A Work Project, presented as part of the requirements for the Award of a Master's degree in Finance from the Nova School of Business and Economics.

THE ROLE OF SOCIAL MEDIA IN THE STOCK MARKET: TWITTER SENTIMENT AS A PREDICTOR OF STOCK RETURNS

DAVIDE GOMES COSTA

Work project carried out under the supervision of:

A. Emanuele Rizzo

17-12-2021

Abstract

The recent surge of emerging technologies, combined with the growth of social media securities-related microblogging, instigated academics to explore new proxies for sentiment. This research dissects the association between 1-month lagged Twitter sentiment and stock returns for the S&P500 constituents from 2008 to 2021 through the sentiment analysis of approximately 34.7 million tweets. Evidence shows a consistent variation pattern of returns across the scope of the anomaly. Furthermore, abnormal returns associated with high Twitter sentiment are pervasive and significant, particularly for value-weight returns. In contrast, there is insufficient evidence on the pervasiveness of abnormal returns for low Twitter sentiment.

Keywords: Asset Pricing, Big Data, Sentiment Analysis, Stock Return, Investor Sentiment.

Acknowledgements

I am grateful to my advisor, Professor A. Emanuele Rizzo, whose expertise and guidance was fundamental to this dissertation's development. Your great advice and questions fostered my learning process and enhanced my work.

I would also like to acknowledge Professor Nicholas H. Hirschey for nurturing my passion for the field of data analytics applied to finance, influencing the elaboration of this thesis. Thank you for giving me the opportunity to learn by working directly alongside you.

In addition, I express my gratitude to Nova School of Business and Economics, including professors, staff members, and colleagues who played a role in my Master's in Finance journey.

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

I. **Introduction**

Investor sentiment, a kernel field of behavioral finance, has long been linked by practitioners and academics to a role in average stock returns patterns and potential mispricing. For example, (C. Lee, Shleifer, and Thaler 1991) findings suggest that changes in individual investor sentiment, adopting closed-end funds discounts as the proxy, are significantly correlated with securities' prices, particularly for small stocks. (de Bondt and Thaler 1985; 1987) dissected the subject from a different standpoint, evidencing a pervasiveness of stock prices underreaction and overreaction to financial news, which is consistent with the overreaction hypothesis, in violation of Bayes' rule. This inclination towards papers supporting the prevalence of behavioral bias in stock returns led (Barberis, Shleifer, and Vishny 1998) to propose a model of investor sentiment based on psychological evidence.

(Fama 1998) argued against this narrative, reasoning that the empirical evidence was insufficient to reject the market efficiency hypothesis. The paper stated three major counterarguments. Firstly, the existence of overreaction or underreaction is consistent with the market efficiency hypothesis if one offsets the other entirely. Secondly, most research focuses on short return windows, not capturing a full view of market inefficiency. Thirdly, most empirical evidence on long-term return anomalies become marginal or non-existent when exposed to different asset pricing models or tested for different statistical approaches.

These arguments motivated academics to improve and explore new proxies for investor sentiment, considering more extensive research periods. (W. Lee, Jiang, and Indro 2002) suggested an interesting approach considering the DIJA, S&P500 and the NASDAQ. This paper is of particular interest, showing pervasiveness and consistency across all the three indices mentioned above, with a more substantial impact on NASDAQ, contradicting previous studies conclusions that investor sentiment solely affects small stocks. Several papers focused on employing more comprehensive proxies for investor sentiment, inferring that the variable consistently impacts stock returns for distinct statistical approaches. For instance, (Baker and Wurgler 2006) derived investor sentiment as a combination of six proxies, and (Schmeling 2009) studied consumer confidence across 18 industrialized countries. The conclusions introduce new findings, suggesting pervasiveness across growth and value stocks, noting a more extreme relation for markets with less integrity.

Finance has always relied heavily on structured and unstructured data, particularly financial markets. Nevertheless, all literature reviewed until this point only used market-based variables (structured data) as proxies to investor sentiment, such as closed-end funds discounts, NYSE share turnover, trading volume, amongst others. These measures are explained by many other economic forces, not capturing the authentic essence of investor sentiment (Da, Engelberg, and Gao 2015). To tackle this limitation, some practitioners relied on survey-based measures, raising several additional limitations inherent to surveys like subjectiveness and low frequency.

The surge of emerging technologies such as machine learning and natural language processing (NLP) revigorated sentiment research, awarding academics the capability to process and derive unstructured data, as text, into numerical variables that could be regressed and tested against stock returns. For instance, (Tetlock 2007) explored the use of NLP to derive investor sentiment, using daily content from the Wall Street Journal, suggesting that high media pessimism foretells downward stress on market prices. Additionally, the surge of the web 2.0 phenomenon, characterized by the emergence of social networks and blogs as an essential big data source, and the growing popularity of finance and financial markets amongst retail investors, formed the opportunity for practitioners to perceive financial-related microblogging as a possible source of investors opinion and sentiment (Pak and Paroubek 2010). Social media platforms unveil enormous potential as relevant sources of mass sentiment, given its intrinsic qualities like high frequency (timeliness), large scale (law of large numbers), and relevancy. Nevertheless, as a recent data source, there are still relevant limitations.

II. Hypotheses Formulation

This paper aims to contribute to the field of behavioral finance by exploring investor sentiment association to stock returns, employing social media securities-related microblogging as the proxy for investor sentiment. Twitter is the data source providing more concise and robust conclusions, highly popular amongst academia.

Despite the very distinct approaches followed by academics on dissecting the Twitter sentiment predictability of stock returns, most papers suggest a clear and significant correlation to some degree. For instance, (Bollen, Mao, and Zeng 2011; Oliveira, Cortez, and Areal 2017; Sprenger, Tumasjan, et al. 2014) consider distinct time periods and stock universes but arrive at similar findings, suggesting a consistent and positive association between Twitter sentiment and future stock returns. (Sul, Dennis, and Yuan 2017) approached the problem from a different stance, conveying the pervasiveness of Twitter sentiment impact on stock returns for the next trading day, next 10 days, and next 20 days. (Sprenger, Sandner, et al. 2014) contested these results encountering a low correlation and Granger causality of the relation, considering the DIJA as the stock universe and a time period of 15-months. However, finding a significant dependence during peaks of Twitter volume.

Based on this evidence and on the intent of testing the relation for a practical and realistic investment scenario, I lay out the main hypotheses as 1-month lagged Twitter sentiment does predict (H1) or does not predict (H0) stock returns.

Moreover, given the evidence suggesting a different consistency for the impact of high values (good sentiment) and low values (bad sentiment) on future stock returns, sub-hypotheses are formulated based on the Twitter sentiment. Scilicet, high 1-month lagged Twitter sentiment does predict (Hα1) or does not predict (Hα0) stock returns; low 1-month lagged Twitter sentiment does predict (Hβ1) or does not predict (Hβ0) stock returns.

III. **Methodology**

A. Data Collection

Data collection and management is an essential step to guarantee the data's relevancy, quality, and reliability. Thus, one must ensure that the source is a suitable social media platform comprising a large and diverse volume of data representing financial-related mass sentiment. Accordingly to (Bukovina 2016), Facebook, Twitter, and Google News are the most relevant and influential media platforms for investor sentiment. However, while Facebook and Google show compelling results in terms of global market sentiment, Twitter produces more robust results at the individual stock level. Additionally, the latter expedites the tracking of stock-related microblogging by providing "cashtags"[1] that identify stock symbols references.

Concerning the universe of stocks to consider, previous research presents robust evidence using DIJA constituents, an index often described as an inadequate representation of the US stock market. Hence, this paper aims to comprise a more extensive and representative stock universe, the S&P500, composed of the 500 largest US securities.

The typical methods used to extract Twitter data are Twitter's API[2] or web scraping[3]. Here, a trade-off between efficiency and relevancy occurs. The first grants limited data with a low execution time, and in contrast, the latter provides unlimited data with high execution time and high computational power. Hence, prioritizing data's relevancy and completeness, this research relied on tools like Python and Selenium[4] to apply the web scraping[3] technique. Nevertheless, to partially offset the efficiency issue and reduce noise, this approach only considers the relevant daily tweets with higher visibility and sentiment impact, filtered accordingly to Twitter's search algorithm.

---

[1] A "cashtag" is a searchable keyword, prefixed with the dollar sign ($), included in a tweet associated with one or multiple securities. E.g.: "$AAPL", "$MSFT" and "$ADBE".
[2] Twitter API (Application Programming Interface) is the Twitter's official API.
[3] Web scraping is a technique used to collect content and data from the internet.
[4] Selenium is a Python library to carry out automated test cases for browsers or web applications.

B. Signal Construction

Following data collection, it is necessary to transform this unstructured data into a sentiment variable. For this purpose, this research employs a Natural Language Processing (NLP) method, especially related to lexical semantics and sentiment analysis, a branch of computer science characterized by the capacity to derive text into a numerical variable depicting intrinsic sentiment and opinion. The signal construction considered the following steps.

a) *NLP Methodology.* Literature suggests that classification with supervised learning (Liu and Zhang 2012) or lexicon-based methods produce similar results, particularly as a mechanism to derive social media sentiment (Dhaoui, Webster, and Tan 2017; Verma and Thakur 2018). Nevertheless, the usage of BoW[1] is notably more straightforward than a machine learning approach (Augustyniak et al. 2014). Furthermore, the latter demands a labelled training dataset with sufficient instances to attain an acceptable accuracy target (Figueroa et al. 2012). Thus, this research employs a lexicon-based approach to derive text sentiment.

b) *Lexicon.* The accuracy of a lexicon-based NLP procedure depends heavily on the lexicon used and its "goodness of fit" for the data text characteristics. The Valence Aware Dictionary and Sentiment Reasoner (VADER) is one of the most well-known and adopted lexicons, specifically developed to tackle the inherent ambiguity of Twitter microblogging (Hutto and Gilbert 2014). Moreover, when tested against other popular sentiment analysis methods, VADER outperforms across different contexts, manifesting adaptability to different datasets, including securities-related text. (Koratamaddi et al. 2021). As a secondary lexicon approach, this research incorporates (Loughran and Mcdonald 2011) financial dictionary, comparing results and driving insights on the usage of financial terminology amongst Twitter microblogging.

[1] Bag of Words is a natural language processing approach that evaluates a text by the bag (multiset) of its words.

c) *Text Cleansing.* Deriving sentiment through lexicon-based models requires the removal of any characters and expressions unassociated with sentiment, minimizing errors. For instance, VADER interprets words and particular characters combinations to identify text emoticons, and thus cleaning punctuation would likely decrease the model's accuracy. Considering such concerns, the process of data cleansing consisted of removing duplicated instances, references to hyperlinks, and any other irrelevant characters identified through regular expressions (regex).

d) *Predicting Sentiment.* The application of the proposed lexicon-based methodologies on the cleansed data, derived 34,683,864 numerical instances representing the sentiment score for each tweet collected. Consequently, as the anomaly variable consists of monthly Twitter sentiment, the monthly averages were computed for each security.

C. Strategy Portfolios and Empirical Approach

Following the formulation of the monthly sentiment score for the correspondent S&P500 constituents between April 2008 and August 2021, the dataset was sorted by sentiment and grouped into terciles, originating three portfolios, top (1st tercile), middle (2nd tercile) and bottom (3rd tercile). Bloomberg datastream was the source of data for monthly stock returns and correspondent market capitalizations, used to derive value-weight and equal-weight returns. Additionally, to study the variable's capability to predict next month stock returns, the long-only (1st tercile) and hedge (1st tercile minus 3rd tercile) portfolios were formed and regressed against the capital asset pricing model (Sharpe 1964; Lintner 2006), Fama-French 3-factor model (Fama and French 1993), and Fama-French 5-factor model (Fama and French 2015). The asset pricing models' portfolios were sourced from Kenneth R.French library.

<ins>Note</ins>: The Appendix exhibits a Methodology Schematic.

IV. **Data**

Table I shows the descriptive statistics on the number of instances collected (number of tweets), for all S&P500 constituents, between April 2008 and August 2021. Additionally, for visualization purposes, the stock universe was sorted by the total number of instances and then divided into deciles, providing a better understanding of the quantity of data extracted across the different stock groups.

34.7 million tweets were collected, from which 14.2 million correspond to 50 stocks, the top decile [1:50]. The evidence that roughly 41% of the instances collected are related to only 10% of the stock universe is not unusual, considering that, similarly to what occurs with securities-related news, social media microblogging is centered around trending and popular stocks. The data supports this rationale, as across the deciles' composition it is observable that a significant part of the data collected is related to large market capitalization and momentum stocks, factors ordinarily positively correlated with higher popularity.

Following the sentiment aggregation by day, the average security has 3386 days of investor sentiment, ranging between 1814 to 4344 days. Consequently, the same method is applied to derive monthly sentiment forming 127 months of data for the average stock, a value that ranges between 82 and 148 months. Moreover, it is observable that standard deviation consistently increase throughout the groups, however dramatically soaring for the last decile [451:504]. The reasoning underlies a few securities, namely BBWI, HWM, VTRS, and AMCR, which are highly unpopular amongst Twitter microblogging, creating a cluster of outliers that are significantly below the median, causing the standard deviation of the group to increase significantly. Given the weak influence of these outliers, it is unlikely that their presence would affect the validity of the results for the S&P500. However, this manifests a significant limitation of the employment of a similar approach to a broader stock universe heavily comprised of equities associated with low capitalization and low popularity.

Table I

Statistics on the number of tweets collected for the stock universe (SPX constituents): April 2008 to August 2021

		Number of Tweets				Number of Daily Signals				Number of Monthly Signals			
Range	# Stocks	Total	Average	Median	Stdev	Total	Average	Median	Stdev	Total	Average	Median	Stdev
All SPX Constituents													
[1 : 504]	505	34683864	68680.92	40328.00	85873.29	1709981	3386.10	3538.00	894.01	64060	126.85	133.00	29.52
Descending Sorting by Number of Tweets, deciles													
[1 : 50][1]	50	14178794	283575.88	231427.00	119918.39	217222	4344.44	4514.00	483.81	7415	148.30	153.00	15.36
[51 : 100][2]	50	5729637	114592.74	110188.50	18938.13	207237	4144.74	4273.00	377.52	7291	145.82	151.00	13.42
[101:150][3]	50	3712181	74243.62	72338.00	8109.32	195545	3910.90	3980.50	440.49	7015	140.30	143.00	15.82
[151 : 200][4]	50	2748731	54974.62	54725.50	3716.03	189404	3788.08	3822.50	299.32	6878	137.56	136.00	10.86
[201 : 250][5]	50	2269288	45385.76	45151.00	2706.05	180540	3610.80	3711.00	553.00	6633	132.66	135.50	18.80
[251 : 300][6]	50	1832128	36642.56	36975.00	2487.17	164785	3295.70	3445.50	631.31	6224	124.48	132.00	26.26
[301 : 350][7]	50	1469998	29399.96	29629.50	1702.21	159666	3193.32	3382.00	590.39	6099	121.98	131.00	24.21
[351 : 400][8]	50	1196117	23922.34	23963.50	1843.15	154190	3083.80	3203.50	495.27	6006	120.12	126.50	21.37
[401 : 450][9]	50	947937	18958.74	18941.50	1299.72	141640	2832.80	2842.00	324.14	5962	119.24	120.50	18.06
[451 : 504][10]	55	599053	10891.87	12309.00	4410.82	99752	1813.67	2081.00	907.87	4537	82.49	96.00	45.21

[1] **[1 : 50] Tickers:** FOX; KIM; EBAY; ABC; ETSY; AFL; AAPL; IBM; GS; AAP; CVS; C; AMD; F; T; CAT; O; TWTR; INTC; GE; WBA; XOM; AMZN; SBUX; WFC; CERN; MSFT; A; BAC; GM; PEG; BA; MCD; CMG; NRG; LOW; NFLX; UPS; ICE; VZ; GILD; MU; MS; ES; QCOM; CVX; TSLA; MRK; FCX; COST;

[2] **[51 : 100] Tickers:** EW; EA; TGT; DFS; HPQ; BMY; NEM; FB; NOW; AMGN; DAL; AAL; AXP; WYNN; MSI; BBY; GOOG; LVS; ATVI; COP; BIIB; DOW; FDX; MO; DIS; LLY; PEP; NKE; MMM; HAL; AIG; LUV; CL; CSCO; PFE; REGN; MGM; UAL; LMT; ABT; JPM; UA; STX; JNJ; AMAT; AVGO; CRM; YUM; FAST; SWKS;

[3] **[101 : 150] Tickers:** MA; MDT; BLK; SLB; ISRG; CCL; TMUS; AMP; GPS; DE; PG; PM; OXY; TEL; WMT; HD; MCO; KMI; ORCL; EOG; KR; KEY; NVDA; TXN; MRO; VRTX; UNH; VLO; ULTA; APA; WDC; MNST; MOS; LEN; PYPL; ABBV; HON; UNP; EXPE; HAS; ALL; ILMN; USB; AME; NXPI; ADBE; DISH; HUM; ACN; ETN;

[4] **[151 : 200] Tickers:** INCY; DD; CME; CMCSA; CF; TTWO; DVN; DPZ; ALK; MET; STZ; TJX; CSX; SO; PNC; BSX; RCL; DG; TSCO; NOC; PENN; ADSK; TSN; MAR; LRCX; FFIV; DLTR; PSX; GIS; COF; IR; ROST; MPC; CZR; MDLZ; JNPR; SCHW; VTR; ESS; DRI; CI; CAG; SPG; CTSH; AKAM; PXD; NEE; NDAQ; NTAP; PRU;

[5] **[201 : 250] Tickers:** HES; D; COG; WMB; DUK; XEL; FOXA; TRV; DHI; GD; ADM; BAX; GLW; XLNX; FTNT; V; MRNA; KMB; L; CHTR; PHM; WHR; INTU; CMI; RF; ADI; ADP; NUE; BK; AZO; ENPH; ORLY; MCK; AMT; EMR; ALGN; CNC; KMX; NWS; CLX; RL; SEE; TAP; STT; IT; K; EXC; HSY; EQIX; EFX;

[6] **[251 : 300] Tickers:** DISCA; PAYX; CPB; URI; FANG; SYK; KHC; ALB; ANTM; FITB; WM; DHR; ED; HCA; NSC; EL; BRK; DXCM; FIS; CTXS; JCI; APH; PPL; COO; FISV; ANET; TMO; IP; BDX; GRMN; TER; MTCH; DLR; CAH; SHW; OGN; HBAN; MAS; PPG; SYY; PVH; HPE; HRL; UAA; PGR; SJM; MCHP; NCLH; JBHT; ARE;

[7] **[301 : 350] Tickers:** TROW; VFC; QRVO; FE; FMC; PTC; NWL; HST; CCI; HBI; BEN; NWSA; AEP; OKE; CMA; ZTS; MKC; WY; APD; ABMD; FLT; TXT; LYB; ITW; PLD; GWW; SYF; BKNG; GPN; ROK; HLT; KLAC; ZION; SWK; PAYC; DOV; HOLX; LH; LNC; EMN; CB; KSU; PSA; PKG; DGX; CBOE; IPG; WU; VIAC; MTB;

[8] **[351 : 400] Tickers:** CTAS; CDNS; DVA; HIG; PCAR; LEG; PH; RMD; TDG; IVZ; ECL; IRM; MLM; RSG; BLL; BWA; SRE; KO; CHD; PWR; AWK; IFF; SNPS; OMC; CHRW; KEYS; ROP; SBAC; VMC; TECH; EQR; HII; ETR; UHS; CNP; GOOGL; AVB; CMS; NTRS; AES; HSIC; MMC; IDXX; CE; EXR; FRC; BXP; SNA; AON; WAT;

[9] **[401 : 450] Tickers:** MSCI; AEE; TRMB; CFG; LYV; BRO; EIX; PFG; CPRT; PNR; PBCT; IPGP; INFO; BR; NI; TFX; SIVB; MHK; AVY; DTE; MTD; FBHS; PKI; VNO; GNRC; EXPD; GPC; REG; NLSN; SPGI; XRAY; CRL; AOS; POOL; RHI; MKTX; RJF; LKQ; WELL; TYL; WAB; VRSN; RE; DRE; ANSS; MAA; XYL; ODFL; AIZ; LDOS;

[10] **[451 : 504] Tickers:** LNT; WEC; WRK; JKHY; NVR; AJG; PNW; TPR; VRSK; ATO; FRT; CINF; BIO; CDW; ZBH; IEX; CTLT; RTX; ROL; FTV; DXC; PEAK; MPWR; WLTW; LIN; UDR; ALLE; WRB; STE; ZBRA; DISCK; TDY; WST; IQV; J; CARR; APTV; LW; TFC; LHX; TT; CBRE; CDAY; BKR; NLOK; LUMN; EVRG; BF; CTVA; OTIS; GL; AMCR; VTRS; HWM; BBWI;

V. Summary Statistics

The methodology considered two dictionaries to derive investor sentiment, intending to proceed with the one that best fits the data. The first acknowledges words and their sentiment based on standard social media syntax, while the alternative is an updated version to include financial terminology. For simplification purposes, this paper refers to the first as "non-financial-lexicon" or "NFL", and to the latter as "financial-lexicon" or "FL".

Figure I and Table II exhibit the sample distributions and the correspondent descriptive statistics. The results suggest that the different approaches lead to significantly distinct distributions, particularly reflecting on the mean, kurtosis, and skewness measures. The NFL sentiment ranges between -0.813 and 0.878 with a mean of 0.123 and a median of 0.117, implying a positive skewness (0.434), indicating that smaller negative sentiments have a lower frequency than larger positive sentiments (longer right-tail), which is similar to the skewness pattern commonly present in stock returns. In contrast, the FL sentiment ranges between -0.689 and 0.66 with an average and median of 0.007 (symmetrical distribution). Additionally, it has a significantly higher kurtosis (18.649), suggesting a more substantial concentration of the distribution around the median, with heavy tails on both sides (leptokurtic distribution).

Thus, in this case, the results suggest that modernizing the dictionary with a financial lexicon weakens the overall sentiment across the data, by predicting a neutral (0.00) sentiment score for a significant percentage of the tweets and removing extreme values. (Loughran and Mcdonald 2011) dictionary processes technical financial expressions, based on 10-X[1] filings, by attributing a sentiment to financially meaningful words and removing sentiment to financially neutral words. The pitfall of this approach is related to the removal of common expressions that are highly meaningful in the context of social media sentiment, even when related to securities and other financial assets microblogging.

[1] 10-X filings: 10-K, 10-K/A, 10-K405, 10-K405/A, 10KSB, 10KSB/A, 10-KSB, 10-KSB/A, amongst others.

Table II shows statistical measures on the variable per group size. Since the stock universe relies on the 500 largest US-listed companies, the most relevant size groups to consider are mega, big, and mid-caps. Big-caps are approximately 86% of all sentiment observations, followed by mega-caps with 9% and mid-caps with 5%. As expected, these proportions correspond to the stock universe characteristics. Furthermore, the distribution for mega-caps seems to be negatively skewed and with significantly high kurtosis. Plausible reasoning is that mega-caps are the most popular stocks amongst social media users, thus reflecting a lot of informative text unrelated to sentiment or opinion. This effect is not discernible for the other size groups.

Figure I provides a detailed view of the average monthly investor sentiment between April 2008 to August 2021 for both methodologies. The NFL sentiment line infers that the average sentiment is always slightly positive and highly volatile. In other words, equities with an average monthly positive sentiment overweight the ones with the opposite sentiment, inducing an overall positive monthly market sentiment. This effect might be related to lexicon-based models' inherent traits and flaws. For instance, lexicons are historically known to predict positive sentiment precisely but inaccurate at predicting negative sentiment because of subjective expressions such as sarcasm, irony, and others.

Based on the sample distributions, the statistics suggest that acknowledging words and their sentiment based only on standard social media syntax (NFL) better represents investor sentiment sourced from Twitter. This does not imply that a financial lexicon (FL) is necessarily a weaker approach to derive investor media-based sentiment for all scenarios. As shown in prior studies, it is the best methodology to follow when the text relies on financial technical terms to express opinion and sentiment, which is generally not the case for microblogging.

Therefore, from this point onwards, the paper’s discussion focuses on the NFL as the primary lexicon-based approach to derive investor sentiment.

Figure I

Sample Distributions and Average Monthly Twitter Sentiment: April 2008 to August 2021

Panel A: Sentiment Distribution — **Panel B:** Average Monthly Sentiment

NFL Sentiment: VADER Lexicon — FL Sentiment: VADER updated from the Financial Dictionary

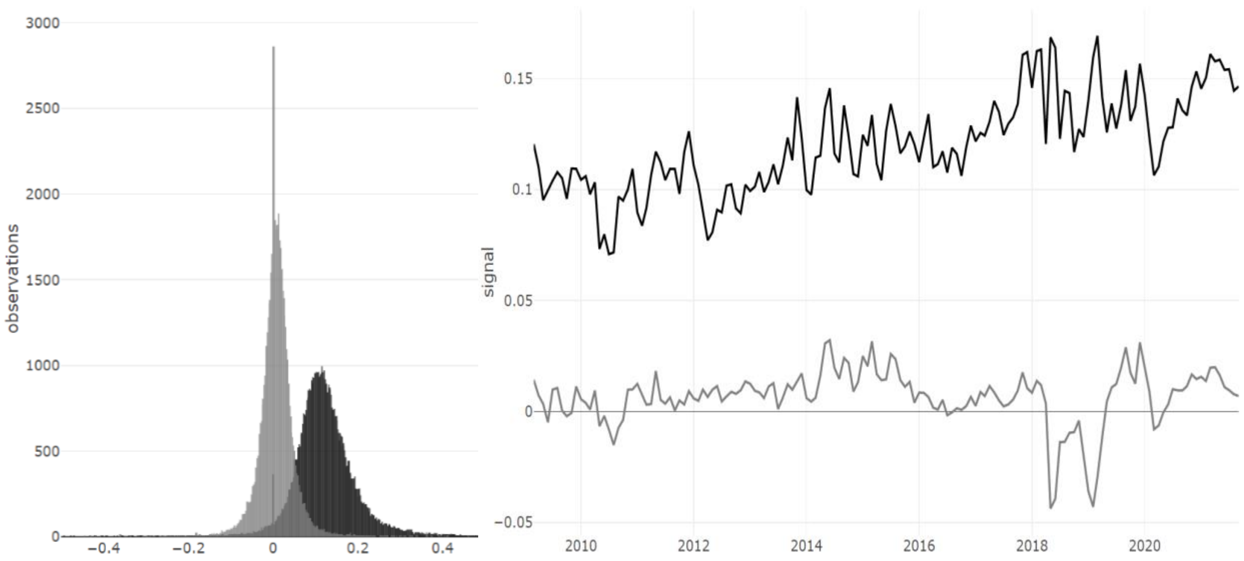


Table II

Twitter Sentiment Descriptive Statistics

The table shows monthly Twitter sentiment statistics regarding the entire sample and for the different size groups. There are two sentiment samples derived from the following lexicons:

i. NFL – Non-Financial Lexicon Sentiment: derived using the VADER lexicon.
ii. FL – Financial Lexicon Sentiment: derived using the VADER lexicon updated from the Financial Dictionary (Loughran and Mcdonald 2011).

The different size groups were derived accordingly to the following expression:

$$sizeGrp(string) = \begin{cases} \text{"}mega\text{"}\ if\ \{mktCap \in \mathbb{Q}: 10 \times 10^9 < mktCap \leq 200 \times 10^9\} \\ \text{"}big\text{"}\ if\ \{mktCap \in \mathbb{Q}: 2 \times 10^9 < mktCap \leq 10 \times 10^9\} \\ \text{"}mid\text{"}\ if\ \{mktCap \in \mathbb{Q}: \text{mktCap} \leq 2 \times 10^9\} \end{cases}$$

	Sample		Per Size Group					
			Mega		Big		Mid	
	NFL	FL	NFL	FL	NFL	FL	NFL	FL
# Observations	64060		5726		54821		3513	
Average	0.123	0.007	0.120	0.005	0.123	0.007	0.131	0.008
Stdev	0.074	0.042	0.066	0.039	0.073	0.042	0.093	0.040
Skewness	0.434	-0.056	-1.072	-0.669	0.353	-0.202	1.365	1.314
Kurtosis	8.936	18.649	25.873	50.751	7.761	16.272	4.872	9.766
Minimum	-0.813	-0.689	-0.802	-0.636	-0.813	-0.689	-0.390	-0.222
Q1	0.082	-0.012	0.085	-0.010	0.082	-0.012	0.081	-0.012
Median	0.117	0.007	0.117	0.006	0.117	0.007	0.115	0.007
Q3	0.158	0.026	0.155	0.021	0.158	0.027	0.157	0.026
Maximum	0.878	0.660	0.802	0.547	0.878	0.660	0.765	0.373

VI. **Sorts**

Sorts on the variable of interest is generally the first approach to reveal a superficial depiction of the variation of the returns across the scope of the anomaly variable, an appropriate preliminary test to the main hypotheses, namely 1-month lagged Twitter sentiment does predict (H1) or does not predict (H0) stock returns. For this purpose, value-weighted and equal-weighted tercile portfolios were formed considering monthly sorts on 1-month lagged Twitter sentiment and the corresponding returns from April 2008 through August 2021. Moreover, first and second half portfolios were developed to test the pervasiveness of the results across the different periods.

Table III shows the average annualized returns, standard deviations, and Sharpe ratios for the portfolios, and Figure II exhibits a graphical representation of the cumulative excess returns. These returns are also normalized for 10% target volatility, providing a risk-adjusted performance comparison.

A. Terciles

Table III results suggest that, in terms of average annualized returns, the top tercile outperforms the middle and bottom terciles consistently, for value-weight and equal-weight returns applied to the entire time period, but also for the first and second halves. Additionally, the bottom underperforms the middle tercile consistently for value-weight returns. Figure II confirms these conclusions, revealing that the top is consistently the best performer in cumulative excess returns, followed by the middle and ultimately by the bottom tercile yielding the worst returns. This clear pattern of the returns variation across the terciles, representing the different spectrums of the explanatory variable, provides preliminary evidence against the null hypothesis (H0), supporting that 1-month lagged Twitter sentiment is positively associated to stock returns.

However, this is not as evident for the equal-weight returns across the different periods. For instance, for the sample period, the equal-weighted bottom tercile yields on average 15% yearly, the same as the middle tercile. Moreover, considering the second half, it underperforms the bottom tercile by 1%.

In finance, it is not satisfactory to infer that a portfolio is better solely based on returns since those returns might associate with unbearable volatility. Therefore, an analysis of risk-adjusted returns is crucial. For this purpose, Table III shows the Sharpe ratio amongst the different terciles and the distinct time periods. Additionally, Figure II shows a graphical representation of the cumulative excess returns for a volatility target of 10%.

The top value-weight and equal-weight returns yield significantly larger Sharpe ratios for all the available periods, showing a pervasive risk-adjusted outperformance compared to the middle and bottom terciles. This suggests that the top tercile not only consistently yields higher returns in absolute terms but also per unit of risk carried by the investor. Furthermore, the Sharpe ratios are consistently higher for the middle compared to the bottom tercile. This leads to particularly interesting conclusions since the middle tercile did not consistently outperform the bottom tercile in terms of average annualized equal-weight returns. Figure II reinforces the results stated, demonstrating that, for the same volatility (10%), there is a clear trend for both value-weight and equal-weight cumulative excess returns, with the top tercile consistently leading to the best risk-adjusted performance, followed by the middle, and then lastly the bottom tercile. This clear risk-adjusted performance pattern reinforces the preliminary evidence against the null hypothesis (H0), initially drawn above.

Observing the average annualized standard deviations, the results show an unexpected insight. The extreme terciles (top and bottom) yield consistently higher standard deviations, a finding suggesting that extreme 1-month lagged Twitter sentiment is associated with higher returns volatility.

B. Value-Weight vs. Equal-Weight Returns

The value-weighted and equal-weighted portfolios produce similar and consistent results. Nevertheless, a direct comparison between both might provide a deeper understanding of the nature of the Twitter sentiment.

Considering the top tercile, the value-weighted approach yields a 24% annual return rate for the entire sample period, significantly larger than the 18% for equal-weight returns. Additionally, the annualized standard deviation is the same for both approaches (0.19), leading to the conclusion that value-weight returns optimize the anomaly variable association to both absolute and risk-adjusted returns, producing consistently higher Sharpe ratios for the entire sample period, first and second halves.

Moreover, the value-weight returns provide a more consistent and clear performance distinction between the terciles. This difference is observable in Figure II, in which, for value-weight returns, there is a clear outperformance of the top tercile compared to the middle tercile, but also of the middle compared with the bottom tercile. In contrast, for equal-weighted returns, while there is still a clear outperformance of the top compared to the middle tercile, the middle yields very similar excess cumulative returns compared to the bottom tercile.

These results indicate that the value-weighted returns foster 1-month lagged Twitter investor sentiment as a predicting factor for stock returns, suggesting that the anomaly variable is better at predicting returns for equities with a higher market capitalization. Plausible reasoning might be related to the data concentration around larger capitalization securities, as described in the data section. Implying fewer instances for smaller capitalization stocks, thus fewer degrees of freedom, and consequently, according to the law of large numbers, an average sample variable that is possibly farther away from the average of the actual population, inducing to less accurate predictions. This would also mean higher accuracy at predicting Twitter sentiment for larger capitalization stocks, exacerbating the discrepancy.

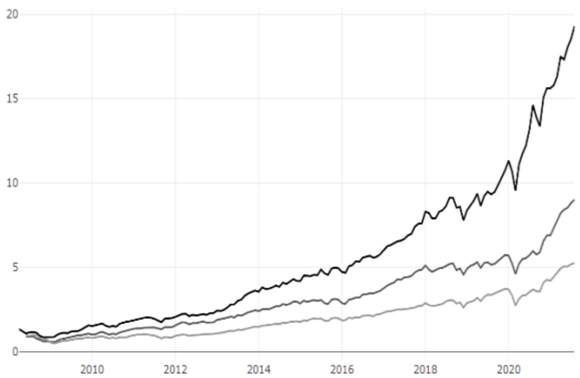
Table III
Average Annualized Returns, Standard Deviations and Sharpe Ratios for the Tercile Portfolios Formed using Sorts on 1- Month Lagged Twitter Sentiment: April 2008 to August 2021

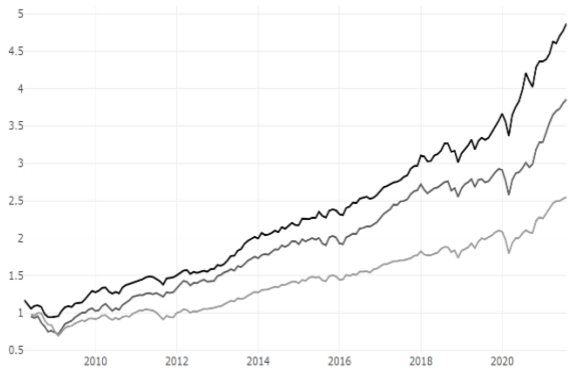
	Average Annualized Returns			Average Annualized Standard Deviations			Sharpe Ratio		
	Bottom	2	Top	Bottom	2	Top	Bottom	2	Top
Value-Weight Returns (VW)									
Sample Period	0.14	0.18	0.24	0.18	0.17	0.19	0.76	1.07	1.24
1st Half	0.11	0.18	0.24	0.22	0.19	0.22	0.51	0.93	1.07
2nd Half	0.18	0.19	0.24	0.17	0.15	0.16	1.07	1.26	1.52
Equal-Weight Returns (EW)									
Sample Period	0.15	0.15	0.18	0.20	0.17	0.19	0.76	0.86	0.97
1st Half	0.15	0.17	0.21	0.22	0.20	0.22	0.69	0.83	0.94
2nd Half	0.15	0.14	0.16	0.17	0.16	0.16	0.87	0.90	1.03

Figure II
Cumulative Excess Returns for the Tercile Portfolios Formed using Sorts on 1- Month Lagged Twitter Sentiment: April 2008 to August 2021

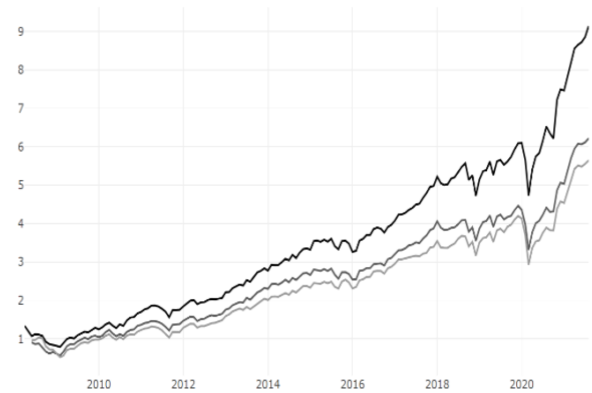
Panel A: No Target Volatility | **Panel B:** 10% Target Volatility

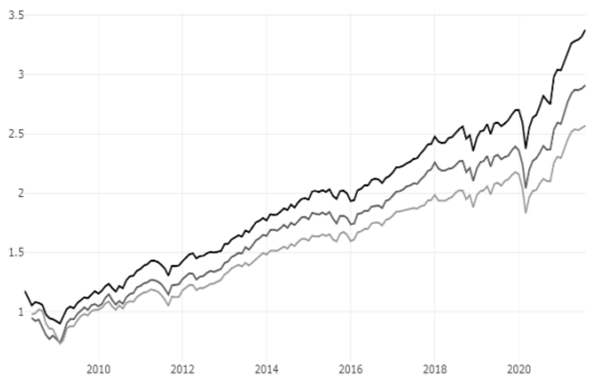
Value-Weight Returns (VW)





Equal-Weight Returns (EW)





— Top Portfolio (1st Tercile) — Middle Portfolio (2nd Tercile) — Bottom Portfolio (3rd Tercile)

After dissecting the Twitter sentiment behavior across the different terciles, a standard procedure relies on the formation of portfolios derived from the extreme terciles to test the sub-hypotheses formulated. Firstly, observing the anomaly variable applied to a long-only portfolio (1st tercile) provides an initial test to the sub-hypotheses that high 1-month lagged Twitter sentiment does predict (Hα1) or does not predict (Hα0) stock returns. Secondly, the formation of an hedge portfolio, comprised of long-short positions in the outer terciles (1st tercile minus 3rd tercile) allows a stronger perception not only on the variable's capability to predict higher excess returns but also of lower excess returns, thus extending the test to the sub-hypotheses that low 1-month lagged Twitter sentiment does predict (Hβ1) or does not predict (Hβ0) stock returns.

Additionally, given the evidence suggesting value-weight returns as the approach that optimizes results, this research emphasizes this approach from this point onwards. Nevertheless, the same results for equal-weight returns are shown in the appendix.

C. Long-Only Portfolio Returns

Figure III provides a comparison between the market and the value-weighted long-only portfolio cumulative excess returns, in which one can clearly observe a significant outperformance of the long-only portfolio over the market excess returns for the sample period. These results are also supported by Table IV, suggesting that the first yields an average annualized return of 24% while the latter returns on average 13%, depicting a very significant yearly average outperformance of 11%. Additionally, for the first half, while the long-only strategy sustained the same average annualized return (24%), the excess market annualized returns decreased to 11%, less than half of the long strategy's performance. In terms of risk-adjusted returns, as shown in Figure III, the outperformance of the long-only is smaller but still very significant, yielding a Sharpe ratio of 1.24 against 0.77 for the market excess returns. Furthermore, Table IV standard deviations suggest that the long-only portfolio returns have consistently higher volatility compared to the market excess returns.

These results show a consistent outperformance, in terms of absolute and risk-adjusted excess returns, of the value-weighted long-only portfolio against the market portfolio, for all the sample periods available. This pervasive and consistent outperformance supports the sub-hypothesis that high 1-month lagged Twitter sentiment does predict stock returns (Hα1).

D. Hedge Portfolio Returns

Figure III suggests that the hedge portfolio is the worst performer, with the market yielding higher excess returns. Table IV supports this conclusion, showing that the hedge portfolio returns on average 7% per year, against the 13% yielded by the market. Additionally, it underperforms consistently for the first and second halves.

In contrast, it is the best strategy in terms of volatility with a standard deviation of 0.12, against the 0.17 implied by the market excess returns. A low volatility in the hedge portfolio implies a high correlation between the extreme terciles (1st and 3rd tercile), as a negative correlation would imply higher volatility. Nevertheless, the benefit in terms of volatility is insufficient when measuring risk-adjusted returns, as shown by Table IV the hedge portfolio offers a significantly lower Sharpe ratio than the market excess returns for the sample period. Figure III supports these conclusions.

In contrast to the evidence shown on the long-only returns, the hedge portfolio consistently underperforms the market for all the periods available in terms of absolute and risk-adjusted excess returns. As mentioned previously, the hedge portfolio is a result of a long-short position in the extreme terciles. Thus, if the top (long) tercile clearly outperforms, the hedge portfolio underperformance must be a direct result of the bottom (short) tercile inability to produce consistently lower returns, showing insufficient preliminary evidence to reject the sub-hypothesis that low 1-month lagged Twitter sentiment does not predict (Hβ0) stock returns. Conceivable reasoning for this conclusion could rely on methodology issues. For instance, lexicon-based NLP is highly accurate at predicting extreme high sentiment; however, it struggles at predicting extreme low sentiment, given the presence of expressions as irony and satire, that adopt words assigned to positive sentiment to express the opposite opinion.

Table IV
Average Annualized Returns, Standard Deviations and Sharpe Ratios for Excess Returns on the Market, VW Long-Only and VW Hedge Portfolios: April 2008 to August 2021

Details[1]

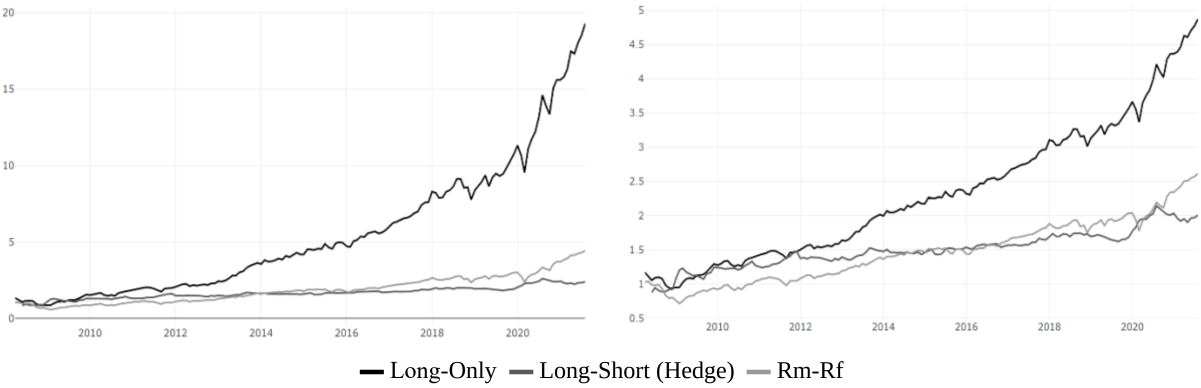
	Average Annualized Returns			Average Annualized Standard Deviations			Sharpe Ratio		
	Long	Hedge	Rm-Rf	Long	Hedge	Rm-Rf	Long	Hedge	Rm-Rf
Sample Period	0.24	0.07	0.13	0.19	0.12	0.17	1.24	0.58	0.77
1st Half	0.24	0.08	0.11	0.22	0.15	0.18	1.07	0.52	0.61
2nd Half	0.24	0.07	0.14	0.16	0.09	0.15	1.52	0.74	0.95

Figure III
Cumulative Excess Returns for the Market, VW Long-Only and VW Hedge Portfolios: April 2008 to August 2021.

Details[1]

Panel A: No Target Volatility	**Panel B:** 10% Target Volatility

Value-Weighted Portfolios (VW)



1

Long-Only Portfolio corresponds to the Top Portfolio, 1st tercile returns (Figure II).
Long-Short (Hedge) Portfolio hedges the Top Portfolio (1st tercile) against the Bottom Portfolio (3rd tercile).
$Hedge\_Returns_n = Top\_Returns_n - Bottom\_Returns_n$.
Rm-Rf from Kenneth R. French: "the excess return on the market, value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t, good shares and price data at the beginning of t, and good return data for t minus the one-month Treasury bill rate."

VII. Cross-Section Regressions

To exhaustively dissect the 1-month lagged Twitter sentiment variable as a predictor of stock returns the approach based on sorts on its own is insufficient. (Fama and French 2008) suggest exploring cross-section regressions as a complementary approach imposing a functional form on the relation between the anomaly variable and returns. With the intent of providing extensive evidence on the sub-hypotheses (Hα and Hβ), long-only and hedge portfolio excess returns for the entire sample, first and second halves, are tested against state-of-the-art asset pricing models, such as the capital asset pricing model (Sharpe 1964; Lintner 2006), Fama-French 3-factor model (Fama and French 1993), and Fama-French 5-factor model (Fama and French 2015). This empirical approach offers insights into the models' predictability of the observed returns and the correspondent correlation with the systematic risk factors. The presence of statistically significant abnormal stock returns (residuals) would imply that the systematic risk factors do not fully predict the realized stock returns produced by the anomaly variable.

The combination of the present models' systematic components derives from different explanatory factors for stock returns, each corresponding to value-weighted portfolios (Fama and French 2015). These portfolios are formed on the market excess returns (Rm-Rf), size (SMB, small-caps minus big-caps), book-to-market values (HML, high B/M minus low B/M, also depicting value minus growth), operating profitability (RMW, robust minus weak), and investment (CMA, conservative minus aggressive). Additionally, RMW and CMA are strongly negatively correlated and jointly decoded as a Quality factor.

An interesting approach for future research leans on employing Carhart's 4-factor model (Carhart 1997), exploring Momentum as a supplementary explanatory variable for the realized returns, as it is often related to behavioral-based explanations directly bonded to investor sentiment.

Table V shows the cross-section residuals, coefficients, and corresponding t-statistics, regarding the CAPM, FF3, and FF5, for the long-only and the hedge portfolio returns, from April 2008 through August 2021, first and second half.

A. *Long-Only Portfolio Returns*

The market excess returns (Rm-Rf) coefficients suggest a strong positive correlation with the long-only portfolio returns, with the models predicting a market beta of 0.961 for the CAPM, 1.05 for the FF3, and 1.007 for the FF5. This suggests a near-perfect positive correlation, conveying that any change in market excess returns, controlling for the remaining explanatory variables, will impact the portfolio's returns in the same direction nearly proportionally. Moreover, the corresponding t-statistics suggest that these slopes are statistically significant at a 0.01 significance level for every asset pricing model tested.

The SMB and HML coefficients, accordingly to FF3 and FF5, suggest a negative but weak correlation with the portfolio returns. The negative correlation with SMB (-0.155 and -0.139) is expected since the dependent variable consists of value-weight realized returns, emphasizing the weight of higher market capitalization stocks. Additionally, the stock universe considered is the S&P500, comprised only by mega, big, and mid-caps. Regarding the HML factor, FF3 and FF5 exhibit coefficients of -0.261 and -0.119. This slight negative correlation might be related to the anomaly variable (Twitter sentiment) concentration on popular stocks, generally growth and large tech stocks, thus possessing a strong presence of low B/M (growth) stocks. Nevertheless, one needs to consider that only the HML slope for the FF3 is statistically significant at a 0.99 confidence level for the sample period. Yet, considering the second half, SMB and HML coefficients are statistically significant at a 0.05 significance level.

Observing the remaining explanatory variables, only comprised on the FF5 model, the RMW coefficient predicts a weak (0.067), inconsistent (direction changes between first and second half), and statistically insignificant correlation. In contrast, the CMA slope shows a moderate negative correlation (-0.513) for the sample period, statistically significant at a 0.99 confidence level. The CMA factor is the return spread of stocks that invest conservatively minus aggressively. Thus, a negative correlation suggests that the long portfolio is more exposed to stocks that invest aggressively rather than conservatively, which supports the same reasoning discussed for HML, that the anomaly variable (Twitter sentiment) is associated with popular stocks, generally growth, and large tech stocks.

The alphas generated by the models suggest that the long-only portfolio realized returns significantly exceed the predicted by the present asset pricing models, beating the CAPM by 12.3%, the FF3 by 10.5%, and the FF5 by 11.2%. Additionally, given the pervasiveness of the residuals across the different periods and its statistical significance, the sub-hypothesis that high 1-month lagged Twitter sentiment does not predict stock returns ($H\alpha0$) is rejected at the 0.01 level of significance, in line with the conclusion drawn in the sorts section.

However, the second half indicates more substantial results, with a significantly higher r-squared across the different models (0.881, 0.923, and 0.928) and better t-statistics, implying a more satisfactory “goodness of fit”. In other words, for the first half, the asset pricing model with the highest r-squared (FF5) is able to explain around 58.4% of the variation in realized returns. Whereas for the second half, the same model is able to predict approximately 92.8% of the observed variation. This suggests that the statistical noise is more prominent for the first half, raising an interesting dilemma, where more data seems to be leading to less accurate information. This effect is likely to be related to the evolution of social media platforms as a relevant source of financial opinion. As shown by (Hentschel and Alonso 2014), there was an exponential growth of financial-related Twitter microblogging at the beginning of the 2010s, suggesting that before that, the data available could still be relatively small to cancel most of the noise incorporated. Nevertheless, the long-only abnormal returns for the second half are still considerably large and statistically significant at a 0.01 significance level, exceeding the CAPM by 10%, the FF3 by 7.7%, and the FF5 by 7.9%.

B. Hedge Portfolio Returns

Results from the cross-sectional regressions of the hedge portfolio returns against the CAPM, FF3, and FF5 suggest that the asset pricing models struggle to predict the variance of the realized stock returns, with consistent low r-squares. CAPM is only able to predict 3.9%, followed by the FF3 with a significant increase to 23.3%, and the FF5 predicting 23.4%.

The market excess returns coefficient suggests a weak negative correlation, given by the high association of the extreme terciles that comprise the hedge portfolio with the market excess returns. However, this coefficient is only statistically significant for the CAPM. Similarly, for SMB, RMW, and CMA factors, evidence shows a weak to zero statistically insignificant correlation. Interestingly, HML has a moderate negative correlation with the portfolio returns, indicating pervasiveness across the FF3 and FF5 and for the different time periods, statistically significant at a 0.01 significance level.

The Alphas produced by the cross-section regressions suggest that the hedge portfolio yields yearly positive stock abnormal returns, above the predicted by the CAPM (0.093), the FF3 (0.061), and the FF5 (0.064). Nevertheless, the results are statistically insignificant at a 0.05 significance level, except for the CAPM. Thus, given the high probability of the alphas being chance results, the sub-hypothesis that low 1-month lagged Twitter sentiment does not predict (Hβ0) stock returns cannot be rejected. Producing similar results and reasonings to the ones stated in the sorts section.

Table V
Cross-Section Regressions Results for Long-Only and Hedge Value-Weight Returns, from April 2008 to August 2021.

	α	$R_M - R_F$	SMB	HML	RMW	CMA	R^2	$Inf.Ratio$
Panel A: Long-Only Value-Weight Returns								
Sample Period: April 2008 to August 2021								
CAPM	0.123*** (3.747)	0.961*** (16.810)					0.641	1.054
FF3	0.105*** (3.272)	1.050*** (17.151)	-0.155 (-1.399)	-0.261*** (-2.906)			0.671	0.939
FF5	0.112*** (3.466)	1.007*** (16.031)	-0.139 (-1.212)	-0.119 (-1.147)	0.067 (0.432)	-0.513*** (-2.695)	0.686	1.027
1st Half								
CAPM	0.144** (2.333)	0.936*** (9.228)					0.522	0.923
FF3	0.131** (2.150)	1.066*** (8.927)	-0.139 (-0.597)	-0.388** (-1.964)			0.550	0.868
FF5	0.161** (2.552)	1.014*** (8.080)	-0.143 (-0.619)	-0.100 (-0.441)	0.105 (0.324)	-1.006** (-2.449)	0.584	1.107
2nd Half								
CAPM	0.100*** (4.453)	0.996*** (23.986)					0.881	0.801
FF3	0.077*** (4.163)	1.071*** (29.468)	-0.196*** (-3.328)	-0.184*** (-3.913)			0.923	0.740
FF5	0.079*** (4.310)	1.048*** (27.069)	-0.203*** (-2.999)	-0.117** (-2.066)	-0.024 (-0.250)	-0.227** (-2.281)	0.928	0.839
Panel B: Hedge Value-Weight Returns								
Sample Period: April 2008 to August 2021								
CAPM	0.093*** (2.614)	-0.155** (-2.513)					0.039	0.737
FF3	0.061* (1.885)	-0.044 (-0.711)	-0.004 (-0.039)	-0.539*** (-5.951)			0.233	0.542
FF5	0.064* (1.914)	-0.045 (-0.698)	-0.024 (-0.206)	-0.541*** (-5.053)	-0.085 (-0.537)	0.027 (0.137)	0.234	0.568
1st Half								
CAPM	0.103* (1.679)	-0.212** (-2.100)					0.054	0.662
FF3	0.075 (1.329)	-0.018 (-0.163)	0.045 (0.211)	-0.779*** (-4.296)			0.241	0.535
FF5	0.086 (1.445)	-0.057 (-0.473)	0.010 (0.045)	-0.855*** (-3.964)	-0.357 (-1.146)	0.174 (0.438)	0.255	0.625
2nd Half								
CAPM	0.077** (2.162)	-0.078 (-1.191)					0.018	0.883
FF3	0.044 (1.451)	-0.005 (-0.085)	-0.092 (-0.963)	-0.375*** (-4.901)			0.321	0.611
FF5	0.047 (1.532)	0.007 (0.114)	-0.147 (-1.291)	-0.336*** (-3.549)	-0.148 (-0.914)	-0.034 (-0.204)	0.329	0.658

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

VIII. **Conclusions**

The inherent complexity of investor sentiment ushered academics to explore a wide range of approaches and proxies, such as social media and specifically Twitter microblogging, recently adopted as a media-based investor sentiment source. This paper contributes new and substantial evidence, dissecting the pervasiveness and consistency of 1-month lagged Twitter sentiment association to stock returns for the S&P500 constituents from 2008 to 2021.

The formation of tercile portfolios considering monthly sorts on the anomaly variable and the corresponding returns shows interesting findings, suggesting a clear and consistent variation pattern of the returns across the scope of the anomaly, supporting a pervasive and positive association with stock returns. Results are more robust for value-weight returns, indicating that this association is stronger for high market capitalization stocks.

Additionally, long-only and hedge portfolios were derived from the extreme terciles and regressed against state-of-the-art asset pricing models, namely the CAPM, FF3, and FF5, imposing a functional form on the relation. The cross-section regressions on the long-only portfolio produce sufficient evidence to support the hypothesis that high 1-month lagged Twitter sentiment predicts stock returns, yielding alphas ranging from 10.5% to 12.3%. In contrast, the hedge returns produce weak results insufficient to reject the hypothesis that low 1-month lagged Twitter sentiment does not predict stock returns.

Limitations and Future Research: This paper encourages future research to extend and challenge these results, tackling the limitations identified. Such as the unrepresentativeness of Twitter microblogging for equities associated with low capitalization, possibly jeopardizing the application of this approach for a broader stock universe such as the Russel 3000. And the low accuracy of lexicon-based NLP methodologies at predicting extreme low sentiment. Another interesting future approach would be to adopt Carhart's 4-factor model as an additional asset pricing model, exploring the relation between Momentum and Twitter sentiment.

References

Augustyniak, Lukasz, Tomasz Kajdanowicz, Piotr Szymanski, Wlodzimierz Tuliglowicz, Przemyslaw Kazienko, Reda Alhajj, and Boleslaw Szymanski. 2014. "Simpler Is Better? Lexicon-Based Ensemble Sentiment Classification Beats Supervised Methods." *ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, October, 924–29. https://doi.org/10.1109/ASONAM.2014.6921696.

Baker, Malcolm, and Jeffrey Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns." *Journal of Finance* 61 (4): 1645–80. https://doi.org/10.1111/J.1540-6261.2006.00885.X.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics* 49 (3): 307–43. https://doi.org/10.1016/S0304-405X(98)00027-0.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2 (1): 1–8. https://doi.org/10.1016/J.JOCS.2010.12.007.

Bondt, Werner F.M. de, and Richard Thaler. 1985. "Does the Stock Market Overreact?" *The Journal of Finance* 40 (3): 793–805. https://doi.org/10.1111/J.1540-6261.1985.TB05004.X.

Bondt, Werner F.M. de, and Richard H. Thaler. 1987. "Further Evidence On Investor Overreaction and Stock Market Seasonality." *The Journal of Finance* 42 (3): 557–81. https://doi.org/10.1111/J.1540-6261.1987.TB04569.X.

Bukovina, Jaroslav. 2016. "Social Media Big Data and Capital Markets—An Overview." *Journal of Behavioral and Experimental Finance* 11 (September): 18–26. https://doi.org/10.1016/J.JBEF.2016.06.002.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *The Journal of Finance* 52 (1): 57–82. https://doi.org/10.1111/J.1540-6261.1997.TB03808.X.

Da, Z., J. Engelberg, and P. Gao. 2015. "The Sum of All FEARS Investor Sentiment and Asset Prices." *Review of Financial Studies* 28 (1): 1–32. https://doi.org/10.1093/rfs/hhu072.

Dhaoui, C., C.M. Webster, and L.P. Tan. 2017. "Social Media Sentiment Analysis: Lexicon versus Machine Learning." *Journal of Consumer Marketing* 34 (6): 480–88. https://doi.org/10.1108/JCM-03-2017-2141.

Fama, Eugene F. 1998. "Market Efficiency, Long-Term Returns, and Behavioral Finance." *Journal of Financial Economics* 49 (3): 283–306. https://doi.org/10.1016/S0304-405X(98)00026-9.

Fama, Eugene F., and Kenneth R. French. 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.

———. 2008. “Dissecting Anomalies.” *Journal of Finance* 63 (4): 1653–78. https://doi.org/10.1111/J.1540-6261.2008.01371.X.

———. 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.

Figueroa, Rosa L., Qing Zeng-Treitler, Sasikiran Kandula, and Long H. Ngo. 2012. “Predicting Sample Size Required for Classification Performance.” *BMC Medical Informatics and Decision Making* 12 (1). https://doi.org/10.1186/1472-6947-12-8.

Hentschel, M., and O. Alonso. 2014. “Follow the Money: A Study of Cashtags on Twitter.” *First Monday* 19 (8). https://doi.org/10.5210/fm.v19i8.

Hutto, C.J., and E. Gilbert. 2014. “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.” In *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014,* 216–25.

Koratamaddi, P., K. Wadhwani, M. Gupta, and S.G. Sanjeevi. 2021. “Market Sentiment-Aware Deep Reinforcement Learning Approach for Stock Portfolio Allocation.” *Engineering Science and Technology, an International Journal* 24 (4): 848–59. https://doi.org/10.1016/j.jestch.2021.01.007.

Lee, Charles, Andrei Shleifer, and Richard H. Thaler. 1991. “Investor Sentiment and the Closed-End Fund Puzzle.” *The Journal of Finance* 46 (1): 75–109. https://doi.org/10.1111/J.1540-6261.1991.TB03746.X.

Lee, Wayne, Christine Jiang, and Daniel Indro. 2002. “Stock Market Volatility, Excess Returns, and the Role of Investor Sentiment.” *Journal of Banking and Finance* 26 (12): 2277–99. https://doi.org/10.1016/S0378-4266(01)00202-3.

Lintner, J. 2006. *Separation Theorems: The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets*. *Stochastic Optimization Models In Finance (2006 Edition)*. https://doi.org/10.1142/9789812773654_0010.

Liu, B., and L. Zhang. 2012. *A Survey of Opinion Mining and Sentiment Analysis*. *Mining Text Data*. Vol. 9781461432. https://doi.org/10.1007/978-1-4614-3223-4_13.

Loughran, T., and B. Mcdonald. 2011. “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks.” *Journal of Finance* 66 (1): 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x.

Oliveira, Nuno, Paulo Cortez, and Nelson Areal. 2017. “The Impact of Microblogging Data for Stock Market Prediction: Using Twitter to Predict Returns, Volatility, Trading Volume and Survey Sentiment Indices." *Expert Systems with Applications* 73 (May): 125–44. https://doi.org/10.1016/J.ESWA.2016.12.036.

Pak, A., and P. Paroubek. 2010. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." In *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 1320–26.

Schmeling, Maik. 2009. "Investor Sentiment and Stock Returns: Some International Evidence." *Journal of Empirical Finance* 16 (3): 394–408. https://doi.org/10.1016/J.JEMPFIN.2009.01.002.

Sharpe, W.F. 1964. "CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK." *The Journal of Finance* 19 (3): 425–42. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x.

Sprenger, Timm O., Philipp G. Sandner, Andranik Tumasjan, and Isabell M. Welpe. 2014. "News or Noise? Using Twitter to Identify and Understand Company-Specific News Flow." *Journal of Business Finance and Accounting* 41 (7–8): 791–830. https://doi.org/10.1111/JBFA.12086.

Sprenger, Timm O., Andranik Tumasjan, Philipp G. Sandner, and Isabell M. Welpe. 2014. "Tweets and Trades: The Information Content of Stock Microblogs." *European Financial Management* 20 (5): 926–57. https://doi.org/10.1111/J.1468-036X.2013.12007.X.

Sul, Hong Kee, Alan R. Dennis, and Lingyao Ivy Yuan. 2017. "Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns." *Decision Sciences* 48 (3): 454–88. https://doi.org/10.1111/DECI.12229.

Tetlock, P.C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance* 62 (3): 1139–68. https://doi.org/10.1111/j.1540-6261.2007.01232.x.

Verma, B., and R.S. Thakur. 2018. *Sentiment Analysis Using Lexicon and Machine Learning-Based Approaches: A Survey*. *Lecture Notes in Networks and Systems*. Vol. 34. https://doi.org/10.1007/978-981-10-8198-9_46.

Appendix

Figure IV
Methodology Schematic

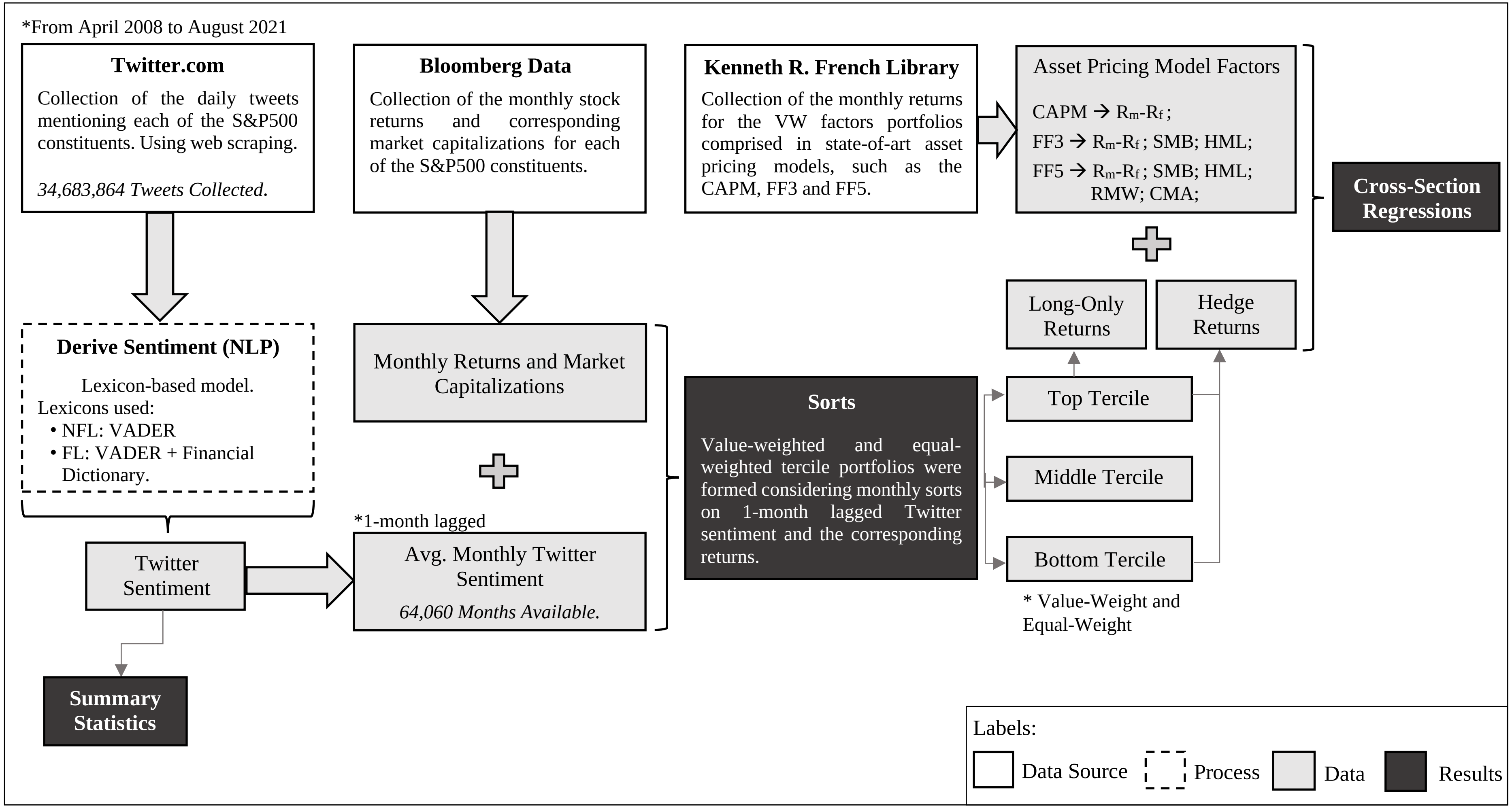


Table VI
Average Annualized Returns, Standard Deviations and Sharpe Ratios for Excess Returns on the Market, EW Long and EW Hedge Portfolios, from April 2008 to August 2021

Details[1]

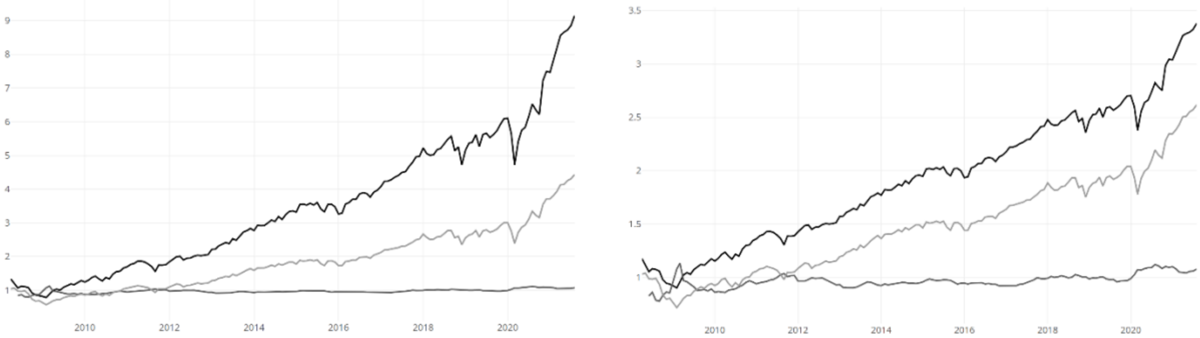
	Average Annualized Returns			Average Annualized Standard Deviations			Sharpe Ratio		
	Long	Hedge	Rm-Rf	Long	Hedge	Rm-Rf	Long	Hedge	Rm-Rf
Sample Period	0.18	0.01	0.13	0.19	0.09	0.17	0.97	0.11	0.77
1st Half	0.21	0.00	0.11	0.22	0.00	0.18	0.94	0.02	0.61
2nd Half	0.16	0.02	0.14	0.16	0.05	0.15	1.03	0.41	0.95

Figure V
Cumulative Excess Returns for the Market, EW Long and EW Hedge Portfolios April 2008 to August 2021.

Details[1]

Panel A: No Target Volatility	**Panel B:** 10% Target Volatility

Equal-Weighted Portfolios (EW)



Long-Only — Long-Short (Hedge) — Rm-Rf

1

Long-Only Portfolio corresponds to the Top Portfolio, 1st tercile returns (Figure II).
Long-Short (Hedge) Portfolio hedges the Top Portfolio (1st tercile) against the Bottom Portfolio (3rd tercile).
$Hedge\_Returns_n = Top\_Returns_n - Bottom\_Returns_n$.
Rm-Rf from Kenneth R. French: "the excess return on the market, value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t, good shares and price data at the beginning of t, and good return data for t minus the one-month Treasury bill rate."

Table VII
Cross-Section Regressions Results for Equal-Weight Long-Only and Hedge Returns, from April 2008 to August 2021.

	α	$R_M - R_F$	SMB	HML	RMW	CMA	R^2	$Info.$
Panel A: Long-Only Equal-Weight Returns								
Sample Period: April 2008 to August 2021								
CAPM	0.059**	0.961***					0.740	0.607
	(2.156)	(16.810)						
FF3	0.061**	1.050***	-0.155	-0.261***			0.740	0.623
	(2.173)	(17.151)	(-1.399)	(-2.906)				
FF5	0.064**	1.007***	-0.139	-0.119	0.067	-0.513***	0.744	0.662
	(2.234)	(16.031)	(-1.212)	(-1.147)	(0.432)	(-2.695)		
1st Half								
CAPM	0.102*	0.999***					0.629	0.762
	(1.926)	(11.491)						
FF3	0.101*	1.044***	-0.141	-0.051			0.632	0.757
	(1.876)	(9.928)	(-0.685)	(-0.291)				
FF5	0.118**	1.012***	0.091	-0.148	0.007	-0.508	0.641	0.899
	(2.073)	(8.915)	(-0.71)	(0.444)	(0.023)	(-1.368)		
2nd Half								
CAPM	0.015	1.035***					0.959	0.457
	(1.13)	(42.968)						
FF3	0.022*	1.008***	0.081**	0.048			0.964	0.720
	(1.724)	(40.802)	(2.014)	(1.516)				
FF5	0.021*	0.981***	0.076**	0.115**	0.088	-0.162**	0.968	0.714
	(1.672)	(37.831)	(2.531)	(2.007)	(1.373)	(-2.425)		
Panel B: Hedge Equal-Weight Returns								
Sample Period: April 2008 to August 2021								
CAPM	0.027	-0.145***					0.069	0.313
	(1.112)	(-3.401)						
FF3	0.01	-0.073	-0.066	-0.274***			0.186	0.124
	(0.433)	(-1.64)	(-0.819)	(-4.191)				
FF5	0.018	-0.078*	-0.277	-0.118***	-0.221*	0.056	0.207	0.224
	(0.753)	(-1.679)	(-1.395)	(-3.618)	(-1.941)	(0.397)		
1st Half								
CAPM	0.022	-0.215***					0.096	0.192
	(0.488)	(-2.88)						
FF3	0.009	-0.058	-0.217	-0.408***			0.220	0.084
	(0.209)	(-0.676)	(-1.308)	(-2.939)				
FF5	0.031	-0.13	-0.548*	-0.283***	-0.663***	0.316	0.305	0.309
	(0.714)	(-1.479)	(-1.763)	(-3.483)	(-2.919)	(1.092)		
2nd Half								
CAPM	0.025	-0.054*					0.039	0.622
	(1.523)	(-1.759)						
FF3	0.011	-0.031	0.006	-0.188***			0.316	0.324
	(0.769)	(-1.081)	(0.132)	(-5.146)				
FF5	0.013	-0.024	-0.162	-0.028***	-0.093	-0.029	0.331	0.385
	(0.897)	(-0.762)	(-0.518)	(-3.593)	(-1.198)	(-0.368)		

***p < 0.01, **p < 0.05, *p < 0.1