The Effects of Social Media Attention and Sentiment on Initial Public Offering Returns

Aiwen Li[1] and Cierra Beck[#]

[1]Pittsford Sutherland High School, Pittsford, NY, USA
[#]Advisor

ABSTRACT

Going public is a monumental step for many companies. Not only does it increase a company's legitimacy in the business community, but it also gives opportunities for them to harness significant amounts of investment capital. However, entrepreneurs and investors often face uncertainty due to the unpredictable nature of initial public offerings (IPOs). This study evaluated the impact of the amount and sentiment of Twitter activity on stock returns using data from domestic companies who went public from June 1, 2020 - May 31, 2021. Overall market behavior, company size, and community population demographics and social economic status of the companies' headquarters locations were controlled. The analyses showed that Twitter activity is associated with higher returns during relatively long-term time frames. The results are relevant for companies and potential IPO investors to predict and maximize their profits in the market. They also open opportunities for future research to investigate more in-depth regarding social factors in relation to IPOs.

Introduction

An initial public offering (IPO) is a significant stepping stone for many companies to increase their legitimacy in the business community, as well as to provide them opportunities to expand their capital to public investors in the market (Deeds et al., 1997; Krinitz & Neumann, 2021). Private companies frequently fail due to insufficient funding and heavy debt loads, but issuing an IPO allows them to overcome these difficulties (Deeds et al., 1997). Returns, or profits, gained from the stock market can be used to pursue projects in research, development, growth, and expansion (Deeds et al., 1997). Unfortunately, companies are typically affected by information asymmetry between themselves and the public, augmenting the difficulties of establishing an appropriate value for their shares and predicting their future market performance (Gian et al., 2020).

In recent years, social media has become an abundant source of information for both companies and potential investors. Companies often share information through social media because of the convenient nature of online platforms; social media sites are easily accessible, cost-effective, and capable of overcoming geographical barriers (Guijarro et al., 2019). Activity by public users, including potential investors, on social media platforms like Twitter are increasingly influencing the behavior of outside systems such as the stock market (Ranco et al., 2015).

The aim of this paper is to determine the association between social media activity on Twitter and companies' IPO stock returns in order to link gaps in the information asymmetry between companies and public investors (Wu et al., 2014). This question is addressed by analyzing Twitter activity using web scraping, natural language processing, and regression analysis with respect to companies' stock market returns, in which companies' sizes and headquarter location's community characteristics are controlled. In doing so, this research has potential implications for the ways in which companies and public investors decide to invest, trade, acquire, and share information, especially through social media platforms.

This paper is structured as follows: Section 2 provides background information regarding the methods used in this study, as well as previous studies on related topics in finance, business, and social media; section 3 outlines the main questions addressed by this research; section 4 describes the utilized data and statistical analysis methods; section 5 discusses the results from the analyses; and section 6 concludes the study and provides outlooks for future research.

Literature Review

Extant Research on IPOs

Over the past twenty years, there have been 2,756 Initial Public Offerings (IPOs) generating a respectable combined gross earning of approximately $665,608,000 (Chang et al., 2017). However, IPOs are a complex process with multiple stages (Fig. 1), typically taking years of preparation (Guilherme, n.d.). Entrepreneurs hire financial specialists, known as underwriters, to work with the company in determining the initial share price based on factors such as the company's profitability, growth trends, competition, and investor interest and confidence (Benveniste & Spindt, 1989). Benveniste and Spindt investigates how information yielded from underwriters' IPO marketing process influence the IPO's initial pricing (Benveniste & Spindt, 1989). In addition, the Securities and Exchange Commission (SEC) requires companies going public to formally register by filing the S-1 and 424B forms, also known as prospectuses, which contain detailed snapshots of the public offering (Tao et al., 2018). Loughran and McDonald (Loughran & McDonald, 2013), Krinitz and Neumann (Krinitz & Neumann, 2021), and Tao et al. (Tao et al., 2018) analyze the language and sentiment of these filings in regards to stock market returns, offer price revisions, and volatility.

Figure 1: Stages of the Initial Public Offering process.

Moreover, companies also need to maintain marketing campaigns in order to generate interest in their offering and maintain a positive image to public investors. The marketing process partially leans on influential social media platforms such as Twitter. Thus, this paper will analyze the impacts of Twitter activity and IPO returns. Surprisingly, there have been limited studies devoted to the relationship between social media and IPO valuation, possibly due to the seemingly different and unrelated fields of mass communication and finance (Gian et al., 2020). Guilherme uses genetic algorithms, based on Twitter sentiment, to propose an investment simulation system for investors to achieve the best gains possible during an IPO period (Guilherme, n.d.). Additionally, Gian et al. use a Stochastic Frontier Analysis (SFA) model to estimate the effects of Twitter activity on IPO first-day offer price revisions (Gian et al., 2020). Offer price revisions are defined as the distance between the maximum achievable price and the actual offer price of a stock, a similar measurement as stock returns (Gian et al., 2020). Furthermore, in a similar vein to this study, Liew and Wang conduct research associating Twitter sentiment with first-day IPO returns (Liew & Wang, 2015). They document a positive relationship between prior days' (1, 2, and 3-day) tweet sentiment and first-day IPO returns from offering price to opening price, as well as a negative relationship between tweet sentiment and first-day IPO returns from opening price to closing price (Liew & Wang, 2015).

Sentiment Analysis
Sentiment analysis, also known as opinion mining, is a computational field of natural language processing (NLP) used to understand and extract opinions, sentiments, attitudes, and emotions on a given subject (Guijarro et al., 2019; Kusumawati et al., 2019). It helps derive information about the positive and negative aspects of a particular entity by

extracting data from text documents (M. et al., 2016; Narayanan et al., 2013). Research using sentiment analysis is carried out among various fields. For instance, Narayanan et al. conduct sentiment analysis to analyze film review opinions by Indian audiences, giving potential insight for production companies, film studios, and movie theaters in their business (Narayanan et al., 2013). Additionally, Athawale and Gohil classify tweets as spam or non-spam, potentially helping to reduce unwanted spam and fraud within the Twitter platform (Ashwini Athawale & Deepali M. Gohil, 2018).

There are currently two main methods for sentiment analysis (M. et al., 2016). The first is a supervised learning approach, which generally involves two sets of documents: a training dataset and a testing dataset. The training set is used to train the classifier and prepare it to analyze the testing set. The Naive-Bayes classifier is a commonly used supervised learning algorithm, and was also used in the two studies above. Another sentiment analysis method is unsupervised learning. Instead of using two separate datasets, the unsupervised approach uses manually collected dictionaries (sometimes referred to as lexicons or corpuses) containing words with a known positive or negative polarity (M. et al., 2016; Jagdale et al., 2016). Sentiment analysis is then performed on a dataset using information from the dictionary.

Regression Analysis

Researchers often use regression analysis to assess whether there is a relationship, or association, between multiple variables, as well as to measure the strength of the association (Vetter & Schober, 2018). There are several different regression analysis techniques based on the number and type of variables, which Vetter and Schober discuss in detail (Vetter & Schober, 2018). Even though the various regression analysis methods do not prove causation, they are still powerful statistical tools which can allow for interpretations of complex multi-factorial data when applied appropriately (Vetter & Schober, 2018). Thus, it is widely applied in various fields. For instance, Gerbershagen et al. use multiple linear and logistic regression analyses to assess the association between age, sex, and postoperative pain intensity among various surgical procedures in German hospitals (Gerbershagen et al., 2014). Additionally, Bartz-Beielstein and Markon use regression analysis and modeling with decision trees to optimize customer service quality with elevator controllers in modern Hong Kong buildings (Bartz-Beielstein & Markon, 2004).

Social Media in Finance

Twitter, a microblogging social network, is one of the most popular communication sources globally with over 350 million active users interacting in more than forty languages. Twitter users can interact with each other by sending short messages known as "tweets," which can contain text, emoticons, links, photos, and videos. They can also include "hashtags" (#) - a symbol indicating a key word or phrase to help other users connect and find similar content. "Cashtags" ($) are a special symbol with a similar keyword functionality as the hashtag, except that the cashtag is used specifically for financial ticker symbols linked to the stock market. Additionally, users can "follow" other accounts, "retweet" information, "reply" to another tweet, and "like" a tweet to show support. The various data that Twitter can provide are therefore often used by companies to collect opinions and feedback from a large audience (Guijarro et al., 2019).

Before launching an IPO, a company typically uses social media platforms, including Twitter, to announce its plans before the IPO date. This often leads to much activity and responses from public social media users, providing a rich data source which can be analyzed in regards to many external variables.

Many studies have been conducted to investigate Twitter influence in the finance sector. Albrecht et al. investigate how different social media networks are linked to the capitalization of Initial Coin Offerings (ICO), a fund-raising method used by companies in the cryptocurrency/blockchain industry similar to an IPO (Albrecht et al., 2019). In addition, Guijarro et al. study the impact of Twitter sentiment on stock market liquidity with the S&P 500 index (Standard & Poor's 500) and trading costs based on bid-ask spreads (Guijarro et al., 2019). Ranco et al. investigate the relationship between Twitter activity, specifically volume and sentiment of tweets, and the behavior of the thirty

stock companies that form the Dow Jones Industrial Average (DJIA) index (Ranco et al., 2015). These studies, however, only focused on historical impacts of Twitter activity on IPOs.

Significance

This paper contributes significantly to the existing literature outlined above. Not only does this research extend the Twitter data timeline to within 30 days before and after the IPO; it also analyzes IPO returns in periods of 1-day, 1-week, and 1-month (instead of limiting to 1-day returns), according to the amount and sentiment of Twitter activity within the same timeline. To control for external factors, this study also includes several unique variables in the regression analyses to control for overall market activity, company size, and the socioeconomic status/demographics of each company's headquarters location by zip code. To the authors' knowledge, no research has been conducted using a similar combination of time frames and control variables.

Methods

Data Sources

This study used published Nasdaq IPO data to capture 723 companies that went public between June 1, 2020 and May 31, 2021. For each company, the company name, ticker symbol and IPO date were collected from the website. Companies' share prices in the stock market within thirty days of their IPOs were then retrieved using the Python Yfinance library. Daily S&P 500 indices, downloaded from the Federal Reserve Economics Data (FRED) database, were also merged with the study's data sample to control for the overall time-variant market activities. After removing 195 companies that did not have complete financial data, 528 companies remained in the study.

      In addition to financial data, the study obtained data of demographics and social economic status in companies' headquarter locations to control for community characteristics. This data was captured from the US Census 2018 American Community Survey database. After 43 international companies were removed from the database, the final data included 485 domestic companies.

      Furthermore, the study scraped Twitter data to capture IPO-associated Twitter activity and sentiments. Raw Tweets were obtained using Python's Selenium package based on keywords with company ticker symbols and dates.

Dependent Variables

To measure each companies' stock returns, three individual day returns were measured as ratios of the closing to the opening prices on day 1, day 7, and day 30 of the IPO, multiplied by 100. In addition, long-term returns were measured as ratios of the closing prices at the end of the first week and first month to the opening price on day 1, multiplied by 100.

Independent Variables

The independent variables in this study are the amount of IPO-associated Twitter activity and their sentiments.

      Using the scraped Twitter data, five stages of text pre-processing were conducted using the Natural Language Toolkit (NLTK) in Python: cleaning, tokenizing, case folding, stop word removal, and lemmatizing (Fig. 2). Since this study only used the textual components for sentiment analysis, the data was *cleaned* of all unnecessary punctuation marks, HTML tags, hyperlinks, Twitter usernames preceded by "@" symbols, hashtag symbols ("#"), and cashtag symbols ("$"). Each tweet was then *tokenized* by splitting the text into separated single words named tokens. *Case folding* was then conducted to convert all characters into lowercase. *Stop words*, or words without a deep meaning (such as "the", "is", "of", etc.), were also removed. Finally, each word was *lemmatized*, or reduced to their stem, if applicable (for instance, "running" is converted to "run"). These pre-processing steps are consistent with several other

studies (Krinitz & Neumann, 2021; Ranco et al., 2015; Kusumawati et al., 2019; Loughran & McDonald, 2013; Guilherme, n.d.).
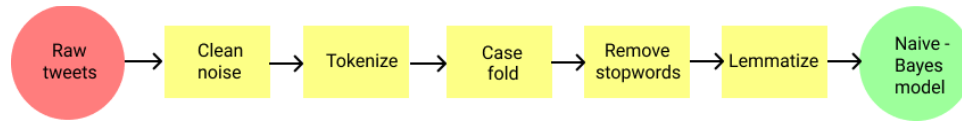
Figure 2: Pre-processing Twitter data for sentiment analysis

The Naive-Bayes classifier, a supervised learning approach and part of the NLTK package in Python, was then used for sentiment analysis to classify each body of text as either "positive" or "negative" (M. et al., 2016). The Naive-Bayes model involves a simplifying conditional independence assumption: Given a class (positive or negative), words are conditionally independent of each other, hence the term "naive" (Narayanan et al., 2013). The model also works with the Bag of Words (BOW) feature extraction which ignores the position of a particular word in the document (M. et al., 2016). The model uses the Bayes Theorem to predict the probability of a given word or feature belonging to a particular class (positive or negative) based on data from the training phase (Kusumawati et al., 2019).

In this study, the Naive-Bayes model was trained using built-in Twitter data from the NLTK package, which includes 10,000 pre-labeled tweets. During the training phase, the frequencies of the words were stored in tables to be used in classifying the testing data (Narayanan et al., 2013). Then, each tweet that was scraped from the Twitter website went through the newly built Naive-Bayes classifier to be labeled as either positive or negative. For each search query, the mean sentiment was calculated by assigning values of +1 for positive tweets or -1 for negative tweets, and then finding the average by summing all the values and dividing by the number of tweets.

By using these features, the sentiment and number of tweets was recorded in the dataset for each company and each time frame. Corresponding to the dependent variables, these independent variables were defined as: total number of tweets and average sentiment up to 30 days before the IPO date, up to 7 days after the IPO date, and up to 30 days after the IPO date. These times are used to associate Twitter activity with IPO returns on day 1, day 7 (or week 1), and day 30 (or month 1). The number of tweets were further converted into a binary variable (1 or 0) with the median being the cutoff due to skewed frequencies.

Covariates

Following previous research, the analyses included several control variables (Deeds et al., 1997; Krinitz & Neumann, 2021; Gian et al., 2020; Albrecht et al., 2019; Tao et al., 2018; Loughran & McDonald, 2013).

The *S&P 500 index*, or Standard & Poor's 500 index, is a weighted index of the 500 largest publicly-traded companies in the United States and is used to control for overall market behavior on companies' IPO returns. Corresponding to the dependent variables, S&P 500 index ratios for day 1, day 7, day 30, week 1, and month 1 were computed.

An IPO's *offer amount* is equal to the product of the IPO price per share multiplied by the total number of shares of stock sold at the IPO. This amount is a measure of an IPO company's size.

IPO companies' headquarters demographics and socioeconomic status were obtained from the 2014-2018 cumulative data collected from the U.S. Census American Community Survey (ACS) database. Specifically, these variables were defined at the zip code level as: percent elderly (aged 65 years and over), percent college graduates, percent poverty, population (per 1,000 residents), and population density (per 1,000 residents per square mile).

Statistical Analyses

Univariate analyses were performed to examine data distributions. Mean and standard deviations were examined for continuous variables; when the distributions of continuous variables were skewed, median and inter-quartile range (IQR) were presented instead. Counts were also presented for categorical variables.

Bivariate analyses were then performed to examine the crude associations between IPO returns and Twitter activity/sentiment, in which simple linear regression models were performed for linear associations, and two-sample independent t-tests were performed for binary outcomes.

The study fitted separate linear regression models to examine the independent relationships between social media activity and IPO stock returns, after controlling for covariates. Specifically, each of the five dependent variables was modeled in separate regression models: IPO return ratios of day 1, day 7, day 30, week 1, and month 1. The independent variables in each model were Twitter activity and sentiment, and covariates included the S&P 500 index ratios of the corresponding time frames, companies' offer amounts, and headquarter locations' characteristics.

All analyses were performed in Python at a 0.05 two-tailed significance level.

Table 1: Study variables, definitions, and data sources.

Variable Type	Variable	Definition	Data Source
Dependent	IPO Returns	Ratio between close and open prices	Yahoo Finance
Independent	Sentiment	Mean sentiment value per company and per time frame	Twitter
Independent	Number of tweets	Number of tweets per company and time frame	Twitter
Control	S&P 500	Change in S&P 500 indices corresponding to each time frame	FRED
Control	Offer amount	IPO price per share multiplied by total number of shares	Nasdaq IPO Calendar
Control	Percent bachelor	Percent residents who are a college graduate or above	U.S. Census ACS
Control	Percent poverty	Percent residents who live below the poverty line	U.S. Census ACS
Control	Percent elderly	Percent residents aged 65 or above	U.S. Census ACS
Control	Population per K	Total population per 1,000 people	U.S. Census ACS
Control	Density per K	Total population density per 1,000 residents, per square mile	U.S. Census ACS

Results and Discussion

Table 2 shows the univariate analyses results. Day 1 IPO returns varied from 70.48% to 165.55% at the closing prices comparing to same-day opening prices, and long-term IPO returns (Month 1) varied from 44.08% to 215.38%. The distributions of IPO returns for all five time periods seem fairly symmetric since the mean and median are close together. However, the variability in the distribution of 1 month returns is relatively large, given the irregularly high IQR and standard deviation (Std). The increase in variability of returns may simply be because the magnitude at which companies grow or decay increases with time.

Past 30-day Tweet counts varied from 2 to 2904 among those IPO companies, 1-week Tweet counts after the IPO varied from 8 to 2417, and post 1-month Tweet counts varied from 9 to 2983. The skewedness in Tweet counts are shown by the large differences between the mean and median. Additionally, the average sentiments were 0.49 ± 0.33, 0.47 ± 0.31, and 0.41 ± 0.27 for past 30-day, post-1-week, and post-1-month Tweets, respectively. The positive

values for Tweet sentiments' mean and median suggest that Twitter tends to have more positive than negative tweets regarding company IPOs.

Table 2: Descriptive Statistics

Variable	Mean	Median	Standard deviation	Interquartile range	Minimum	Maximum
Day 1 Return	100.83	100.00	10.45	2.69	70.48	165.55
Day 7 Return	100.06	99.90	4.19	1.79	84.95	129.48
Day 30 Return	99.57	99.95	3.60	2.05	81.85	115.75
Week 1 Return	101.65	100.50	14.76	4.33	61.33	184.07
Month 1 Return	101.32	100.30	18.50	7.41	44.08	215.38
Tweet count past 30 day	154.52	29.00	397.89	73.00	2	2904
Tweet count 1 week	134.73	38.00	314.23	85.00	8	2417
Tweet count 1 month	231.58	54.00	445.92	158.00	9	2983
Sentiment past 30 day	0.49	0.45	0.33	0.49	-0.44	1.00
Sentiment 1 week	0.47	0.45	0.31	0.45	-0.63	1.00
Sentiment 1 month	0.41	0.38	0.27	0.40	-0.79	1.00
SP500 day 1 ratio (%)	0.10	0.13	1.02	1.17	-6.26	2.55
SP500 day 7 ratio (%)	0.42	0.70	1.81	2.05	-6.28	4.90
SP500 day 30 ratio (%)	2.63	2.67	2.50	3.54	-5.92	9.88
SP500 week 1 ratio (%)	0.11	0.13	1.03	1.20	-6.26	2.16
SP500 month 1 ratio (%)	0.14	0.17	0.94	1.23	-3.64	2.32
Offer amount (million)	0.39	0.22	1.41	0.19	0.00	28.71
% Bachelor	57.16	60.90	17.90	23.40	3.60	100.00
% Poverty	10.41	8.30	7.37	7.60	0.00	45.00
% 65+	15.81	14.60	7.97	7.70	0.00	84.40
Pop. per 1000	26.10	25.76	16.10	21.00	0.01	122.81
Dens per 1000	18.19	5.41	25.94	16.19	0.00	141.79

Table 3 shows the crude associations between IPO returns and corresponding Twitter activities and sentiments. IPO companies having greater numbers of Tweets before the IPO date had higher one-week returns (p=0.024) as well as individual-day returns at day 30 of the IPO date (p=0.017). Furthermore, IPO companies having more one-

week Twitter activity after the IPO date also had higher one-week (p=0.020) and day 30 returns (p=0.036). Finally, IPO companies having more one-month Twitter activities after the IPO date had higher day 30 returns (p=0.025). The results suggest that more Twitter activity is generally associated with higher profits, specifically for one-week and day 30 returns, possibly due to greater investor interest. More social media activity may also lessen the information asymmetry between companies and public investors, helping IPO companies establish appropriate values for their shares with the goal of raising profit.

As for sentiment, companies with more positive Tweets within one week after their IPOs saw lower individual day 7 returns (p=0.043) and 1 week returns (p=0.042). These results, however, may be confounded by covariates not accounted for in bivariate analyses.

Table 3: Bivariate Analyses between Twitter Activity and IPO Returns

A: Past 30-day Tweets

Variable	Num Tweets < Median: Mean ± Std	Num Tweets > Median: Mean ± Std	t-value (num tweets)	p-value (num tweets)	Sentiment coef.	Sentiment p-value
Day 1 Return	100.1 ± 3.2	101.5 ± 14.3	-1.48	0.139	-0.6	0.665
Day 7 Return	99.8 ± 1.3	100.3 ± 5.8	-1.50	0.135	-0.7	0.221
Day 30 Return	100.0 ± 1.4	99.2 ± 4.9	2.40	0.017	0.5	0.335
Week 1 Return	100.1 ± 4.8	103.1 ± 20.1	2.28	0.024	-0.9	0.357
Month 1 Return	99.9 ± 6.0	102.7 ± 25.3	-1.64	0.103	-1.4	0.599

B: Post-1-week Tweets

Variable	Num Tweets < Median: Mean ± Std	Num Tweets > Median: Mean ± Std	t-value (num tweets)	p-value (num tweets)	Sentiment coef.	Sentiment p-value
Day 1 Return	-	-	-	-	-	-
Day 7 Return	99.8 ± 1.1	100.3 ± 5.8	-1.21	0.226	-1.2	0.043
Day 30 Return	99.9 ± 1.2	99.2 ± 4.9	2.10	0.036	0.5	0.327
Week 1 Return	100.1 ± 4.4	103.2 ± 20.3	-2.35	0.020	-4.4	0.042
Month 1 Return	100.1 ± 5.4	102.5 ± 25.6	-1.41	0.159	-2.6	0.329

C: Post-1-month Tweets

Variable	Mean ± Std: Fewer Tweets	Mean ± Std: More Tweets	t-value (num tweets)	p-value (num tweets)	Sentiment coef.	Sentiment p-value
Day 1 Return	-	-	-	-	-	-
Day 7 Return	-	-	-	-	-	-
Day 30 Return	99.9 ± 1.0	99.2 ± 5.0	2.26	0.025	0.8	0.189
Week 1 Return	-	-	-	-	-	-
Month 1 Return	100.7 ± 5.4	101.9 ± 25.6	-0.72	0.472	-1.7	0.584

Table 4 shows the results of the multivariate regression models of IPO returns with Twitter activities, when both company- and community-level confounders were controlled. The findings show that Twitter sentiment value is not associated with IPO returns, suggesting that the negative associations found in the bivariate analyses were confounded. However, the amount of past 30-day Twitter activity was independently associated with one-month IPO returns (P<0.001). This suggests that more Twitter activity leading up to an IPO tends to be associated with higher long-term profits, independent of the type of activity (positive or negative).

In addition, IPO companies with headquarters located in senior areas with more clustered populations had higher day 30 returns. Seniority in a population may indicate relatively greater prosperity, while clustered populations may indicate a well-known area, thus giving the company a better foundation to build off of.

Table 4: Summary of Multivariate Regression Analyses Results (stated: coefficients)

Variable	IPO Returns: Day 1	IPO Returns: Day 7	IPO Returns: Week 1	IPO Returns: Day 30	IPO Returns: Month 1
Past Tweets	1.196	0.739	0.202	-0.276	6.724**
Past Sentiment	0.541	0.393	4.608	-0.493	-0.132
Tweet + 7d	-	-0.517	2.715	-	-
Sentiment + 7d	-	-1.616	-5.695	-	-
Tweet + 30d	-	-	-	-0.342	-4.369
Sentiment + 30d	-	-	-	0.378	-0.067
SP500 d1	-19.594	-	-	-	-
SP500 d7	-	22.504	-	-	-
SP500 w1	-	-	-29.115		-
SP500 d30	-	-		13.420	-
SP500 m1	-	-	-	-	-109.206
Offer amt (per billion)	-0.449	0.035	-0.239	-0.105	-0.738
% Bachelor	-0.036	0.013	0.007	0.000	-0.005
% Elderly	-0.058	-0.004	-0.013	0.060***	-0.023
Pop. per 1000	-0.004	0.025	0.032	0.025**	0.037

***$p < 0.01$; **$p < 0.05$

Conclusion

In recent years, the impact of social media attention and sentiment on capital markets has developed as a mainstream of finance (Gian et al., 2020). This study examined how the financial success of companies' initial public offerings, or IPOs, are linked to social media activity. The analyses also control for several other variables to garner further insights into the overall market activity, IPO company size, and demographics and socioeconomic factors of company headquarters locations. In doing so, the study addresses several research gaps regarding social factors and associated stock market returns as outlined in previous research.

Bivariate and multivariate analyses were conducted to measure the associations between the variables. The bivariate analyses showed evidence of crude associations between the amount and sentiment of related Twitter activity with relatively long-term IPO returns. These results add to conclusions made by Gregori et.al and Liew and Wang who find that positive Twitter sentiment is associated with higher short-term IPO profits (Gian et al., 2020; Liew &

Wang, 2015). The multivariate analyses, when including the control variables, only found that the amount of Twitter activity within 1 month before the IPO date is associated with IPO returns after 1 month, again suggesting a more long-term association. Two control variables relating to demographics and socioeconomic statuses of company head-quarters were also found to be significant for day 30 returns.

The findings in this paper are relevant for public investors, entrepreneurs, business owners, and companies. To maximize profits from the public market, companies can use the results outlined above to develop an IPO strategy and predict their profits based on social factors relevant to their company. Similarly, public investors can maximize their stock market returns by strategizing their timing and amount of capital invested based on factors from the results. For instance, to reflect the results of this study, investors aiming for long-term success may strategize based on the amount of social media activity or any related community characteristics, while those aiming for relatively short-term success may focus more other factors not related to social media.

Limitations and Future Work

Entrepreneurs and public investors may take caution to the results due to limitations in the insights that this study pulled. Since using a dummy variable to compensate for skewedness in the number of tweets leads to results of a binary form, enhancing data on Twitter activity in more depth and detail may lead to different results. Additionally, a small portion of tweets are not made by users at all, but by stock market Twitter bots, and do not necessarily reflect overall public sentiment. Constructing a classifier to identify and filter out posts from Twitter bots would further maximize the quality of the Twitter data. Nonetheless, Twitter bots typically post about all companies who become public, which somewhat evens out the distributions of the Twitter variables.

Furthermore, this study only focuses on companies in the United States, so future research may extend the sample to include companies of other countries. After all, the nature of the relationship between social factors and business appears to be complex, and future research may add on to the progress presented in this paper (Liew & Wang, 2015).

Finally, as IPOs are characterized by high uncertainty, their returns can be unpredictable. For instance, in conventional markets, an increase in positive social media sentiment might be considered as beneficial to the company and be associated with positive returns. Nevertheless, an increased mean sentiment score was found to have negative effects on day 7 and 1-week IPO returns in the bivariate analyses, but was found to be not significantly associated with IPO returns in multivariate analyses. Additionally, short-term IPO returns were found to be unpredictable due to no evidence of its associations with any independent variables. These uncertainties may provide opportunities for future studies to evaluate IPOs with respect to other potential variables, which can include deeper analysis into different social media platforms or looking into entrepreneurs' and companies' detailed backgrounds (Albrecht et al., 2019).

Acknowledgments

References

1.   Albrecht, S., Lutz, B., & Neumann, D. (2019). *How Sentiment Impacts the Success of Blockchain Startups – An Analysis of Social Media Data and Initial Coin Offerings*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2019.551

2.   Ashwini Athawale, M., & Deepali M. Gohil, M. (2018). Spam Detection on Collection of Twitter Data Using Naive Bayes Algorithm. *International Journal of Innovative Research in Science, Engineering and Technology*, *7*(6). http://www.ijirset.com/upload/2018/june/60_10_Spam.PDF

3.   Bartz-Beielstein, T., & Markon, S. (2004). Tuning search algorithms for real-world applications: A regression tree based approach. *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)*, 1111–1118. https://doi.org/10.1109/CEC.2004.1330986

4.   Baucom, C., Bate, J., Ochoa, S., Santos, I., Sergios, A., Lorentzen, L., & Reilly, K. (2019). The Epidemiology of the AIDS Pandemic: Historical, Cultural, Political, Societal Perspectives and Knowledge of HIV. Journal of Student Research, 8(2). https://doi.org/10.47611/jsr.v8i2.781

5.   Benveniste, L. M., & Spindt, P. A. (1989). How investment bankers determine the offer price and allocation of new issues. *Journal of Financial Economics*, *24*(2), 343–361. https://doi.org/10.1016/0304-405X(89)90051-2

6.   Byrne, I., Kanaoka, Y., Pollack, N. E., Rhee, H. J., & Sommers, P. M. (2019). An Analysis of Airport Delays Across the United States, 2012-2018. Journal of Student Research, 8(2). https://doi.org/10.47611/jsr.v8i2.775

7.   Chang, C., Chiang, Y.-M., Qian, Y., & Ritter, J. R. (2017). Pre-market Trading and IPO Pricing. *The Review of Financial Studies*, *30*(3), 835–865. https://doi.org/10.1093/rfs/hhw032

8.   Deeds, D. L., Decarolis, D., & Coombs, J. E. (1997). The impact of firmspecific capabilities on the amount of capital raised in an initial public offering: Evidence from the biotechnology industry. *Journal of Business Venturing*, *12*(1), 31–46. https://doi.org/10.1016/S0883-9026(97)84970-1

9.   Gerbershagen, H. J., Pogatzki-Zahn, E., Aduckathil, S., Peelen, L. M., Kappen, T. H., van Wijck, A. J. M., Kalkman, C. J., & Meissner, W. (2014). Procedure-specific Risk Factor Analysis for the Development of Severe Postoperative Pain. *Anesthesiology*, *120*(5), 1237–1245. https://doi.org/10.1097/ALN.0000000000000108

10.   Gian, L. G., Luca, M., Camilla Mazzoli, & Sabrina, S. (2020). The social side of IPOs: Twitter sentiment and investors attention in the IPO primary market. *African Journal of Business Management*, *14*(12), 529–539. https://doi.org/10.5897/AJBM2020.9135

11.   Guijarro, F., Moya-Clemente, I., & Saleemi, J. (2019). Liquidity Risk and Investors' Mood: Linking the Financial Market Liquidity to Sentiment Analysis through Twitter in the S&P500 Index. *Sustainability*, *11*(24), 7048. https://doi.org/10.3390/su11247048

12.   Guilherme, B. (n.d.). *Investing in Stock IPOs with Sentiment Analysis from Twitter optimized by Genetic Algorithms*. https://www.semanticscholar.org/paper/Investing-in-Stock-IPOs-with-Sentiment-Analysis-by-Guilherme/23a12f6ab6cdb79b277d97476795ac1f3bb3ad9c#paper-header

13.   Jagdale, R., Shirsat, V., & Deshmukh, S. (2016). Sentiment Analysis of Events from Twitter Using Open Source Tool. *International Journal of Computer Science and Mobile Computing*, *5*(4), 475–485.

14.   Krinitz, J., & Neumann, D. (2021). *Market Engineering: Insights from Two Decades of Research on Market and Information*. Springer. https://link.springer.com/content/pdf/10.1007%2F978-3-030-66661-3.pdf

15.   Kusumawati, R., D'arofah, A., & Pramana, P. A. (2019). Comparison Performance of Naive Bayes Classifier and Support Vector Machine Algorithm for Twitter's Classification of Tokopedia Services. *Journal of Physics: Conference Series*, *1320*, 012016. https://doi.org/10.1088/1742-6596/1320/1/012016

16.   Liew, J. K.-S., & Wang, G. Z. (2015). Twitter Sentiment and IPO Performance: A Cross-Sectional Examination. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2567295

17.   Loughran, T., & McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, *109*(2), 307–326. https://doi.org/10.1016/j.jfineco.2013.02.017

18.   M., V., Vala, J., & Balani, P. (2016). A Survey on Sentiment Analysis Algorithms for Opinion Mining. *International Journal of Computer Applications*, *133*(9), 7–11. https://doi.org/10.5120/ijca2016907977

19.   Narayanan, V., Arora, I., & Bhatia, A. (2013). Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model. In H. Yin, K. Tang, Y. Gao, F. Klawonn, M. Lee, T. Weise, B. Li, & X. Yao (Eds.), *Intelligent Data Engineering and Automated Learning – IDEAL 2013* (Vol. 8206, pp. 194–201). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41278-3_24

20.   Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. *PLOS ONE*, *10*(9), e0138441. https://doi.org/10.1371/journal.pone.0138441

21.   Tao, J., Deokar, A. V., & Deshmukh, A. (2018). Analysing forward-looking statements in initial public offering prospectuses: A text analytics approach. *Journal of Business Analytics*, *1*(1), 54–70. https://doi.org/10.1080/2573234X.2018.1507604

22.   Vetter, T. R., & Schober, P. (2018). Regression: The Apple Does Not Fall Far From the Tree. *Anesthesia & Analgesia*, *127*(1), 277–283. https://doi.org/10.1213/ANE.0000000000003424

23.   Wu, D. D., Zheng, L., & Olson, D. L. (2014). A Decision Support Approach for Online Stock Forum Sentiment Analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *44*(8), 1077–1087. https://doi.org/10.1109/TSMC.2013.2295353