9004	253	10611	2.0822	4.5661
	TNC	83	1805.3735	1540.0971	184	7709	1.781	3.285
	NC	83	11738.0482	5914.7775	3459	26836	0.5004	-0.7143
	NPC	83	260.5301	263.8644	0	1368	1.5658	2.6963
	NNC	83	213.7229	201.0061	0	965	2.0451	4.6452

Table 37: Summary statistics - NFLX.

B Linear Regression Models

TO DO: INDICATE THE TIME INTERVAL IN THE TOP LEFT OF THE LINEAR REGRESSION MODEL STATIONARY TESTS.

B.1 Stationarity Results

Daily	**AMZN**	**AAPL**	**META**	**GOOGL**	**MSFT**	**NFLX**
$\Delta P - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta P - KPSS$	Yes	No	Yes	No	No	Yes
$ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 38: ADF & KPSS stationarity test results for daily linear regression variables. "Yes" indicates that the ADF or KPSS test concludes the time-series is stationary.

Weekly	**AMZN**	**AAPL**	**META**	**GOOGL**	**MSFT**	**NFLX**
$\Delta P - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta P - KPSS$	Yes	No	Yes	No	No	Yes
$ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 39: ADF & KPSS stationarity test results for weekly linear regression variables. "Yes" indicates that the ADF or KPSS test concludes the time-series is stationary.

Monthly	**AMZN**	**AAPL**	**META**	**GOOGL**	**MSFT**	**NFLX**
$\Delta P - ADF$	Yes	Yes	Yes	No	No	Yes
$\Delta P - KPSS$	Yes	No	Yes	No	No	Yes
$ATTN - ADF$	Yes	No	Yes	Yes	Yes	No
$ATTN - KPSS$	Yes	Yes	No	Yes	Yes	Yes
$SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - ADF$	Yes	Yes	Yes	No	Yes	Yes
$Return - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 40: ADF & KPSS stationarity test results for monthly linear regression variables. "Yes" indicates that the ADF or KPSS test concludes the time-series is stationary.

C Orthogonalized Data

C.1 Macroeconomic Data - Stationarity Results

Monthly	AMZN	AAPL	META	GOOGL	MSFT	NFLX
$\Delta P - ADF$	Yes	Yes	Yes	No	No	Yes
$\Delta P - KPSS$	Yes	No	Yes	No	No	Yes
$ATTN - ADF$	Yes	No	Yes	Yes	Yes	No
$ATTN - KPSS$	Yes	Yes	No	Yes	Yes	Yes
$SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - ADF$	Yes	Yes	Yes	No	Yes	Yes
$Return - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 41: ADF and KPSS test results for regression variables constructed using $SENT_\perp$ and $ATTN_\perp$ at the monthly time-interval. "Yes" indicates an ADF or KPSS test points toward time-series stationarity. "No" indicates an ADF or KPSS test points toward time-series non-stationarity.

C.2 RW Model - Stationarity Results

	AAPL	AMZN	GOOGL	META	MSFT	NFLX
			Daily			
Residual - ADF	Yes	Yes	Yes	Yes	Yes	Yes
Residual - KPSS	Yes	Yes	Yes	Yes	Yes	Yes
			Weekly			
Residual - ADF	Yes	Yes	Yes	Yes	Yes	Yes
Residual - KPSS	Yes	Yes	Yes	Yes	Yes	Yes
			Monthly			
Residual - ADF	Yes	Yes	Yes	No	Yes	Yes
Residual - KPSS	Yes	Yes	Yes	Yes	Yes	Yes

Table 42: ADF and KPSS test results for random-walk model residuals at all intervals. "Yes" indicates an ADF or KPSS test points toward time-series stationarity. "No" indicates an ADF or KPSS test points toward time-series non-stationarity.

C.3 RW - Regression Results Weekly Interval

	Model 1.1_{RW}		**Model 1.2_{RW}**		**Model 1.3_{RW}**	
Proxy	$\boldsymbol{ATTN}$	$\boldsymbol{SENT}$	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$	$\boldsymbol{\%\Delta ATTN}$	$\boldsymbol{\%\Delta SENT}$
AAPL	208.83	114.3	75.69	10.29	-1.85	-15.29
	(1.14)	(0.63)	(0.67)	(0.09)	(-0.24)	(-1.05)
AMZN	-241.31	300.43	-2.34	79.52	-10.21	-0.4
	(-0.97)	(1.21)	(-0.01)	(0.5)	(-0.78)	(-0.14)
GOOGL	11.01	-127.91	18.08	-112.24	0.34	-5.55
	(0.07)	(-0.87)	(0.2)	(-1.24)	(0.13)	(-0.65)
META	797.97	-1210.63	198.53	-366.59	-5.35	-1.58
	(1.66)*	(-2.51)**	(0.66)	(-1.22)	(-0.47)	(-0.06)
MSFT	-373.95	567.24	-149.96	213.08	5.56	-0.26
	(-1.26)	(1.91)*	(-0.83)	(1.2)	(1.35)	(-0.04)
NFLX	-1276.72	1647.09	-1637.7	1578.31	-39.0	62.51
	(-0.96)	(1.24)	(-1.95)*	(1.79)*	(-0.34)	(0.75)

Table 43: Random-walk model results at the weekly time interval. Model 1.1_{RW} results are generated using $Residual_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot ATTN_{i,t-1} + \beta_{i,s} \cdot SENT_{i,t-1} + \epsilon_{i,t}$, model 1.2_{RW} results are generated using $Residual_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \Delta SENT_{i,t-1} + \epsilon_{i,t}$, and model 1.3_{RW} results are generated using $Residual_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \%\Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \%\Delta SENT_{i,t-1} + \epsilon_{i,t}$. Coefficients are scaled by 10^3.

C.4 RW Model - Regression Results Monthly Interval

	Model 1.1_{RW}		Model 1.2_{RW}		Model 1.3_{RW}	
Proxy	$\boldsymbol{ATTN}$	$\boldsymbol{SENT}$	$\boldsymbol{\Delta ATTN}$	$\boldsymbol{\Delta SENT}$	$\boldsymbol{\%\Delta ATTN}$	$\boldsymbol{\%\Delta SENT}$
AAPL	911.83	42.77	32.41	67.14	66.78	-23.84
	(1.27)	(0.06)	(0.08)	(0.16)	(1.53)	(-1.58)
AMZN	-1855.08	-1605.73	-1155.39	-709.95	-2.15	-33.84
	(-1.7)*	(-1.46)	(-1.72)*	(-1.08)	(-0.22)	(-1.39)
GOOGL	-223.38	1094.12	-94.2	710.32	-64.47	2.37
	(-0.4)	(1.97)*	(-0.29)	(2.25)**	(-0.75)	(0.14)
META	2400.54	-1109.15	126.73	54.14	11.2	5.17
	(1.32)	(-0.63)	(0.09)	(0.05)	(0.26)	(0.05)
MSFT	-2522.5	1270.86	-1874.42	763.38	-120.07	-42.88
	(-2.65)***	(1.33)	(-2.96)***	(1.41)	(-1.27)	(-0.19)
NFLX	-3893.5	2718.37	-2452.21	1497.93	-415.38	56.19
	(-0.86)	(0.6)	(-0.94)	(0.57)	(-0.56)	(0.69)

Table 44: Random-walk model results at the monthly time interval. Model 1.1_{RW} results are generated using $Residual_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot ATTN_{i,t-1} + \beta_{i,s} \cdot SENT_{i,t-1} + \epsilon_{i,t}$, model 1.2_{RW} results are generated using $Residual_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \Delta SENT_{i,t-1} + \epsilon_{i,t}$, and model 1.3_{RW} results are generated using $Residual_{i,t} = \alpha_{i,0} + \beta_{i,a} \cdot \%\Delta ATTN_{i,t-1} + \beta_{i,s} \cdot \%\Delta SENT_{i,t-1} + \epsilon_{i,t}$. Coefficients are scaled by 10^3.

D Fama French Data - Stationarity Results

Daily	AMZN	AAPL	META	GOOGL	MSFT	NFLX
$\Delta P - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta P - KPSS$	Yes	No	Yes	No	No	Yes
$ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 45: ADF and KPSS test results for Fama French model residuals at the daily interval .“Yes” indicates an ADF or KPSS test points toward time-series stationarity. “No” indicates an ADF or KPSS test points toward time-series non-stationarity.

Weekly	AMZN	AAPL	META	GOOGL	MSFT	NFLX
$\Delta P - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta P - KPSS$	Yes	No	Yes	No	No	Yes
$ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 46: ADF and KPSS test results for Fama French model residuals at the weekly intersal ."Yes" indicates an ADF or KPSS test points toward time-series stationarity. "No" indicates an ADF or KPSS test points toward time-series non-stationarity.

Monthly	AMZN	AAPL	META	GOOGL	MSFT	NFLX
$\Delta P - ADF$	Yes	Yes	Yes	No	No	Yes
$\Delta P - KPSS$	Yes	No	Yes	No	No	Yes
$ATTN - ADF$	Yes	No	Yes	Yes	Yes	Yes
$ATTN - KPSS$	Yes	Yes	No	Yes	Yes	Yes
$SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta ATTN - KPSS$	Yes	No	Yes	Yes	Yes	Yes
$\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$Return - ADF$	Yes	Yes	Yes	No	Yes	Yes
$Return - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta ATTN - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$\%\Delta SENT - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 47: ADF and KPSS test results for Fama French model residuals at the weekly interval ."Yes" indicates an ADF or KPSS test points toward time-series stationarity. "No" indicates an ADF or KPSS test points toward time-series non-stationarity.

E Exponential Regression Models

E.1 Stationarity Results

Daily	**AMZN**	**AAPL**	**META**	**GOOGL**	**MSFT**	**NFLX**
$R_{log} - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$R_{log} - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - KPSS$	Yes	Yes	Yes	No	Yes	Yes
$ln(\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - KPSS$	Yes	Yes	Yes	No	Yes	Yes
$ln(\%\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 48: Stationarity for daily log values

Weekly	**AMZN**	**AAPL**	**META**	**GOOGL**	**MSFT**	**NFLX**
$R_{log} - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$R_{log} - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	No
$ln(\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - KPSS$	Yes	No	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 49: Stationarity for weekly log values

Monthly	AMZN	AAPL	META	GOOGL	MSFT	NFLX
$R_{log} - ADF$	Yes	Yes	Yes	No	Yes	Yes
$R_{log} - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - ADF$	Yes	No	Yes	Yes	Yes	Yes
$ln(ATTN) - KPSS$	Yes	No	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - ADF$	Yes	No	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 50: Stationarity for monthly log values

E.2 Stationarity Results - Orthogonalized Data

Monthly	AMZN	AAPL	META	GOOGL	MSFT	NFLX
$R_{log} - ADF$	Yes	Yes	Yes	No	Yes	Yes
$R_{log} - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - ADF$	Yes	No	Yes	Yes	Yes	Yes
$ln(ATTN) - KPSS$	Yes	Yes	Yes	Yes	No	Yes
$ln(\Delta ATTN) - ADF$	Yes	No	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - KPSS$	Yes	No	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 51: Stationarity for daily log values when computed from proxies regressed over macroeconomical data

Daily	AMZN	AAPL	META	GOOGL	MSFT	NFLX
$R_{log} - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$R_{log} - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - KPSS$	Yes	Yes	Yes	No	Yes	Yes
$ln(\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - KPSS$	Yes	Yes	Yes	No	Yes	Yes
$ln(\%\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 52: Stationarity for daily log values when computed from proxies regressed over Fama French 3 Factor Model

Weekly	AMZN	AAPL	META	GOOGL	MSFT	NFLX
$R_{log} - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$R_{log} - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - KPSS$	Yes	No	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 53: Stationarity for weekly log values when computed from proxies regressed over Fama French 3 Factor Model

Monthly	AMZN	AAPL	META	GOOGL	MSFT	NFLX
$R_{log} - ADF$	Yes	Yes	Yes	No	Yes	Yes
$R_{log} - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 54: Stationarity for monthly log values when computed from proxies regressed over Fama French 3 Factor Model

Daily	AMZN	AAPL	META	GOOGL	MSFT	NFLX
$R_{log} - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$R_{log} - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - KPSS$	Yes	Yes	Yes	No	Yes	Yes
$ln(\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - KPSS$	Yes	Yes	Yes	No	Yes	Yes
$ln(\%\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 55: Stationarity for daily log values when computed from proxies regressed over Fama French 5 Factor Model

Monthly	**AMZN**	**AAPL**	**META**	**GOOGL**	**MSFT**	**NFLX**
$R_{log} - ADF$	Yes	Yes	Yes	No	Yes	Yes
$R_{log} - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(ATTN) - KPSS$	No	Yes	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - ADF$	Yes	Yes	Yes	Yes	Yes	Yes
$ln(\%\Delta ATTN) - KPSS$	Yes	Yes	Yes	Yes	Yes	Yes

Table 56: Stationarity for monthly log values when computed from proxies regressed over Fama French 5 Factor Model

CONTINUE FROM HERE

F Fama French 3 Factor Model - Regression Results

Computing the logarithmic values from the ratios computed from the residual values of Fama French 3 Factor Model results in the same prognosis for stationarity as when orthogonalizing over macroeconomical data. All three models that will be denoted with a FF_3 are fit for linear regressions once transformed logarithmically.

F.1 Model 2.1_{FF_3}

Frequency	Daily		Weekly		Monthly	
Proxy	***ATTN***	***SENT***	***ATTN***	***SENT***	***ATTN***	***SENT***
AAPL	-4.39	-0.39	3.79	-0.08	22.9	7.63
	(-1.26)	(-0.79)	(0.37)	(-0.03)	(0.75)	(0.75)
AMZN	2.34	-0.29	2.93	3.33	22.77	-12.49
	(0.77)	(-0.59)	(0.35)	(1.54)	(0.81)	(-1.3)
GOOGL	-5.03	0.58	8.82	-1.36	14.97	12.24
	(-2.02)**	(1.38)	(1.14)	(-0.63)	(0.7)	(1.56)
META	-5.24	-0.78	18.04	-7.8	47.3	-2.19
	(-1.57)	(-1.59)	(1.35)	(-3.19)***	(1.3)	(-0.25)
MSFT	-3.0	0.55	1.49	3.4	12.47	15.14
	(-0.65)	(1.27)	(0.14)	(1.87)*	(0.66)	(2.47)**
NFLX	4.01	0.57	0.31	2.36	-2.35	-7.85
	(0.55)	(0.68)	(0.02)	(0.58)	(-0.06)	(-0.53)

Table 57: Regression Results for Model 2.1_{FF_3}, $R_{i,t} + 1 = \alpha_i \cdot (ATTN_{i,t-1} - ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot SENT_{i,t-1}}$. Coefficients are scaled by 10^3.

F.2 Model 2.2_{FF_3}

Frequency	Daily		Weekly		Monthly	
Proxy	$\Delta ATTN$	$\Delta SENT$	$\Delta ATTN$	$\Delta SENT$	$\Delta ATTN$	$\Delta SENT$
AAPL	-1.28	0.03	-5.14	-1.41	-16.57	7.72
	(-0.41)	(0.09)	(-0.57)	(-1.02)	(-0.54)	(1.26)
AMZN	2.92	-0.31	11.07	1.18	38.74	-4.84
	(1.01)	(-0.98)	(1.23)	(0.84)	(1.58)	(-0.84)
GOOGL	-4.4	0.29	7.45	-1.18	12.07	10.28
	(-1.63)	(1.12)	(1.07)	(-0.9)	(0.65)	(2.26)**
META	-7.67	-0.41	1.52	-2.33	7.38	0.15
	(-1.85)*	(-1.28)	(0.14)	(-1.45)	(0.23)	(0.03)
MSFT	-6.16	0.4	-2.27	1.55	11.84	9.38
	(-1.3)	(1.48)	(-0.25)	(1.35)	(0.58)	(2.61)**
NFLX	-2.49	-0.05	-17.03	2.06	-12.52	-4.17
	(-0.47)	(-0.1)	(-0.97)	(0.82)	(-0.33)	(-0.51)

Table 58: Regression Results for Model 2.2_{FF_3}, $R_{i,t}+1 = \alpha_i \cdot (\Delta ATTN_{i,t-1} - \Delta ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot \Delta SENT_{i,t-1}}$. Coefficients are scaled by 10^3.

F.3 Model 2.3_{FF_3}

Frequency	Daily		Weekly		Monthly	
Proxy	$\%\Delta ATTN$	$\%\Delta SENT$	$\%\Delta ATTN$	$\%\Delta SENT$	$\%\Delta ATTN$	$\%\Delta SENT$
AAPL	-2.85	0.04	0.61	-0.01	-29.92	0.28
	(-1.17)	(2.74)***	(0.08)	(-0.02)	(-1.18)	(1.74)*
AMZN	1.39	-0.0	7.97	-0.09	-9.94	0.78
	(0.54)	(-0.29)	(1.05)	(-0.82)	(-0.76)	(1.01)
GOOGL	3.96	0.05	3.02	0.03	-7.27	-0.54
	(1.52)	(2.63)*	(0.47)	(0.87)	(-0.42)	(-0.2)
META	-2.64	0.0	-1.06	-0.27	-30.0	-2.33
	(-1.0)	(0.44)	(-0.23)	(-1.07)	(-1.44)	(-2.18)**
MSFT	-0.07	-0.0	9.98	0.02	-1.8	-2.34
	(-0.04)	(-0.23)	(1.41)	(1.27)	(-0.2)	(-2.6)**
NFLX	-17.31	-0.0	3.78	-0.01	15.67	0.09
	(-5.12)***	(-0.39)	(0.44)	(-0.1)	(0.81)	(0.19)

Table 59: Regression Results for Model 2.3_{FF_3}, $R_{i,t}+1 = \alpha_i \cdot (\%\Delta ATTN_{i,t-1} - \%\Delta ATTN_{i,min} + 1)^{\beta_{A,i}} \cdot e^{\beta_{S,i} \cdot \%\Delta SENT_{i,t-1}}$. Coefficients are scaled by 10^3.

G Summary statistics for proxy regressions on residual of price

G.1 $Res = \alpha + \beta \cdot proxy_t$

Daily	AAPL	AMZN	GOOGL	META	MSFT	NFLX
TC	0.0	-0.01	0.01	0.01	0.01	0.0
	(0.02)	(-0.28)	(1.58)	(0.23)	(0.37)	(0.03)
TPC	0.04	0.19	0.27	0.51	0.06	1.53
	(0.57)	(1.75)*	(3.37)***	(2.24)**	(0.72)	(2.69)***
TNC	0.07	0.22	0.05	0.83	0.1	2.67
	(1.21)	(1.73)*	(0.56)	(2.97)***	(0.56)	(3.25)***
NC	0.01	-0.01	0.06	0.14	0.15	0.01
	(0.6)	(-0.12)	(1.48)	(0.88)	(1.66)*	(0.03)
NPC	0.39	1.37	4.75	9.32	2.02	15.86
	(0.77)	(1.46)	(5.79)***	(3.09)***	(1.29)	(2.89)***
NNC	1.0	4.48	-0.03	0.91	5.29	33.49
	(1.82)*	(2.53)**	(-0.08)	(0.87)	(1.58)	(4.63)***
VMP	-2.8	-2.6	-2.92	-7.6	-7.12	-5.59
	(-22.16)***	(-15.18)***	(-28.51)***	(-23.07)***	(-35.58)***	(-9.7)***
PCR	-1.75	-2.65	-1.28	-5.58	-1.24	-12.07
	(-8.7)***	(-9.49)***	(-10.48)***	(-9.14)***	(-5.21)***	(-12.68)***
SVI	4.89	3.11	5.18	-6.63	4.14	11.78
	(2.63)***	(3.37)***	(0.84)	(-2.75)***	(3.04)***	(2.29)*
MTR	-4.55	0.08	-0.59	-4.23	-14.71	3.71
	(-3.57)***	(0.06)	(-0.61)	(-2.07)**	(-3.7)***	(2.67)***

Table 60: $Res = \alpha + \beta \cdot proxy_t$ - Daily. Coefficents for VMP and MTR are scaled by 10^{-2}. Coefficient for PCR is scaled by 1. Remaining coefficients are scaled by 10^3.

Weekly	AAPL	AMZN	GOOGL	META	MSFT	NFLX
TC	-0.0	0.02	0.0	-0.08	0.01	0.15
	(-0.38)	(0.55)	(0.06)	(-2.98)	(0.2)	(0.99)
TPC	0.06	0.44	0.07	-0.48	0.03	2.11
	(0.75)	(2.06)	(0.54)	(-1.78)	(0.31)	(2.74)
TNC	0.04	0.34	0.12	0.87	0.15	1.98
	(0.66)	(1.31)	(1.07)	(3.61)	(0.65)	(1.89)
NC	0.01	0.17	0.0	-0.5	-0.02	0.54
	(0.3)	(1.24)	(0.06)	(-3.44)	(-0.16)	(0.89)
NPC	0.48	4.81	2.86	-6.84	0.19	23.35
	(0.88)	(2.86)	(2.77)	(-2.22)	(0.09)	(3.77)
NNC	0.66	4.57	0.44	2.97	3.4	24.08
	(1.05)	(1.67)	(0.76)	(3.39)	(0.78)	(3.04)
VMP	-2.44	-1.59	-3.04	-2.92	-6.15	-2.23
	(-7.23)	(-3.86)	(-11.3)	(-3.96)	(-10.4)	(-1.71)
PCR	-6.43	-12.39	-4.96	-16.79	-5.42	-39.96
	(-5.36)	(-6.94)	(-7.05)	(-4.92)	(-3.55)	(-7.06)
SVI	1.04	1.14	-4.72	-13.12	-0.61	2.14
	(0.81)	(1.44)	(-0.93)	(-5.29)	(-0.71)	(0.59)
MTR	-1.53	-0.69	-1.1	-12.69	-12.27	2.11
	(-1.31)	(-0.56)	(-1.18)	(-7.74)	(-3.43)	(1.69)

Table 61: $Res = \alpha + \beta \cdot proxy_t$ - Weekly. Coefficents for VMP and MTR are scaled by 10^{-2}. Coefficient for PCR is scaled by 1. Remaining coefficients are scaled by 10^3.

Monthly	**AAPL**	**AMZN**	**GOOGL**	**META**	**MSFT**	**NFLX**
TC	0.0	-0.01	-0.0	-0.06	0.01	-0.13
	(0.08)	(-0.28)	(-0.15)	(-1.69)	(0.14)	(-0.72)
TPC	0.16	0.42	0.18	0.23	-0.01	1.15
	(1.04)	(0.99)	(0.76)	(0.44)	(-0.05)	(1.1)
TNC	0.07	0.22	0.35	0.88	-0.13	1.59
	(0.55)	(0.64)	(1.41)	(2.44)	(-0.36)	(1.05)
NC	-0.02	0.18	0.01	-0.36	0.21	0.1
	(-0.34)	(0.78)	(0.13)	(-1.6)	(1.3)	(0.13)
NPC	1.93	3.53	3.4	10.76	4.45	10.74
	(1.84)	(1.48)	(1.97)	(1.67)	(1.84)	(1.43)
NNC	1.54	1.5	1.62	2.93	8.91	20.67
	(1.42)	(0.37)	(1.67)	(2.13)	(1.4)	(1.84)
VMP	-0.46	-0.9	-2.6	-6.77	-4.65	-1.91
	(-0.49)	(-1.15)	(-4.4)	(-3.65)	(-3.9)	(-0.96)
PCR	-28.8	-27.87	-14.7	-80.9	-18.75	-74.42
	(-5.52)	(-3.75)	(-3.78)	(-6.07)	(-2.82)	(-2.97)
SVI	2.52	0.19	-4.67	-5.64	0.32	0.68
	(1.82)	(0.23)	(-1.08)	(-1.88)	(0.38)	(0.19)
MTR	-3.79	-2.58	-4.18	-5.43	-11.78	-1.02
	(-2.94)	(-1.68)	(-3.04)	(-2.09)	(-2.87)	(-0.68)

Table 62: $Res = \alpha + \beta \cdot proxy_t$ - Monthly. Coefficents for VMP and MTR are scaled by 10^{-2}. Coefficient for PCR is scaled by 1. Remaining coefficients are scaled by 10^3.

G.2 $\boldsymbol{Res = \alpha + \beta \cdot proxy_{t-1}}$

Daily	**AAPL**	**AMZN**	**GOOGL**	**META**	**MSFT**	**NFLX**
TC	-0.0	-0.01	0.0	-0.02	-0.0	-0.05
	(-0.63)	(-0.45)	(0.22)	(-0.62)	(-0.06)	(-0.4)
TPC	0.01	-0.02	0.02	-0.07	-0.01	-0.1
	(0.17)	(-0.17)	(0.21)	(-0.31)	(-0.15)	(-0.17)
TNC	0.02	0.06	0.01	0.16	-0.07	-0.05
	(0.34)	(0.48)	(0.15)	(0.56)	(-0.41)	(-0.06)
NC	0.01	-0.07	0.04	-0.12	-0.12	0.03
	(0.52)	(-0.77)	(0.94)	(-0.77)	(-1.27)	(0.06)
NPC	0.44	-0.96	0.17	-3.81	0.51	1.82
	(0.87)	(-1.02)	(0.2)	(-1.26)	(0.33)	(0.33)
NNC	-0.1	0.29	0.48	1.06	2.19	-1.54
	(-0.18)	(0.16)	(1.11)	(1.02)	(0.65)	(-0.21)
VMP	0.74	0.51	0.59	1.32	2.43	1.54
	(5.22)	(2.78)	(4.79)	(3.51)	(9.48)	(2.6)
PCR	0.05	-0.54	0.01	0.49	0.01	1.25
	(0.25)	(-1.89)	(0.11)	(0.79)	(0.03)	(1.26)
SVI	7.35	-1.16	19.58	3.4	4.25	0.65
	(3.96)	(-1.25)	(3.17)	(1.41)	(3.12)	(0.13)
MTR	1.08	-1.04	1.35	0.14	4.51	0.35
	(0.85)	(-0.87)	(1.4)	(0.07)	(1.13)	(0.25)

Table 63: $Res = \alpha + \beta \cdot proxy_{t-1}$- Daily. Coefficents for VMP and MTR are scaled by 10^{-2}. Coefficient for PCR is scaled by 1. Remaining coefficients are scaled by 10^{3}.

Weekly	AAPL	AMZN	GOOGL	META	MSFT	NFLX
TC	0.0	-0.02	-0.0	0.02	-0.01	-0.11
	(0.31)	(-0.68)	(-0.27)	(0.79)	(-0.38)	(-0.72)
TPC	0.02	-0.14	-0.03	0.41	-0.02	-0.32
	(0.31)	(-0.64)	(-0.27)	(1.53)	(-0.14)	(-0.41)
TNC	0.03	0.07	0.13	0.33	-0.01	0.4
	(0.52)	(0.27)	(1.17)	(1.33)	(-0.03)	(0.38)
NC	-0.0	-0.08	-0.03	0.04	-0.03	-0.36
	(-0.12)	(-0.62)	(-0.66)	(0.24)	(-0.24)	(-0.59)
NPC	0.22	-1.11	0.14	8.41	3.66	-7.29
	(0.41)	(-0.65)	(0.14)	(2.73)	(1.72)	(-1.16)
NNC	0.68	0.49	0.72	0.4	12.43	-1.32
	(1.08)	(0.18)	(1.24)	(0.45)	(2.88)	(-0.16)
VMP	0.04	0.03	0.35	-4.73	0.17	1.91
	(0.12)	(0.07)	(1.12)	(-6.63)	(0.25)	(1.46)
PCR	0.48	0.04	1.48	-14.42	-0.08	-6.82
	(0.39)	(0.02)	(1.98)	(-4.19)	(-0.05)	(-1.13)
SVI	-0.64	-0.26	-1.45	4.36	0.28	-6.08
	(-0.5)	(-0.33)	(-0.29)	(1.7)	(0.32)	(-1.67)
MTR	-0.32	-0.66	-0.76	0.96	-0.77	-0.8
	(-0.27)	(-0.54)	(-0.81)	(0.54)	(-0.21)	(-0.64)

Table 64: $Res = \alpha + \beta \cdot proxy_{t-1}$ - Weekly. Coefficents for VMP and MTR are scaled by 10^{-2}. Coefficient for PCR is scaled by 1. Remaining coefficients are scaled by 10^3.

Monthly	**AAPL**	**AMZN**	**GOOGL**	**META**	**MSFT**	**NFLX**
TC	-0.01	0.01	-0.01	0.01	0.01	-0.04
	(-0.67)	(0.2)	(-0.53)	(0.32)	(0.28)	(-0.2)
TPC	-0.05	0.09	0.09	-0.28	0.06	-0.17
	(-0.36)	(0.22)	(0.38)	(-0.53)	(0.28)	(-0.16)
TNC	0.06	0.02	-0.05	-0.02	0.04	0.12
	(0.46)	(0.07)	(-0.21)	(-0.05)	(0.11)	(0.08)
NC	-0.04	-0.3	0.07	0.26	0.09	-0.06
	(-0.79)	(-1.32)	(0.72)	(1.14)	(0.54)	(-0.08)
NPC	0.05	-1.79	0.4	-1.21	1.26	1.45
	(0.05)	(-0.74)	(0.23)	(-0.18)	(0.51)	(0.19)
NNC	0.28	4.34	-2.23	-0.91	-10.73	6.44
	(0.26)	(1.06)	(-2.32)	(-0.64)	(-1.68)	(0.56)
VMP	-1.66	0.04	0.4	-0.42	0.89	-0.82
	(-1.78)	(0.05)	(0.6)	(-0.21)	(0.69)	(-0.41)
PCR	-1.75	10.52	0.73	18.19	17.17	25.2
	(-0.28)	(1.29)	(0.17)	(1.13)	(2.54)	(0.94)
SVI	-3.46	1.88	-3.51	-0.39	-0.19	1.35
	(-2.52)	(2.34)	(-0.81)	(-0.11)	(-0.23)	(0.38)
MTR	-0.53	2.66	2.31	0.15	6.0	0.61
	(-0.38)	(1.72)	(1.61)	(0.06)	(1.4)	(0.4)

Table 65: $Res = \alpha + \beta \cdot proxy_{t-1}$ - Monthly. Coefficents for VMP and MTR are scaled by 10^{-2}. Coefficient for PCR is scaled by 1. Remaining coefficients are scaled by 10^3.