ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
Bachelor Thesis Economics & Business

The Effect of Absolute Search Volumes on Initial Public Offering Underpricing in the United States: An Empirical Analysis

Author: Zhang, M.
Student number: 531180
Thesis supervisor: Dr. Li, Y.
Second reader: Dr. de Bliek, R.
Email: 531180mz@eur.nl
Finish date: 28 June 2023

ABSTRACT

This dissertation examines the relationship between Google's absolute search volumes and initial public offering (IPO) underpricing in the United States (US) between 2020–2021. By analyzing daily data on IPOs, a regression analysis was conducted, considering both absolute and relative search volumes. The findings show that absolute search volumes have predictive power on IPO underpricing, with both significant positive and negative effects observed 28 days prior and 7 days before IPO releases. However, relative search volumes do not seem to exhibit the same relationship. These results suggest that the impact of search volumes on IPO underpricing varies depending on the sample, variables used and economic conditions. While high absolute search volumes are shown to significantly affect IPO underpricing, it may not apply uniformly to all IPOs, especially considering long-term implications.

Keywords: IPO, underpricing, search volumes, investors

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 Introduction

Initial public offerings (IPO) have traditionally suffered from 'underpricing', which is described by the difference between the initial offer price and the closing price on the first day of public trading for securities (Ibbotson, 1975). IPOs refer to the process in which private companies decide to offer their shares to the public, becoming a publicly traded company (Pastor & Veronesi, 2005). Underpricing in IPOs is a common phenomenon since underpricing was examined to be present in 48.9% of IPOs in 2022 in the United States (US) (Ritter, 2022). This leaves 'money on the table' for firms that go public in IPOs. This empirical study is conducted based on the years 2020–2021. This time horizon was taken as these years mark the consecutive breaking of the record of most IPOs annually after and during COVID-19. According to Mazunder and Saha (2021), COVID-19 fear decreased initial returns (IR) for IPOs in 2020, which begs the question of what caused the abnormality in IPO pricing during that time. Additionally, media attention and internet user engagement have gathered momentum, which could potentially explain deviations in IPO IRs in 2020-–2021 (Tetlock, 2015).

Prior research has found the relevance of using Search Volume Indexes (SVI), or search traffic, as a predictor of IPO underpricing. The effect of search volume is defined as the number of searches on an IPO before its inception and its influence on post-launch underpricing (Vakrman & Kristoufek, 2015). Additional studies have been conducted based on the effect of media attention on potential IPO underpricing. Bajo and Raimondo (2017) study this relationship between media attention in terms of underlying connotations of news items on IPO performance, whereas Vakrman and Kristoufek (2015) examined the plausible effect of search engine indexes on IPO underpricing. These papers found that an increase in media attention and SVIs lead to significant overperformance on first-day returns of IPOs. Particularly, Vakrman and Kristoufek (2015) have proved the effect of positive media sentiment on underpricing and the predictive power of SVIs.

In this paper, the strategy by Vakrman and Kristoufek (2015) is replicated, using absolute search volumes (ASV) input as compared to SVI in the original study. SVIs are used as a comparative measure in the analysis of this paper. The use of ASV input enables one to look at the existing research in a different light. Although absolute search volume is not commonly used in financial research, it can however extend our knowledge of between-firm variation. Using the SVI, we are inhibited to only research the within-firm variation as the values generated are normalized to a given time interval for a given query/firm (Bank, Larch & Peter, 2011). There is reason to believe that the results obtained by Vakrman and Kristoufek (2015) have changed over time and especially in this period as the market displayed disruptive characteristics caused by COVID-19. Baig and Chen (2022) argue that COVID-19 and its regulations have harmed the IPO market, which is defined as an increased IPO pricing volatility as well as the degree of underpricing for 'COVID-19 IPOs'. This was aggravated when the severity of both the pandemic and government regulations increased. If the findings do not hold in the given window, that means that IPO overperformance could be susceptible to shocks in relation to an increased number of IPOs, showing yet another variable affecting underpricing. Additionally, studying the effect of ASVs compared to SVIs adds to the understanding of indicators of IPO underpricing since ASVs are statistically different from the commonly used SVIs. As mentioned previously, ASVs allow us to look at the between-firm variation of IPOs which was not possible by using SVIs, since the latter does not produce any analyzable data. ASV thus shows a nuanced position in factors affecting underpricing, which may reveal more latent factors with a significant impact. Therefore, this research analyzes both SVI inputs and ASV inputs as variables of interest in two separate regressions. Ultimately, given the previous research on this topic, or the lack thereof, an unanswered question remains. *How do absolute search volumes affect IPO underpricing and is it different from using relative search volumes?*

The research method employed to answer this question is Ordinary Least Squares (OLS) over a cross-section of IPOs for the period 2020–2021 to determine the effect of both ASVs and SVIs on IPO underpricing. In addition to checking for the OLS assumptions, control variables are added to the regression to account for potential omitted variable bias (OVB). The unit of analysis is the firm level in the US. Furthermore, the operationalization of IPO underpricing and ASV was conducted as follows: IPO underpricing is determined based on the absolute difference between the offer price and the first-day closing price of individual IPOs in the US from 2020-2021 (in US Dollar terms). ASV is based on the 28 days before an IPO, as this has been concluded as a sensible window by Bajo and Raimondo (2017). One Google search query counts towards the total ASVs for each respective IPO. Furthermore, an emphasis is put on the 7 days before an IPO since media attention displays signs of doubling before its inception (Bajo & Raimondo, 2017). This window is operationalized in a separate variable to account for any differences that might occur. The sample used in this study contains around 427 IPOs from 2020-2021 as a similar selection approach was taken and found feasible by Bajo and Raimondo (2017) and Vakrman and Kristoufek (2015). The data on IPO IRs is retrieved from a combination of Yahoo Finance, Stock Analysis, and Scoops. Yahoo Finance is a reputable database of financial information, whereas Stock Analysis and Scoops are growing sources of financial knowledge. Yahoo Finance provides data on the offer price, financing type, and other inputs to ensure quality. Stock Analysis and Scoops provide IRs of IPOs between 2020–2021 in the US and both are consolidated by using the Yahoo Finance dataset. Additionally, Keywords Everywhere contains data on ASVs and is thus used as a proxy for investor attention.

This dissertation found that an increase in ASVs pre-IPO is causing an increase in IPO underpricing in the US between 2020–2021, if search volumes increased 7 days prior to release. The opposite effect holds for search volume numbers 28 days before release. The results found in this study do not correspond entirely to results found in similar studies by Vakrman and Kristoufek (2015) and Da, Engelberg and Gao (2011). Conclusively, it was found that SVIs do not have predictive power on the movements of initial returns by IPOs and thus levels of underpricing. Contrary to expectations, these results show significant differences in using ASVs as compared to SVIs. One thing to notice is that ASVs are not common practice since financial research prefers using relative values as compared to absolute ones, as relative values have comparative power, whereas absolute values touch more upon the intrinsic values of firms in more detail. Relative values are also taken against a certain benchmark in a prespecified time horizon, which enables a comparative insight into the intrinsic meaning of the percentage changes in Google SVIs.

Next, an outline of the contents of each chapter in this dissertation is presented. Chapter 1 briefly introduces the topic of initial public offering underpricing. Chapter 2 includes the theoretical framework, considering relationships between the variables of interest and conclusions from past financial research. Consequently, chapter 3 describes the data set and variables used. Chapter 4 describes the methodology employed in this research. Finally, chapters 5 and 6 explore the results, leading to a discussion and a conclusion on the effect of Google search volumes (both relative and absolute) on underpricing levels.

CHAPTER 2 Theoretical Framework

2.1 IPO Underpricing

Underpricing is described by Allen and Faulhaber (1989) as an immediate loss to the initial owners of a firm offering its IPO. This is caused by a difference between the price of the IPO and the first closing price. Underpricing is a type of valuation for companies that decide on going public to raise funds for future operations but pricing their shares below market value. This happens through IPOs, where IPOs represent an exit strategy for initial investors to realize returns on investment (ROI). The unit of analysis is the firm level taking the US as a measure of context.

2.2 Absolute Search Volume and Relative Search Volume

Search volumes are defined as search queries for terms put into the Google Search Engine (Da, Engelberg & Gao). They use relative search volumes as a proxy for investor attention, whereas this research uses ASVs instead. ASVs can be considered part of internet traffic, where ASV can serve as a proxy for firm recognition and investor attention (Bank, Larch & Peter, 2011). Search volume indexes (SVIs) however are used as means of comparison to determine whether differences exist in the effects of both measures on IPO underpricing. SVIs are measured by dividing the total searches per geography and time frame to show relative popularity, after which the output is normalized into values ranging from 0–100 based on the proportion of the topic to all searches. I utilized worldwide searches in this research as investing has become increasingly popular and increasingly accessible for all kinds of investors. When mentioning ASVs or SVIs, both measures relate to Google's search traffic.

2.3 Relationship between Search Volumes, Investor Attention and IPO Underpricing

Vakrman and Kristoufek (2015) describe the effect of Abnormal Search Volume Indexes (ASVI) on IRs of IPOs. Where ASVIs are a logarithmic deviation of the 26-day median value of GSV. They find that IRs are higher for IPOs that receive increased investor attention. However, the results seem to hold for IPOs going public in positive media sentiment periods. This conclusion was drawn from the investor sentiment theory popularized by Aggarwal, Krigman and Womack (2002), Demers and Lewellen (2003) and Loughran and Ritter (2002). The research by Vakrman and Kristoufek (2015) considers both attention (idiosyncratic firm-specific) and market sentiment (market level) when considering the first-day returns. The researchers assessed the data on a fundamental basis first. The firms in the sample were divided into three groups based on their respective ASVIs pre-IPO, specifying: high, medium and low attention groups. The initial view yields a positive return for high-attention IPOs of 22.85%, while low-attention groups yield around 12.23% only. Thus, at first glance, it seems that investor attention is likely to significantly affect first-day IRs, according to Vakrman and Kristoufek (2015).

Additionally, the authors found that Google's SVIs can produce relatively accurate short-run forecasts. They show that the higher the attention before an IPO, the higher the IRs. The results presented are highly significant and have a magnitude of 41.4%, corresponding to an incremental standard deviation of ASVI. In contrast, investor sentiment and offering size do not have predictive power over IRs. The insignificant predictive power is also in contrast with Da, Engelberg and Gao (2011). Therefore, it is likely that the predictive effect of the size of the offering is highly dependent on the quality of search information as well as the selected sample of firms.

The paper by Ma and Tsai (2002) argues that underpricing and IRs cannot be used interchangeably. Price revision or correction occurs approximately half a year after going public. IRs are comprised of True Discount (TD) and Market Reaction (MR), where TD corresponds to underpricing. Ma and Tsai (2002) find that positive values of MR correspond to the overreaction of investors, whereas the opposite holds for negative values. The TD corresponds, as aforementioned, to the actual underpricing. Additionally, TD is not influenced by attention, whereas MR and attention are strongly interdependent. After careful and deliberate research on this setting introduced by Ma and Tsai (2002), Vakrman and Kristoufek (2015) confirm that Google SVIs can predict part of the IRs (market overreaction), while the true IPO discount (i.e. underpricing) cannot be predicted by utilizing SVIs.

The predictor or the independent variable in their respective research is similar to the one used in this paper, although it differs in nature using absolute and relative natures of the input data. Furthermore, the outcome is identical/very similar since the first day IRs of the IPO are accounted for in the dependent variable. However, the research by Vakrman and Kristoufek (2015) also considers long-run fluctuations caused by SVIs, which is not touched upon in this paper.

A related influential (seminal) piece by Da, Engelberg and Gao (2011) introduced the phenomenon of using Google SVIs as a proxy for investor attention and consequently as a potentially influential variable on IPO first-day returns and long-run underperformance of IPOs. They conclude that Google SVIs are a direct measure of investor attention as opposed to indirect measures, like turnover, extreme returns, news and advertising expense. Taking a sample of 3000 stocks ranging from 2004 – 2008, it is shown that SVIs are correlated but differ in nature from existing proxies used for investor attention. By testing the attention-induced price pressure hypothesis of Barber and Odean (2008), Da, Engelberg and Gao (2011) came to the following result. The result of this study is that an increase in SVIs for Russell 3000 stocks predicts larger first-day IRs for IPOs as well as a robust long-run underperformance for IPOs in the sample. By conducting this research, Da, Engelberg and Gao (2011) show the usefulness of search query data in financial applications. However, to the best of their knowledge, they were not the first to employ SVIs as a proxy for investor attention in financial literature. Mondria, Wu and Zhang (2010) were the first to employ internet search query volumes in financial research. They do not understate the importance of additional research on the effectiveness of internet search volumes in future research. Their closing note states that search volumes can be an objective way of viewing quantifications of investor interests and search volumes could and should see many more applications in financial research.

The final study discussed in this framework is the aforementioned study by Bajo and Raimondo (2017). They collected data from 2,814 IPOs in US stock markets between 1995 and 2013. The stark difference in their respective approach was that they utilized news articles as a proxy for investor attention to explain the workings of IPO underpricing. They conclude that news outlets can influence retail investors' belief in stock performance. In other words, positive newspaper tones are correlated with first-day IRs. They use textual analysis, introduced by Loughran and McDonald (2011), to examine the sentiment (i.e. tone) from close to 30,000 news articles from close to 500 individual newspaper outlets. The results gathered are statistically significant, as a shift of one standard deviation results in an approximate 2.5% increase in the level of underpricing. Furthermore, they discuss that the timing of coverage of the newspaper and its reputation are equally important. To be more exact, they discovered that positive sentiments ahead of the IPO are not associated with increased IRs, but positive sentiments close to the IPO date, that is 4 days before the release, are indeed associated with higher positive IRs. This also marks the period where the IPO coverage is at its peak. This finding is then interpreted as being a confirmation that media sentiment is important, but only if it is paired with investor attention. In addition, by bi-partitioning, Bajo and Raimondo (2017) found that only renowned (national) newspapers possess the ability to shape investor beliefs and therefore only these sources of significance are associated with levels of underpricing. This research is similar, but it takes a nuanced approach to the matter of proxies for investor attention. Sentiment is heavily relied upon, whereas in this paper the focus is on the sheer amount of volume towards IPOs, but also considering control variables to account for latent restrictions.

2.4 Corresponding Literature and Hypotheses to the Relationship between Investor Attention and Google Search Traffic

IPO underpricing has known a long history of trial and error, trying to explain the workings that result in the initial loss of IPOs. Muscarella and Vetsuypens (1989) conducted research on one of the models used to determine IPO underpricing, namely *Baron's model*. Baron's model relied on information misalignments between the issuers and underwriters, resulting in lower offer prices than would be if there were no such asymmetries of information. This was initiated to better comprehend what caused underpricing and what its repercussions meant for issuers and stakeholders. And even before Muscarella and Vetsuypens (1989) touched upon underpricing, IPOs were conceptualized and tested empirically (Reilly & Hatfield, 1969). To elaborate, Daily, Certo, Dalton and Roengptiya (2003) conducted a meta-analysis on the consensus on causes of IPO underpricing, after which they found that much of the attention catered towards IPO releases can be attributed to the dot com bubble and the rise of information dissemination technologies. This review article aggregates previous research into a comprehensive holistic view of the workings behind IPO underpricing, resulting in a nuanced view at the time the research was published. Although significant, it was not the first research published on underpricing mechanisms. One of the first researches on underpricing was conducted by Ritter (1991) stating that investors tend to have a periodic overoptimism towards novel IPO releases on which issuers can capitalize by leveraging deliberate timing of offerings. In addition, Loughran and Ritter (1995) mention that IPOs tend to overperform whenever the underlying market finds itself in a growth state. This research focused more on behavioral aspects. Ritter and Welch (2002) even go to the extent of stating that high first-day returns are a reward for institutional investors.

Although it has been clear that IPO workings have been long debated in existing financial literature, a consensus on the effect of information asymmetries has been present in the existing literature. The upper hand gained by investors with more information was modeled in the seminal work by Rock (1986). Rock (1986) argues that uninformed investors tend to overweigh their respective portfolios with low-value stocks as compared to well-informed investors, caused by information asymmetry as a primary cause. Additionally, Bank, Larch and Peter (2011) agree and found that an increase in search queries is correlated to an increase in trading activity of a stock and its liquidity, however for mostly uninformed investors. Adding to this view, Aggarwal, Prabhala and Puri (2002) state that institutional investors outperform retail investors, extending the existing literature by providing a more profound analysis of the types of investors.

The causes of IPO underpricing are related, but not uniquely so, to information asymmetries. Therefore, causes like greater uncertainty about the future proceedings of IPOs tend to increase the first-day underpricing (Ritter, 1984). In addition, signaling practices have become increasingly important in conveying IPO value by issuers. For instance, venture capitalist backing and the hiring of reputable underwriters have been used to signal the value and quality of prospective IPOs (Booth & Smith, 1986; Lee & Wahal, 2004). Greater uncertainty is one of the many reasons for IPO underpricing and these factors are disclosed in the analysis part of this paper to ensure even coverage of factors and to isolate the effect of ASVs on IPO underpricing for the given timeframe.

As investor attention was attributed to the dot com bubble by Daily, Certo, Dalton and Roengptiya (2003), the internet has evolved into a more advanced and sophisticated era now. All individual search queries in Google's search engine are being saved and its statistics are available to the public. Therefore, this databank is timely, reliable and freely available. Jun, Sun Yoo and Choi (2018) go above and beyond by providing a comprehensive overview of 657 research papers utilizing Google Trends as a big data source. Search query volumes have seen a shift from being used to recognizing trend behavior to being measures for forecasting purposes.

Its use has been extended to more areas of research, such as product sales (Choi & Varian, 2009), real estate sales projections (Dietzel, Braun & Schäfers, 2014), surveillance of disease outbreaks (Carneiro & Mylonakis, 2009), epidemiology (Seifter, Schwarzwalder, Geis & Aucott, 2010) and many more. In terms of financial markets, a more recent study by Chen (2017) examined Google search volumes as a proxy for investor attention and found that increased investor attention causes a decline in stock returns, especially those of indexes. The seminal study by Da, Engelberg and Gao (2011) goes beyond that notion by stating that investor attention is better captured by using SVIs because it tends to capture investor attention in a timelier manner whilst also accurately measuring retail investors' attention. SVIs are even argued to predict higher stock prices in the first 2 weeks and even yield higher initial first-day returns and consequently: long-run underperformance. Follow-up research by Goddard, Kita and Wang (2015) expanded upon the rationale offered by Da, Engelberg and Gao (2011). They investigated the link between investor attention span and the dynamics of currency prices using *limited attention theory*. Essentially, they added to the growing library of scientific research dedicated to the effects of online search intensity through the Google search engine. They conclude that empirical evidence is limited and findings require the development of more robust models on the role of investor attention. Perhaps using ASVs will be a revelation.

To add to the growing literature on search intensity Swamy, Dharani and Takeda (2019) provided a profoundly extending piece of knowledge. This study elaborated on the investor attention effect caused by Google search intensity for emerging economy equity markets. This study performed quantile regression analysis to increase understanding of why online search intensity generates insightful information in predicting stock returns using SVIs. The study finds that high search intensity is highly correlated with predictive power on significant returns in the 3 subsequent weeks of trading of stocks. Using a quantitative approach, SVIs are proven to be insightful in predicting the direction and magnitude of excess returns. In conclusion, the paper states that SVIs can be a significant tool in determining profitable trading strategies. A different take on the effects of the Google search engine on financial market outcomes was introduced by Volyublennaia (2014). The author introduced Google search probability (GSP) as a proxy for investor attention and found that increased investor attention was only short-lived in most cases. The research concludes by stating that the impact of attention is short-lived in nature and that, conversely, changes in returns are significant and long-lasting in nature. The study emphasizes past index returns as an indicator for future returns as this data reveals the nature of information procured by investors. Additionally, Volyyublennaia (2014) argues that investor attention diminishes the predictability of index returns, whereas lagged values of indexes are weaker in terms of impact if more attention is present. Overall, more revealed information caused by increased investor attention leads to more efficient markets according to the paper.

On the other side, using SVIs as a predictive source of big data has seen many difficulties recently. Challenges of sorts range from overgeneralization to low-quality data, where the utilization of big data is not a prerequisite to accurate forecasting and predictions. To use SVIs successfully, one ought to refine through filtering big data. Jun, Sun Yoo and Choi (2018) provided a nuanced view of the possibilities, characteristics and shortcomings of using Google SVIs as a source of big data, where the conclusion is that the possibilities outperform the limitations if data collection is refined and performed critically.

The consensus that can be drawn from the existing literature on both Google SVIs and IPO underpricing, is that an increased rate of search intensity is an accurate measure of investor attention, however, the effect is short-lasting in nature and does therefore not affect long-term stock prices. Moreover, underpricing is defined as an inefficiency of financial markets in most if not all financial literature. Moreover, underpricing has many causes for its existence, which makes it increasingly difficult to locate its true drivers.

2.5 Expectations and Hypotheses

This research is conducted keeping the original study by Vakrman and Kristoufek (2015) in mind. Therefore, I expect that an increase in ASVs pre-IPO is partly causing an increase in the initial underpricing of IPOs in the US between 2020–2021. Additionally, the results will most likely not reign true in the long run, as most research on this topic concludes that investor attention gains traction in the short run and exhibits no similar result in the long run. Differences may reside in the different time horizons, considering the (potential) effect of COVID-19 on the results as well as a different predictor (ASVs). Thus, the following hypotheses are tested as such:

H1: *Google's absolute search volumes will positively and significantly affect the initial returns of IPOs in the US from 2020–2021 as a proxy for investor attention.*

H2: *Google's search volume indexes will positively and significantly affect the initial returns of IPOs in the US from 2020–2021 as a proxy for investor attention.*

CHAPTER 3 Data

3.1 Data set

The data in this research was mainly collected from Stock Analysis and Yahoo Finance. The total number of IPOs in 2020 and 2021 was 1,515. However, after filtering the data only emerging growth firms are accounted for in this research. The reason is that emerging growth firms are unbiased in effect as spin-offs of established companies have some type of recognition before going public, hence potentially offsetting the effect of search volumes on the IRs. Additionally, Vakrman and Kristoufek (2015) also used this approach. Moreover, special purpose acquisition companies (SPAC) and holding firms are not considered as these are not emerging growth firms. Therefore, the final sample contains 427 IPOs. *Keyword Everywhere* is used to uncover the ASVs of search terms by using Google Trends. This medium has proven to be reputable, as this extension has been used by over a million paid subscribers for many applications. The entries of ASVs into the data were added manually. Only firms from the NASDAQ and NYSE were taken as these are the most prominent exchanges in the US. This is measured within a dummy variable. This variable takes on a value of 1 if the IPO was offered on the NYSE and 0 if it was offered on the NASDAQ. The respective industries of the firms in the sample belong to range from biotechnology to waste management and were founded between 1851 and 2021. Moreover, the number of employees ranges from 3 to 69,252.

Stocks and corresponding tickers where data were not available on Yahoo Finance for the first day of public trading were not considered in the research and the data were altered accordingly. Additionally, some stocks moved to foreign exchanges, which were then removed from the sample for this research. Furthermore, over-the-counter market exchanges are also not considered in the research since a lack of regulation and normality exists within these respective markets.

In this study, four Application Programming Interfaces (API) were created manually in Python using Pandas and Matplotlib to retrieve data from Yahoo Finance and Google Trends on first-day closing prices. The second API is utilized to retrieve the SVIs of tickers in the sample using Google Trends as a primary source. The API includes a built-in break function to ensure its progress and not be detected as bot activity. The code can be found in my Github Repository (Zhang, 2023). Two benchmarks are used for the SVIs and ASVs: the 28-day and the 7-day mark. Therefore, the values on both the 28^{th} and the 7^{th} day are taken as inputs for the regression as these values represent the popularity of a query over the specified time horizon in the data collection method.

3.2 Variables and Control Variables

Multiple variables are used for both regressions, hereafter the variables are introduced and explained. *SVIs* are used to measure the search traffic on each respective name/ticker of each IPO in 2020–2021. The same applies to *ASVs*.

Initial Return (IR), which is defined as the log return of the IPO calculated from the logarithm of the offering price to the first-day closing price. The log of *offering size, which indicates the size of the initial offering, measured in US Dollars. Offering size x employees is* defined as an interaction term between the value of the offering and the total number of employees. This applies to the SVI regression as well. This interaction term was believed to be effective as venture capitalists are well known to prefer well-rounded teams in terms of human resources when choosing which businesses to invest in. The quality of the team, as well as the size of the offering, are believed to have a bilateral effect on IPO underpricing through its venture capitalist backing (Johnson & Welbourne, 2000). *NYSE is* defined as a dummy variable that equals 1 if the IPO belonged to the NYSE and 0 otherwise. The respective industry in which an IPO operates is used to account for correlated residuals. *Employees* is used as the number of employees a firm has at its IPO. Lastly, the variables of interest are denoted within a 28-day and 7-day interval prior to the IPO for both ASVs and SVIs, denoted as *ASV 28, ASV 7, SVI 28 and SVI 7.* All the variables were collected from Stock Analysis and Yahoo Finance and the closing prices and SVIs were collected by using APIs. Furthermore, ASVs were collected manually as mentioned before. The number of observations is equal across variables with N = 427, except for employee count and the interaction term.

3.3 Summary Statistics

Table 3.1 Descriptive statistics of the variables used for the OLS-regression
This table includes observations, mean, standard deviation, minimum value and maximum value of the variables used in the OLS-regression. This shows an overview of the dimensions of all variables.

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
IR	427	0.26	0.57	-0.91	4.41
ASV 28	427	57,059.84	378,171.13	0	5,450,000
ASV 7	427	47,563.35	320,642.68	0	4,550,000
SVI 28	427	38.16	36.69	0	100
SVI 7	427	44.37	39.98	0	100
Size	427	18.84	1.39	14.83	24.08
NYSE	427	0.21	0.41	0	1
Size x Employees	411	2,910,000,000,000	18,400,000,000,000	21,800,000	287,000,000,000,000
Employees	411	2,110.89	6,501.47	3	69,252
TV C	427	91.50	49.77	1	173
TV NYSE	427	89.86	48.61	1	172
Industry	427	33.51	22.12	1	73

Note. See Appendix A for a full description of the variables. Trading volumes are denoted in millions.

The *Initial Returns (IR)* display a positive mean with a skewness to the right of the distribution. The *ASV* for both the 28-day and 7-day intervals shows high standard deviations from the positive means. This means high volatility among IPOs. The size of the offering is positive (which is expected) but has a wide range of values. This can deem outlier management to be necessary for the remainder of this research. The employee count seems to adhere to expectations however, it must be mentioned that the trading volume variables are measured in millions. Since all the trading volume values are refined to have the same format, the effect should remain unchanged. Lastly and noticeably, the number of industries of IPOs in this sample is 73, which is used to account for within-group correlation.

For additional information on the frequencies of country and industry categories, please refer to appendix C.

Table 3.2 Pairwise correlations
This table shows the correlations between all variables used in the OLS-regression. Correlations may take on values between 0 and 1, where 0 indicates no particular correlation and 1 articulates a perfect correlation.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) IR	1.000								
(2) ASV 28	0.047	1.000							
	(0.330)								
(3) ASV 7	0.049	1.000*	1.000						
	(0.314)	(0.000)							
(4) SVI 28	0.032	-0.031	-0.033	1.000					
	(0.511)	(0.528)	(0.498)						
(5) SVI 7	0.020	-0.042	-0.045	0.664*	1.000				
	(0.685)	(0.382)	(0.356)	(0.000)					
(6) Size	-0.013	0.346*	0.346*	0.134*	0.087	1.000			
	(0.792)	(0.000)	(0.000)	(0.005)	(0.073)				
(7) Size x Employees	0.009	0.552*	0.548*	-0.040	-0.056	0.345*	1.000		
	(0.848)	(0.000)	(0.000)	(0.416)	(0.261)	(0.000)			
(8) NYSE	-0.068	0.042	0.044	0.114*	0.125*	0.332*	0.132*	1.000	
	(0.160)	(0.386)	(0.364)	(0.018)	(0.010)	(0.000)	(0.007)		
(9) Employees	-0.023	0.160*	0.161*	0.033	0.014	0.319*	0.609*	0.223*	1.000
	(0.644)	(0.001)	(0.001)	(0.506)	(0.775)	(0.000)	(0.000)	(0.000)	

Note. *** p<0.01, ** p<0.05, * p<0.1. See Appendix A for a full description of the variables.

Table 2.2 shows a significant correlation between ASV 28 and ASV 7, which was expected since ASV 28 contains ASV 7 values by default. The same can be said about the SVI 28 and SVI 7 variables since both variables measure a similar phenomenon. It is however interesting that the offering size as well as the interaction term of offering size and employee count are correlated to the ASVs, albeit only significant at the 10% level. Furthermore, the NYSE dummy variable seems to be correlated with SVI levels but not with ASV. Lastly, the offering size and the interaction term of offering size and employee count show a correlation that is according to expectations.

CHAPTER 4 Methods

The ASV data was found by searching the following structure(s) for each IPO:

Stock name + 'stock' or *Ticker + 'stock'*

Furthermore, a logarithmic transformation of the IR variable is used, as popularized by Vakrman and Kristoufek (2015). Therefore, the IRs look as follows:

$$IR = log(Pclose) – log(Poffer)$$

Where Pclose denotes the closing price and Poffer denotes the offering price.

The variables of interest are studied in two separate regression equations for both SVIs and ASVs. The regressions take the following form:

$$IR_1 = \beta_0 + \beta_1 * ASV7 + \beta_2 * ASV28 + \beta_3 * \boldsymbol{Control\ Variables}_i + \epsilon_i$$

$$IR_2 = \beta_0 + \beta_1 * SVI7 + \beta_2 * SVI28 + \beta_3 * \boldsymbol{Control\ Variables}_i + \epsilon_i$$

Control Variables contain all the variables in the equation to ensure no omitted variable bias (OVB). The variables include *employee count*; *NYSE*; *interaction term size x employee count* and the *offering size*. OLS is used to test the significance of the variables of interest as well as the statistical significance of the remaining control variables. 5 conditions must be satisfied to ensure the validity of the statistical method employed (which is OLS in this case). A synopsis is provided on all 5 conditions below:

A1 E($\boldsymbol{u_i}$) = 0

The mean independence assumption is satisfied whenever the expected value of the error term equals zero. This must be true to ensure that the error term is not related to the independent variables in the sample. However, there is no statistical test for this assumption, so the possibility remains that the constant in this model is biased. However, I assume the opposite to be true.

A2 Var($\boldsymbol{u_i}$) = $\boldsymbol{\sigma^2}$ < ∞

We want to check if our model contains homoscedasticity meaning that the error terms have the same variance. If our model suffers from heteroskedasticity then the standard errors are wrong and we would need new standard errors for hypothesis testing. I utilize two formal tests to test for potential heteroskedasticity, namely the Breusch-Pagan test and the White test. The tests show the rejection of homoskedasticity in favor of heteroskedastic standard errors. The results can be found in Appendix B. This is solved by using robust standard errors in the regression. Not using robust standard errors results in having biased estimates of the coefficients in the regression. The tests convey p-values lower than the critical value of 0.05, which indicates the existence of heteroskedasticity. Therefore, I use robust standard errors to account for this assumption.

A3 Cov(u_i, u_j) = 0

Working with cross-sectional data renders it near impossible to detect correlation among the model's residuals. However, a solution exists in using sector/industry clustering of the IPOs in the sample to account for potentially correlated errors. Although within-group error clustering is allowed by using this method, across-group clustering is not supported (Brooks, 2019). Therefore, using clustered errors, I am enabled to assume uncorrelated errors.

A4 Cov(xi, ui) = 0

By using an IV regression, one should be able to control for omitted variable bias (OVB). The first stage is committed to explaining the relationship between the endogenous variable and the instrumental variable(s). The second stage is then used to estimate the relationship between the predicted values and the dependent variable. The NYSE is the largest stock exchange in the world by size. Although market sentiment is traditionally believed to encourage herding effects by investors (Blasco, Corredor & Ferreruela, 2011), I was only able to collect monthly data on consumer sentiment, which does not represent investor attention sufficiently. Thus, the trading volume of the NYSE is used alongside the trading volume based on the trading volume of stocks that are both traded on the NYSE and the NASDAQ. This data was retrieved using an API from NYSE's official website. The data can be found in my GitHub Repository once again (Zhang, 2023). Trading volume is believed to be a relevant instrument since trading volume is traditionally believed to be a sufficient indicator of investor engagement and the popularity of retail investing. Aouadi, Arouri and Teulon (2013) found that investor attention, proxied by search volume, is strongly correlated with trading volume. Therefore, we can ascertain that trading volume has a potential effect on IPO underpricing by its correlation with ASVs and SVIs, without having a direct effect on the levels of underpricing as trading volume does not affect the fluctuations of IPOs before release.

From the perspective of institutional investors, they are driving factors of IPO-values of firms as they signal value to retail investors. Chiang, Chan and Sherman (2010) find that institutional investors are adequate bidders in terms of the ultimate price of an IPO, whereas retail investors seem to be less well-informed and less precise in bidding close to the IPO's intrinsic value. Trading volume is correlated with online search volume as an increased trading volume indicates a higher attention rate for financial investing in general, which in turn potentially influences online searches for stocks. However, the trading volume of existing stocks is not proven to have a significant effect on specific IPOs' performances after initialization. The volume might indicate a higher liquidity and volatile market, however, it need not affect the fluctuations of IPO pricing levels. Although the trading volume of IPOs after release is associated with higher levels of underpricing, especially when a lock-up period is applicable, this does not prove the relationship between the levels of underpricing and trading volume of major stock markets before release (Zheng, Ogden & Jen, 2005). Furthermore, there are more factors at play when it comes to the fluctuations in trading volume of the NYSE and the NASDAQ. Economic factors such as overall market health, investor attention and economic climate may influence the volatility as well. Furthermore, restricted trading is also a significant determinant of trading volume, as some IPOs are not publicly available to all types of investors but only to a select few. This comes down to the preferences of the issuer and the underwriter employed. In addition, Nimalendran, Ritter and Zhang (2007) found that IPO post-allocations were mainly affected by commission per share to underwriters and not so much by trading volume. Therefore, trading volume will prove to be a relevant instrument in the regression. Lastly, exogeneity cannot be tested and is therefore assumed for now.

After realizing both conditions, a two-stage IV regression was run. The following models were specified:

$$IR_1 = \beta_0 + \beta_1 * ASV7 + \beta_2 * ASV28 + \beta_3 * \boldsymbol{Control\ Variables}_i + \epsilon_i$$

$$IR_2 = \beta_0 + \beta_1 * SVI7 + \beta_2 * SVI28 + \beta_3 * \boldsymbol{Control\ Variables}_i + \epsilon_i$$

$$Instrumental\ variables = NYSE\ trading\ volume,\ NYSE\ Tape\ C$$

Where NYSE Tape C comprises stocks traded on both the NASDAQ and NYSE.

A5 $u_i \sim N(0, \sigma^2)$

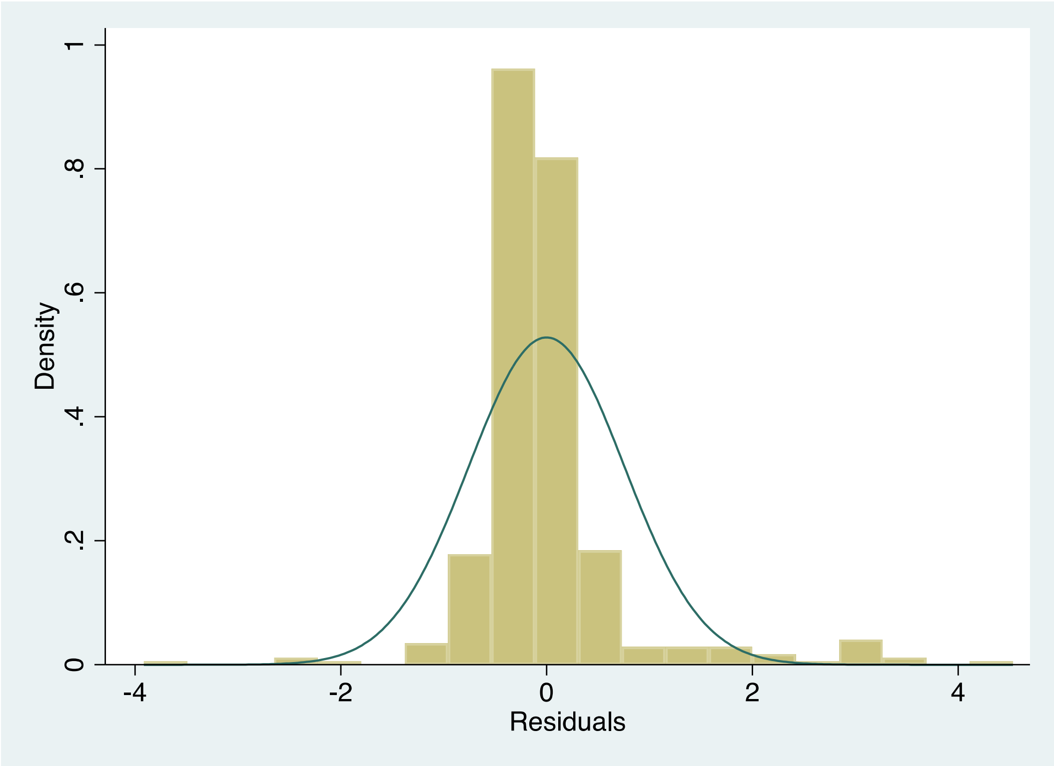

Figure 4.1 Distribution of the residuals of the effect of absolute search volumes on IPO underpricing

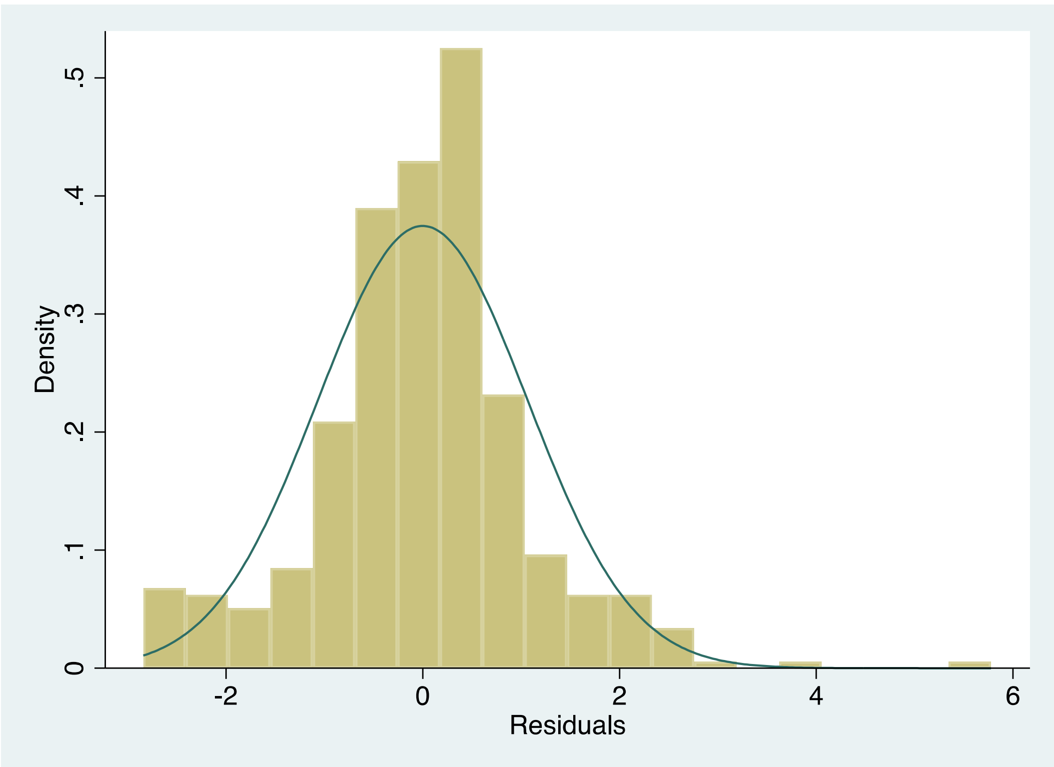

Figure 4.2 Distribution of the residuals of the effect of relative search volumes on IPO underpricing

The residuals seem to be approximately normally distributed, although it must be stated that the absolute search volume regression residuals in Figure 4.1 display high center values, whereas the tails are relatively flat. The normal distribution is still the best description that fits the overall trend of the residuals, although arguments do exist in favor of a student t-distribution. The residuals of the relative search volumes are close to normal, however, a slight right skew can be detected by analyzing the histogram. Some outliers are present and visible, however, they pose no significant threat to the normality assumption of the error terms. Hence, normality of residuals is assumed and the research is carried out accordingly.

All the statistical analysis is conducted in STATA, which is a powerful statistical analysis tool. Occasionally, Python (a statistical programming language) is of assistance in retrieving multidimensional data for research, alongside packages included in the software, such as Pandas and Matplotlib. All the scripts are be linked to my GitHub repository (Zhang, 2023).

CHAPTER 5 Results & Discussion

The model is estimated using an IV regression. Therefore, since the dependent variable IRs is measured in logarithms and the independent variables ASV and SVI are measured in absolute and relative terms, the coefficient of X can be interpreted as follows: when X changes incrementally and marginally (i.e., by 1 unit), the percentage change of Y is the coefficient of X.

Table 5.1 IV-regression results from regressing initial returns on Google absolute search volume

This table shows the effect of absolute search volumes on initial returns while controlling for several other variables. Each regression incrementally increases the variables used, ultimately arriving at the final regression equation on the far right.

			Dependent Variable		
			Initial Returns		
	(1)	(2)	(3)	(4)	(5)
ASV 7	0.00007 (0.00004)	0.00007* (0.00004)	0.00007* (0.00004)	0.00007* (0.00004)	0.00006** (0.00003)
ASV 28	-0.00006 (0.00003)	-0.00006* (0.00003)	-0.00006* (0.00003)	-0.00006* (0.00003)	-0.00005** (0.00003)
Size		-0.01588 (0.04014)	0.00662 (0.04071)	0.005504 (0.04322)	-0.00145 (0.04149)
NYSE			-0.20996** (0.09837)	-0.19838* (0.10525	-0.18047* (0.09497)
Employees				0.00000 (0.00000)	-0.00002* (0.00001)
Size * Employees					0.00000 (0.00000)
Constant	0.30182*** (0.03457)	0. 60137 (0.76029)	0.22679 (0.76927)	0.26451 (0.81274)	0.33922 (0.05310)
Observations	427	427	427	411	411
First-stage adjusted R_2	0.9992	0.9992	0.9992	0.9992	0.9993

Note. *p<0.1, **p<0.05, ***p<0.01. Standard errors are reported in parentheses. The second-stage R_2 is omitted due to its insignificance in IV-regression application. Offering size is measured in US Dollars. Full description of the variables can be found in Appendix A.

In the IV-regression, the aim is to minimize OVB and to account for endogeneity. The R^2 in the IV is often referred to as 'the weak R^2' or 'the partial R^2', which does not provide any insightful information on the goodness of fit of the model. The R^2 does not directly represent the variance explained by the instruments or the independent variables in the model. However, the adjusted R^2 is reported for comprehensiveness purposes. Other tests exist to test for the significance of both the regression and its instruments and these are utilized in the remainder of this paper. For instance, an endogeneity post-IV-regression is used to prove the existence and the necessity for using instruments to account for OVB. The results show that endogeneity is present indeed, which means that the instruments prove to be justified in trying to account for any missing variables in the error term that might be correlated with the variables of interest. To see the result of the post-regression test, please refer to Appendix B.

Seeing the results of the ASV-regression in column 5 of table 5.1, it shows a significant effect on the 5% level, which indicates that the effect of ASVs is statistically significant on IRs of IPOs from 2020–2021. One can tell that the effect of absolute search volume 7 days before the release of an IPO is .00006 for every incremental increase in the ASV. Therefore, an increase of 10,000 in search volume would indicate a 60% increase in the IPO underpricing. Conversely, the ASV 28 days before release have the opposite effect, which is surprising because a similar effect to ASV 7 days prior was expected. This is opposite to what Da, Engelberg and Gao (2011) found as they only found that investor attention close to release was expected to have a significant effect. In the end, the coefficient indicates that a 10,000-search incremental increase causes an approximate 50% decrease in the level of underpricing for IPOs. Furthermore, being traded on the NYSE only results in an approximate 18.05% decrease in IPO underpricing, albeit at a 10% significance level. In addition, the control variables offering size and the interaction term of employee count and offerings size do not seem to be having a significant effect on IPO underpricing, both in magnitude and statistical significance. This is in contrast with the findings of Da, Engelberg and Gao (2011). Furthermore, the significance of the employee count variable is lacking, which makes the size of the coefficient negligible for interpretation due to the 5% confidence level taken in this study. Finally, the constant/intercept cannot be interpreted as some variables cannot take on the value of 0, leaving the intercept without a logical interpretation. In addition, the constant is not statistically significant. Although there seems to be significance in the results, due to the potential existence of OVB, the exclusion restriction cannot be discarded (Brooks, 2019). Further research on this topic would be necessary to find more causes of underpricing. Although this IV regression has its limitations, by using these fitted values, the aim is to be closer to the true estimates underlying the phenomenon of underpricing in US IPOs. Lastly, it must be mentioned that values that equal 0.00000 in Table 5.1 do not represent a zero-coefficient of the variable, instead, the formatting is done was a way that would best capture the effect of each variable since the effects of each respective variable were relatively small.

Therefore, Hypothesis 1 is hereby not rejected based on the 5% significance level taken in this paper. This is in line with the findings by Vakrman and Kristoufek (2015), as they found a significant effect of above-average attention proxied by Google search volumes on initial returns, however by using relative

Google search volumes. They also conclude that above-average attention is only influential in periods of positive market sentiment. This is not touched upon in this research. Ultimately, the first hypothesis cannot be rejected and therefore, I conclude that absolute search volumes have predictive power on IPO underpricing in this sample based on US IPOs ranging from 2020 to 2021. Next, I test whether the findings by Vakrman and Kristoufek (2015) hold by using the same IV-regression approach while replacing ASVs with SVIs. This replicates the variables of interest used by Vakrman and Kristoufek (2015) and the results are used to create a verdict on whether a difference exists between using ASVs as input and SVIs as input in financial research on IPO underpricing as well as a final verdict on the effect of Google search volumes on overall IPO first-day performance.

Table 5.2 IV-regression results from regressing initial returns on Google relative search volume

This table shows the effect of relative search volumes on initial returns while controlling for several other variables. Each regression incrementally increases the variables used, ultimately arriving at the final regression equation on the far right.

			Dependent Variable		
			Initial Returns		
	(1)	(2)	(3)	(4)	(5)
SVI 7	0.05537 (0.07365)	0.05589 (0.07250)	0.05669 (0.07448)	0.03306 (0.03411)	0.03044 (0.03161)
SVI 28	-0.03956 (0.05420)	-0.03992 (0.05330)	-0.04017 (0.05461)	-0.02325 (0.02589)	0.02120 (0.02397)
Size		-0.00357 (0.05712)	0.03596 (0.08073)	0.01794 (0.05201)	0.00708 (0.04730)
NYSE			-0.42138 (0.49076)	-0.24930 (0.20961)	-0.22871 (0.19029)
Employees				0.00000 (0.00000)	-0.00000 (0.00000)
Size * Employees					0.00000* (0.00000)
Constant	-0.69161 (1.20416)	-0.63369 (1.72377)	-1.3174 (2.42164)	-0.62301 (1.34066)	-0.38087 (1.19099)
Observations	427	427	427	411	411
First-stage adjusted R_2	0.4377	0.4364	0.4379	0.4422	0.4417

Note. *p<0.1, **p<0.05, ***p<0.01. Standard errors are reported in parentheses. The second-stage R_2 is omitted due to its insignificance in IV-regression application. Offering size is measured in US Dollars. A full description of the variables can be found in Appendix A.

At first sight, it becomes obvious that, contrary to expectations, SVIs do not seem to have the same predictive power on the level of IPO underpricing as ASVs in this research. This is in stark contrast with the findings of Da, Engelberg and Gao (2011) and Vakrman and Kristoufek (2015). Namely, both papers conclude with remarks on the increased pricing levels of IPOs post-release considering a heightened level of attention accumulated by proxying Google search volumes (relative values). Therefore, it is expected to find similar results to both the highly influential paper by Da, Engelberg and Gao (2011) and Vakrman and Kristoufek (2015), respectfully. Looking at the data in column 5 of Table 5.2, SVIs do not seem to display the same characteristics as the ASV coefficients. Both the SVI 7 days prior and the SVI 28 days before publication exhibit insignificance at the 5% level. This means that SVIs do not have any predictive power over the IPO underpricing in this sample. Most variables seem to display statistical insignificance, which does not allow for the interpretation of any of the findings as potential correlations with either a positive or negative movement or co-movement. The interaction term between offering size and employee count is however significant, albeit at the 10% level. Next, offering size and employee count do not account for much explanatory power either as both are also statistically insignificant. The same interpretation of the intercept as mentioned for the ASV regression applies here as well. Additionally, the 0.00000 values need not necessarily be equal to zero, however, the coefficients are small enough to disregard their respective effects in the post-regression analysis. This applies to the interaction effect between offering size and employee count as well. Although significant at the 10% level, the size of the coefficients is negligible.

Hypothesis 2 can therefore be rejected as there is no sufficient evidence to support the original claim made by Vakrman and Kristoufek (2015). There seems to be no direct correlation between relative search volumes and IPO underpricing for this sample.

Thus, relative Google search volumes do not seem to have predictive power on the sample used in this research. This may reside in the set of variables used in their research and in this paper. They included long-term effects and sentiment analysis in their respective research. This was not feasible in this research as resources were limited. Moreover, Vakrman and Kristoufek (2015) had access to complex data on sentiment analysis, which was transformed to fit their regression accordingly and accurately. SVIs might also be subject to endogeneity, even after conducting an IV regression, since there are many more variables that cause or are correlated with levels of IPO underpricing.

As Mondria, Wu and Zhang (2010) were among the first to use Google search volumes in financial research, both Da, Engelberg and Gao (2011) and Bajmondo and Raimondo (2017) found that investor attention, proxied by Google SVIs, displayed predictive characteristics over IPO underpricing. While one paper assessed the effect of SVIs directly on IPO underpricing, the other used text-based sentiments to predict the fluctuations of respective IPOs. The focus of this research is the use of ASVs application in financial literature and research. The results show that ASVs have predictive power when it comes to IPO underpricing, however, the SVIs do not seem to have the same ability considering the sample used in this paper. Although ASVs are not commonly used in research, it shows potential for future application in financial literature as it proves to be insightful and significant in a nuanced way compared to the traditionally used relative values of search volume.

The results in this paper show that increased attention before an IPO, proxied by Google ASVs, drives up prices after the release of IPOs in this sample. In other words, more attention indicates higher levels of IPO underpricing. This proven effect might be limited to the US or to the sample used in this research, which means that in different contexts or countries, the results might differ. However, considering the size of the US IPO market, one can believe that the effect of search volumes can indeed be extended to other fields of financial research. This study shows that an increase in IPO underpricing is likely positively influenced by higher levels of ASVs 7 days before release, whereas higher levels of ASVs 28 days before release cause a significant negative effect on IPO underpricing. This effect might be dependent on the use of ASVs, or the sample employed in this research. Therefore, it is suggested to test the use of ASVs in IPO underpricing research more extensively, using different methods and data in future applications.

CHAPTER 6 Conclusion

6.1 Concluding Remarks

In this study, I have researched the effect of absolute search volumes on IPO underpricing, specifically for a sample gathered from IPOs residing in the US from 2020–2021. These years were marked as ‘COVID-19’ years, which also saw record-breaking numbers of IPOs consecutively. Based on previous research, it was believed that heightened investor attention causes higher levels of IPO underpricing, especially when these increased numbers of attention were close to the release of a particular IPO. This notion was popularized by seminal and smaller-scale studies alike, but they shared the commonality of having seen no disruption in, not only the financial markets but in entire economies and societies around the world. Also, because the US is home to the largest IPO market in the world, as well as hosting the largest stock exchanges in the world, I found it intriguing to see whether COVID-19 and record-breaking numbers would result in different conclusions based on existing research methods. Additionally, I looked at Google's absolute search volumes instead of relative values, which has been the benchmark for similar studies on this topic before. Moreover, the effect of ASVs on IPO underpricing has never been studied before. By comparing both ASVs and SVIs, I would be able to draw a conclusion on each respective effect and therefore conclude on the research question in this paper, namely: *How do absolute search volumes affect IPO underpricing and is it different from using relative search volumes?*

To answer this question, an OLS regression was conducted, more specifically an IV regression since omitted variable bias was believed to exist within the model. The results showed statistically significant coefficients for both absolute search volumes 28 days before release and 7 days before release. The same does not apply to the relative search volumes regression as none (except for the interaction effect) of the variables were shown to be significant, despite having used equal data collection, data filtering, data cleaning and data analysis methods (including the same regression equation). Therefore, I conclude that absolute search volumes are believed to have influenced IPO underpricing in US stocks that went public on either the NASDAQ or NYSE (or both) between 2020–2021. A caveat that needs mentioning is that the effects might be partially related to the COVID-19 upset. The effects are however opposite as the search volumes 28 days before release are negatively correlated with IPO underpricing and whereas the search volumes 7 days before initialization are positively correlated with IPO underpricing. Thus, this research concludes by stating an existing effect of ASVs on IPO underpricing, a notion that has not been shown before, which might enable further use of ASVs, not only in financial literature but also extending the use onto more topics within the discipline as well as onto other disciplines. Although the results in this paper suggest a significant influence of ASVs on IPO underpricing, the effects might be dependent on the sample quality; the sample size; statistical methodology; data collection and many more confounding factors that lay at the foundation of the results shown in this paper. As there are more factors influencing the level of pricing for IPOs in the US, more extensive research must be done to ensure that the effect is indeed robust and significant for other samples as well.

6.2 Implications

To conclude, many factors are influential in the search for variables influencing IPO underpricing. This concerns both institutional investors as well as retail investors. Institutional investors are believed to have asymmetric information benefits as compared to retail investors. This is in favor of institutional investors as they have more timely and accurate information. The applications of the results in this research paper suggest that heightened levels of anticipation for IPOs result in 'money left on the table' opportunities for all investors, which can be leveraged by scrutinizing the levels of anticipation or investor attention using Google's absolute search volumes. As this study suggests, Google search volumes are significant predictors for underpricing levels of IPOs in the US. Whether this holds for other samples remains to be seen. For now, these insights can be used by retail and institutional investors alike to predict post-IPO movements. For instance, increased search volume 7 days before an IPO release indicates a probability of an underpricing occurrence, while increased search volume attention 28 days before release causes the opposite effect. These results can then be used by investors seeking to invest in IPOs.

6.3 Limitations

This research has been conducted to the best of my abilities. Data were retrieved and consolidated in a way that would be robust to biases. The first limitation lies in the foundation of the study: absolute search volumes are not commonly used in financial market research, as absolute values are liable to biases, display lower comparability (compared to relative values) and are harder to generalize and apply on samples other than the one used. A second limitation lies within the data collection method. First, Google Trends Supercharged (or Glimpse) did not suffice as a reliable source since the official Google extension provided no means to examine daily data on absolute search volumes. Data was available however only when looking at a 5-year window. This could be improved in the future as the extension is relatively short-lived and could see updates improving its usability for scientific research. In addition to its technical limitations, Google Glimpse Supercharged is not free to the public and requires a monthly subscription costing over 200 Euros, which would not be reimbursed by Erasmus University Rotterdam. Therefore, when the software improves, researchers could potentially utilize this extension over the Keywords Everywhere extension, which is estimated based on Google Trends input. Although proven to be accurate, it is not equal to an internally created extension such as Google's own Google Glimpse Supercharged. The last limitation is influenced by the consolidation procedure. Scoops was used to consolidate the information gathered from Stock Analysis and Yahoo Finance; however, Scoops has not been updated to include IPOs after 2020. A solution to this would be to use other databases for both data collection and data consolidation to ensure that the proven effects in this study, consequently, hold in other samples as well.

REFERENCES

Aggarwal, R. K., Krigman, L., & Womack, K. L. (2002). Strategic IPO underpricing, information momentum, and lockup expiration selling. Journal of financial economics, 66(1), 105-137.

Aggarwal, R., Prabhala, N. R., & Puri, M. (2002). Institutional allocation in initial public offerings: Empirical evidence. The Journal of Finance, 57(3), 1421-1442.

Alexander Dietzel, M., Braun, N., & Schäfers, W. (2014). Sentiment-based commercial real estate forecasting with Google search volume data. Journal of Property Investment & Finance, 32(6), 540-569.

Baig, A. S., & Chen, M. (2022). Did the COVID-19 pandemic (really) positively impact the IPO Market? An Analysis of information uncertainty. Finance Research Letters, 46, 102372.

Bajo, E., & Raimondo, C. (2017). Media sentiment and IPO underpricing. Journal of Corporate Finance, 46, 139-153.

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. Financial markets and portfolio management, 25, 239-264.

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

Blasco, N., Corredor, P., & Ferreruela, S. (2012). Market sentiment: a key factor of investors' imitative behaviour. Accounting & Finance, 52(3), 663-689.

Booth, J. R., & Smith II, R. L. (1986). Capital raising, underwriting and the certification hypothesis. Journal of financial economics, 15(1-2), 261-281.

Brooks, C. (2019). Introductory econometrics for finance. Cambridge university press.

Carneiro, H. A., & Mylonakis, E. (2009). Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clinical infectious diseases, 49(10), 1557-1564.

Chen, T. (2017). Investor attention and global stock returns. Journal of Behavioral Finance, 18(3), 358-372.

Chiang, Y. M., Qian, Y., & Sherman, A. E. (2010). Endogenous entry and partial adjustment in IPO auctions: are institutional investors better informed?. The Review of Financial Studies, 23(3), 1200-1230.)

Choi, H., & Varian, H. (2009). Predicting initial claims for unemployment benefits. Google Inc, 1(2009), 1-5.

Cyr, L. A., Johnson, D. E., & Welbourne, T. M. (2000). Human resources in initial public offering firms: do venture capitalists make a difference?. Entrepreneurship Theory and Practice, 25(1), 77-92.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The journal of finance*, *66*(5), 1461-1499.

Daily, C. M., Certo, S. T., Dalton, D. R., & Roengpitya, R. (2003). IPO underpricing: A meta–analysis and research synthesis. Entrepreneurship Theory and Practice, 27(3), 271-295.

Demers, E., & Lewellen, K. (2003). The marketing role of IPOs: evidence from internet stocks. Journal of Financial Economics, 68(3), 413-437.

Goddard, J., Kita, A., & Wang, Q. (2015). Investor attention and FX market volatility. Journal of International Financial Markets, Institutions and Money, 38, 79-96.

Ibbotson, R. G. (1975). Price performance of common stock new issues. Journal of financial economics, 2(3), 235-272.

Jun, S. P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. Technological forecasting and social change, 130, 69-87.

Katti, S., & Phani, B. V. (2016). Underpricing of initial public offerings: A literature review. Universal Journal of Accounting and Finance, 4(2), 35-52.

Lee, P. M., & Wahal, S. (2004). Grandstanding, certification and the underpricing of venture capital backed IPOs. Journal of Financial Economics, 73(2), 375-407.

Loughran, T., Ritter, J. R., & Rydqvist, K. (1995). Initial public offerings: International insights. Pacific-Basin Finance Journal, 1(3), 139-140.

Loughran, T., & Ritter, J. R. (2002). Why don't issuers get upset about leaving money on the table in IPOs?. The Review of Financial Studies, 15(2), 413-444.

Lundmark, L. W., Oh, C., & Verhaal, J. C. (2017). A little Birdie told me: Social media, organizational legitimacy, and underpricing in initial public offerings. Information Systems Frontiers, 19, 1407-1422.

Ma, T., & Tsai, P. R. (2001). Are initial return and IPO discount the same thing? A comparison of direct public offerings and underwritten IPOs. A Comparison of Direct Public Offerings and Underwritten Ipos (November 2001).

Mazumder, S., & Saha, P. (2021). COVID-19: Fear of pandemic and short-term IPO performance. Finance Research Letters, 43, 101977.

Mondria, J., Wu, T., & Zhang, Y. (2010). The determinants of international investment and attention allocation: Using internet search query data. Journal of International Economics, 82(1), 85-95.

Pástor, Ľ., & Veronesi, P. (2005). Rational IPO waves. *The Journal of Finance*, *60*(4), 1713-1757.

Ritter, J. R. (2022). Initial public offerings: Underpricing. University of Florida: Warrington College of Business, 1-20.

Ritter, J. R., & Welch, I. (2002). A review of IPO activity, pricing, and allocations. The journal of Finance, 57(4), 1795-1828.

Rock, K. (1986). Why new issues are underpriced. Journal of financial economics, 15(1-2), 187-212.

Tetlock, P. C. (2015). The role of media in finance. Handbook of media Economics, 1, 701-721.

Seifter, A., Schwarzwalder, A., Geis, K., & Aucott, J. (2010). The utility of Google Trends for epidemiological research: Lyme disease as an example. Geospatial health, 4(2), 135-137.

Swamy, V., Dharani, M., & Takeda, F. (2019). Investor attention and Google Search Volume Index: Evidence from an emerging market using quantile regression analysis. Research in International Business and Finance, 50, 1-17.

Vakrman, T., & Kristoufek, L. (2015). Underpricing, underperformance and overreaction in initial public offerings: Evidence from investor attention using online searches. SpringerPlus, 4, 1-11. https://doi.org/10.1186/s40064-015-0839-4

Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. Journal of Banking & Finance, 41, 17-35.

Yahoo Finance. (2019). Negative Enterprise Value: Does It Really Mean Money for Nothing? Retrieved from: https://finance.yahoo.com/news/negative-enterprise-value-does-really-215939487.html.

Zhang, M. (2023). APIxYahooFinance. Github Repository. Retrieved from: https://github.com/MellaBI/APIxYahooFinance/blob/0d71d18b1d8c048590c83cdc66a91e944e2b1a1e/API%20Yahoo%20Finance%20Closing.

Zheng, S. X., Ogden, J. P., & Jen, F. C. (2005). Pursuing value through liquidity in IPOs: underpricing, share retention, lockup, and trading volume relationships. Review of Quantitative Finance and Accounting, 25, 293-312.

Appendix A Variable Descriptions

Table 8.1 Variable Descriptions

A complete overview of all the variables used in this research along with their respective definition and full name.

Variable	Title	Definition
IR	Initial Returns	Log(closing price) – log(offer price) of US IPOs in 2020-2021.
ASV 28	Absolute Search Volume 28 before IPO release	Absolute daily search volume numbers retrieved manually from Google.
ASV 7	Absolute Search Volume 7 before IPO release	Absolute daily search volume numbers retrieved manually from Google.
SVI 28	Relative Search Volume 28 before IPO release	Relative daily search volume numbers retrieved from Google using an API.
SVI 7	Relative Search Volume 7 before IPO release	Relative daily search volume numbers retrieved from Google using an API.
Employees	Employee Count	Number of employees at time of inception.
NYSE	New York Stock Exchange	Equals 1 if a stock goes public on the NYSE and 0 otherwise.
Size	Offering Size	Defined as the product of offering price x shares offered.
Size x Employees	Interaction term between offerings size and employee count	Believed to possess co-determining predictive power on IPO underpricing.
TV C	Trading Volume NYSE tape C	Trading volume from stocks trading both on the NYSE and NASDAQ.
TV NYSE	Trading Volume NYSE	Trading volume from stocks traded on the NYSE.

Appendix B Endogeneity Post-test, White Tests and Breusch-Pagan Tests for Heteroskedasticity

Table 9.1 Endogeneity post-IV regression test for absolute search volumes

The results of the post-IV endogeneity test for absolute search volumes.

Tests of endogeneity

H0: Variables are exogenous.
Robust regression F(1,69) = 9.03456 (p = 0.0037) (Adjusted for 70 clusters in industry)

Table 9.2 Endogeneity post-IV regression test for relative search volumes

The results of the post-IV endogeneity test for relative search volumes.

Tests of endogeneity

H0: Variables are exogenous.
Robust regression F(1,69) = 4.65902 (p = 0.0344) (Adjusted for 70 clusters in industry)

Table 9.3 Breusch-Pagan test for heteroskedasticity of the absolute search volume regression

The results of the heteroskedasticity Breusch-Pagan test for absolute search volumes.

Breusch‚ÄìPagan/Cook‚ÄìWeisberg test for heteroskedasticity for absolute search volumes

Assumption: Normal error terms Variables: mc asv28 asv7 NYSE inter_mc_employees employees tv_c tv_ny
H0: Constant variance chi2(9) = 190.46 Prob > chi2 = 0.0000

Table 9.4 Breusch-Pagan test for heteroskedasticity of the relative search volume regression.

The results of the heteroskedasticity Breusch-Pagan test for relative search volumes.

Breusch‚ÄìPagan/Cook‚ÄìWeisberg test for heteroskedasticity

Assumption: Normal error terms

Variables: mc svi28 svi7 NYSE inter_mc_employees employees tv_c tv_ny

H0: Constant variance

chi2(9) = 198.67

Prob > chi2 = 0.0000

Table 9.5 White Test for heteroskedasticity of the relative search volume regression

The results of the White test concerning heteroskedasticity for relative search volumes.

White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity
chi2(43) = 62.34
Prob > chi2 = 0.0284
Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	62.340	43	0.028
Skewness	25.440	8	0.001
Kurtosis	5.900	1	0.015
Total	93.680	52	0.000

Table 9.6 White Test for heteroskedasticity of the absolute search volume regression

The results of the White test concerning heteroskedasticity for absolute search volumes.

White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity
chi2(43) = 59.97
Prob > chi2 = 0.0443
Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	59.970	43	0.044
Skewness	22.810	8	0.004
Kurtosis	5.740	1	0.017
Total	88.520	52	0.001

Appendix C Industry and Country Frequency Tables

Table 10.1 Industry frequencies in sample and percentage of total
Frequencies and percentages of the frequency total for industries in the sample taken from the US.

Industry	Frequency	Percentage
Advertising Agencies	2	.48
Aerospace & Defense	1	.72
Apparel Manufacturing	1	.96
Apparel Retail	3	.72
Asset Management	5	1.20
Auto & Truck Dealerships	1	1.20
Auto Manufacturers	5	0.72
Banks – Regional	5	.24
Beverages - Non-Alcoholic	3	30.38
Beverages - Wineries & Distilleries	1	.24
Biotechnology	127	.72
Building Products & Equipment	1	.24
Capital Markets	3	.48
Communication Equipment	1	.48
Computer Hardware	2	.24
Consulting Services	2	.48
Consumer Electronics	1	.24
Credit Services	2	.48
Diagnostics & Research	5	1.20
Drug Manufacturers - Specialty & Generic	7	1.67
Education & Training Services	8	1.91
Electronic Components	1	.24
Electronic Gaming & Multimedia	3	.72
Engineering & Construction	1	.24
Entertainment	1	.24
Farm Products	3	.72
Food Distribution	1	.24
Furnishings, Fixtures & Appliances	1	.24
Gambling	3	.72
Grocery Stores	3	.72
Health Information Services	12	2.87
Healthcare Plans	3	.72
Home Improvement Retail	2	.48
Household & Personal Products	1	.24
Industrial Distribution	1	.24
Information Technology Services	5	1.20
Insurance – Diversified	1	.24
Internet Content & Information	6	1.44
Internet Retail	10	2.39
Leisure	1	.24
Luxury Goods	2	.48
Marine Shipping	1	.24
Medical Care Facilities	4	.96
Medical Devices	24	5.74
Medical Instruments & Supplies	6	1.44
Mortgage Finance	3	.72
Other Industrial Metals & Mining	1	.24
Packaged Foods	5	1.20
Packaging & Containers	1	.24

Personal Services	3	.72
REIT – Mortgage	3	.72
REIT – Retail	1	.24
REIT – Specialty	1	.24
Real Estate – Development	1	.24
Real Estate Services	2	.48
Recreational Vehicles	1	.24
Residential Construction	1	.24
Restaurants	5	1.20
Semiconductors	4	.96
Shell Companies	6	1.44
Software – Application	50	11.96
Software – Infrastructure	29	6.94
Solar	3	.72
Specialty Business Services	3	.72
Specialty Industrial Machinery	1	.24
Specialty Retail	4	.96
Staffing & Employment Services	2	.48
Tobacco	1	.24
Travel Services	1	.24
Utilities – Diversified	1	.24
Utilities – Renewable	2	.48
Waste Management	1	.24

Table 10.2 Country frequencies and percentages in the sample
Frequencies and percentages of the frequency total for countries in the sample taken from the US.

Country	**Frequency**	**Percentage**
Australia	3	.72
Belgium	1	.24
Brazil	4	.96
Canada	12	2.87
Cayman Islands	2	.48
China	22	5.26
Denmark	2	.48
France	3	.72
Germany	8	1.91
Ireland	2	.48
Israel	13	3.11
Italy	2	.48
Japan	1	.24
Jersey	1	.24
Netherlands	2	.48
Russia	1	.24
Singapore	3	.72
South Korea	2	.48
Sweden	1	.24
Switzerland	4	.96
United Kingdom	7	1.67
United States	321	76.79
Uruguay	1	.24