Bank Parikrama – A Journal of Banking & Finance
Volume 50, No. 1, June 2025 (pp. 181-203)
ISSN: 1019-7044
DOI: http://doi.org/10.64204/BP1019-7044/50(1)/A20254

Bid-Ask Spread in Financial Market: Insights from Search Engine Query Data

- Mohd. Anisul Islam*
- Noyon Islam**

Abstract

This paper investigates the capabilities of query data for 'company name' to provide insights into the movement of bid-ask spread of stock, which is a basic component of transaction cost. The magnitude of bid-ask spread has an impact on measuring trading performance. Results from econometric techniques on a sample of 497 stocks reveal that the bid-ask spread of a stock is correlated with the search volume of the corresponding company name. Furthermore, we find that the stocks of more searched companies are likely to be traded at a lower bid-ask spread. However, if search is motivated by negative sentiment, the bid-ask spread will rise. This finding illustrates that the fluctuation of bid-ask spread can be anticipated by query data that will assist investors to make trading decisions prudently.

Keywords: Bid-Ask Spread, Query Data, Company Name, Negative Sentiment, Transaction Cost
JEL Classification: G11, G12, G24

1. Introduction

Bid-ask spread – the difference between the highest price buyers are willing to pay and the lowest price sellers are willing to accept- is one of the significant components of transaction costs and a fundamental measure of liquidity in the financial market. This simple measure captures complex market dynamics like volatility, information asymmetry, order processing cost, inventory cost, and other factors. Moreover, a higher spread, while compensating market makers for financial intermediation services, can dramatically reduce the desired return for any portfolio of investors. Since understanding of this spread has implications for market efficiency, price discovery, and overall cost of capital, both market makers and investors are concerned about the cross-section of the bid-ask spread

---

* Mohd. Anisul Islam is an Assistant Professor, Department of Finance, University of Dhaka, Bangladesh, Email: ai.fin@du.ac.bd; **Noyon Islam is a Lecturer, Department of Finance, University of Dhaka, Bangladesh, Email: noyon@du.ac.bd. ** The views expressed in this paper are the author's own. The authors gave thanks to Neha Gupta, module leader and teaching assistants at the Big Data Analytics Module, MSc Finance Course 2019-20, Warwick Business School, University of Warwick for their helpful suggestions.

of securities. Considering the proliferation of search engine-based information gathering, This study examines whether variations in search engine query data for a company's name can explain the bid-ask spread of its securities.

The bid-ask spread indicates the cost of any transaction service. The procedure, process, and performance of transaction services are discussed in the domain of market microstructure theory. Regarding financial intermediation, assets cannot be transferred from one party to another without mentioning the bid and ask price. Conventional market microstructure theory implies that this spread is primarily determined by liquidity, volatility, information asymmetry, order processing cost, and inventory cost, among other factors. Empirical evidence finds that price level, market activity, return variance, and competition among market makers explain the cross-sectional variation in bid-ask spread (Menyah and Paudyal, 1996). However, the digital revolution and proliferation of internet-based information gathering have introduced a new dimension of information flow and investor behavior, which remain underexplored in the context of market liquidity.

Most of the spread-determining factors, such as market activity, price level, return variance etc., are nowadays significantly influenced by the behavior of investors who often seek stock-related information using search engines to make decisions. The advent of technology has fundamentally transformed how investors acquire and process information about financial securities. Google searches, Bing queries, and other search engine activities are likely to provide unprecedented real-time insights into investor attention, information demand, and market sentiment. The behavioral data reflects investor attention and the process of gathering information that typically occurs before making trading decisions. As such, it may offer a useful way to anticipate market liquidity conditions. Therefore, we can suggest that analyzing search engine query data could provide valuable insights into changes in the bid-ask spread.

In this perspective, Investors nowadays utilize search engines for searching public information about companies to update their expectations about future outcomes from investment in a stock. Public announcements of events are an essential means of informing traders of information. Many investors calculate the fundamental values of assets using this newly acquired information. But very few steps are taken to quantify the information and use the information to maximize return. Griffin et al (2011) quantify the effect of financial media on stock prices and measure the cross-country differences in stock price reactions to public announcements information in the financial media. They also find that public news outlets offer an enormous amount of information regarding shifts in company value. So, non-professional investors use Google for obtaining public information available in different websites. This makes Google the number one search engine. Hence, the Search Volume Index (SVI) released by search engines like Google is likely to reflect investors' attention and sentiment. Search volume data indicates interest and sentiment of investors since people usually search for the term that creates interest or draws attention for specific reasons.

This study intends to discern the relationship between bid-ask spread of stocks and the query data for company name, contributing to the growing evidence on digitally mediated investor behavior and market microstructure. We explore whether search intensity, reflected by the search volume index of Google, can explain variations in bid-ask spreads across different securities. We address the primary research question: does increased search activity predict changes in bid-ask spread? What is the channel through which search behavior affects market liquidity? Does this effect vary across different types of securities, like whether variation in market capitalization of securities intervenes this relationship between query data and bid-ask spread? Searching by company name returns both financial and non-financial information of the company which assists investor in making transaction decision. As higher searching of company name reduces information asymmetry and signals investor attention, it can result in reduced bid-ask spread. So negative relationship between query data and bid-ask spread is desirable.

The bid-ask prices serve as critical determinants of the equilibrium market price for a wide range of securities, thereby facilitating efficient financial intermediation. In the short run, both the pattern of transactions and the fluctuation in stock prices are influenced by these bid-ask quotes. Investors typically purchase securities at the ask price and sell them at the bid price. Market makers, including brokers and dealers, play a pivotal role in setting the equilibrium market prices by considering the bid-ask spread. Their trading decisions are guided by market orders, which reflect the best available price at any given moment. By matching prices and executing trades, market makers can ascertain the bid-ask spread from the available data, which aggregates the current bids and asks within the financial market.

In anonymous market conditions, if a transaction occurs at the bid price, this can lead to a decrease in both the bid and ask prices, potentially discouraging the seller from further selling and instead motivating them to purchase more securities. Conversely, purchases at the ask price may trigger additional selling activities (HR Stoll, 2003). Research suggests a negative correlation between the bid-ask spread and trading activity of specific securities within competitive markets. However, increased price volatility may lead to a wider bid-ask spread, as volatile stocks typically embody greater risks of loss and correspondingly higher return potential. Studies examining the relationship between bid-ask spread and price levels have produced mixed findings; some researchers assert a positive correlation, while others argue for an inverse relationship (Aitken and Frino, 1996).

To address market anomalies attributed to investor behavior, Robert Shiller (2003) developed the theory of behavioral finance, which intersects economics, finance, and psychology. The advancement of technology has enabled broader access to information, allowing investor sentiment to be forecasted using trends in technology. For instance, the mood identified on social media platforms, such as Twitter, can potentially predict movements in the Dow Jones Industrial Average Index (DJIA) with considerable accuracy (Bollen et al., 2011). Additionally, during bearish market conditions, there is a noted increase in traffic on platforms like Wikipedia as investors seek detailed information regarding various companies (Moat et al., 2013).

Li et al. (2014) found a significant relationship between user engagement with the media and stock price movements, proposing a quantitative media strategy to forecast stock prices. Fang and Peress (2009) investigated the relationship between media coverage of firms and their predicted returns, highlighting that firms with less media coverage tend to exhibit higher stock returns, especially among smaller firms. They also explored the effect of media coverage on mutual funds, revealing that mutual funds often gravitate towards stocks with extensive media coverage, although this inclination varies significantly across different mutual funds. Information regarding stock movements is disseminated through various media channels and social platforms, enabling investors to better comprehend price dynamics in competitive markets. For example, Atkins et al. (2018) demonstrated that stock price movements could be accurately predicted based on information from news sources, noting that the impact of official news announcements on the Dow Jones Industrial Average and NASDAQ typically manifests with a time lag of about 20 minutes, evaluated through Reuters Stock Market text data.

In terms of data analysis, considerable text data can be quantified to predict future market trends. Joseph et al. (2011) found that text data generated from online search intensity could be indicative of trading volume and stock returns. They suggested that more volatile equities, which are inherently difficult to arbitrage, tend to react more strongly to search frequency compared to less volatile stocks. Specifically, search volumes related to individual stocks demonstrate a positive relationship with both trading volumes and return volatility.

The Google search engine, widely recognized as a premier search tool, offers extensive information across various industries through its "Search Volume Index" (SVI) feature available via Google Trends. Da, Engelberg, and Gao (2011) posited that the SVI embodies a more precise measurement of investor attention compared to traditional measures. They also argued that Google Trends is indicative of "revealed attention," as queries regarding a stock in Google stem from an interest in it. The SVI consists of time-series data, allowing for observation of the frequency with which certain terms are searched. Moreover, Yang, Santillana, and Kou (2015) utilized Google Flu Trends to enhance influenza tracking methods, signifying the robustness and adaptability of Google’s data in capturing public sentiment.

Da et al. (2015) proposed a daily sentiment index designed to measure the financial concerns of American households through Google search query data. Ding and Hou (2015) explored GSV to assess investor sentiment, concluding that heightened investor attention correlates with reduced bid-ask spreads, thereby enhancing liquidity. They noted limitations, however, as the GSV for smaller firm tickers often returned zero values.

Swamy and Dharani (2019) contended that Google Search Volume (GSV) can capture both the directional and quantitative aspects of excess returns in the market. They found that a higher GSV correlates with higher excess returns, indicating the critical role of investor sentiment under these circumstances.

Investor sentiment is a nuanced reflection of anticipated future risks, which may not always align with factual data; however, it plays a significant role in influencing stock price movements. Search data linked to trading symbols serves as a reliable proxy for gaugeing investor sentiment (Z He et al., 2019; Mian and Sankaraguruswamy, 2012). The stock market is, to a certain extent, directed by investor sentiments, which in turn affects trading behaviors. Increased search volumes for specific stocks often signal heightened investor attention. When investor sentiment turns negative, a trend emerges where stocks are typically sold, and this tendency intensifies for riskier securities. Accordingly, Ana Brochado (2020) developed dual sentiment measures based on Google data—positive and negative indices. The findings indicated a correlation between stock market returns and trading volumes with these sentiment-based indices; notably, the positive index exerted a greater influence on stock returns compared to the negative index. Investor sentiment stems from factors such as noise trading and risk aversion, with the intensity of the Search Volume Index (SVI) correlating with security prices.

An elevated SVI coupled with increased transaction volumes reflects heightened trading activity and improved liquidity for specific stocks. Such conditions foster market efficiency and, consequently, lower bid-ask spreads. Therefore, it is likely to observe a lower bid-ask spread in conjunction with a higher SVI for a given stock. Joseph, Babajide Wintoki, and Zhang (2011) suggested that high SVI can predict anomalous stock returns. They advanced a long-short investment strategy, advocating for shorting low-SVI stocks while going long on those with high SVI, rebalancing the portfolio weekly. Their research posits that analyzing search volume can provide actionable insights into the overall performance of financial markets, though this is contingent on the specific search terms used. High purchasing interest in certain stocks suggests positive market sentiment, predicting outperformance of these stocks. Conversely, declining trend indications can be observed when search terms lose popularity as market conditions deteriorate.

Preis et al. (2013) explored the utilization of generic finance-related search keywords or firm-specific terms to predict stock market fluctuations. Their goal was to develop a portfolio management strategy to outperform the stock market index via Google Search Data analysis. Over a span of seven years, they discovered that trading based on Google searches for specific keywords could outperform market indices by up to 310%. In contrast, Joseph et al. (2011) cautioned that high transaction costs associated with frequent portfolio rebalancing might negate the profitability of trading strategies based on search intensity. They acknowledged, however, that the volume of online searches is a reliable indicator of anomalous trading activity and abnormal stock returns. Similarly, Bijl et al. (2016) asserted that profitable trading strategies could yield abnormal returns, provided transaction costs are considered. Their analysis of S&P 500 data from 2008 to 2013, during the global financial crisis, revealed a counterintuitive finding whereby high Google Search Volume correlated with lower stock returns on a weekly basis, with subsequent reversals.

This analysis aims to enhance the existing literature in several keyways. First, it demonstrates how the act of searching for stock-related information on search engines can diminish information asymmetry, resulting in lower transaction costs. Second, it elucidates how search engine-provided insights can shape investor attention and sentiment, which subsequently impacts the bid-ask spread. Finally, accessing such insights from query data equips investors with a pre-trade estimation of transaction costs, aiding in the optimal placement of trade orders. The succeeding section will delve into the methodological framework and present the findings of this research.

The primary objective of this paper is to establish causal relationships between search behavior and spread dynamics while controlling for traditional spread determinants identified in prior literature. This study leverages comprehensive datasets combining high-frequency bid-ask spread data with search volume index data available from Google for five-year time horizons. It is anticipated that similar patterns will manifest in developing markets, including Bangladesh. Notably, the Bangladesh Securities and Exchange Commission (BSEC) has instituted a policy limiting daily stock price declines to no more than 3%. Such circuit breakers in developing economies like Bangladesh are expected to temper speculative activities among investors (MF Ahmed, 2013). Consequently, a negative correlation between the SVI and bid-ask spread is expected in the Bangladeshi context.

2. Methods and Results

2.1 Data

We use closing bid and ask prices on Monday of every week, of 497 US stocks out of 505 stocks of S&P 500 index as of March 1, 2020. This weekly dataset also includes transaction volume and return data for the period from January 1, 2015, to December 31, 2019. All series have been obtained from the CRSP database[1] which provides historical data of stock market. For query data in search engine, We use search data for 500 company names from ‘‘Google Trends’’. It is to be noted that we retrieved search volume data for the period from January 1, 2015, to December 31, 2019, by accessing the Google Trends website (http://www.google.com/trends) on March 7, 2020, using gtrendsR package on R studio. However, when we attempted to retrieve more recent query data for 497 stocks together from “Google Trends” using the same gtrendsR package, we could not retrieve recent data because recent updates to Google Trends have limited access to “Google Trends” data, particularly for bulk keyword queries. However, retrieval of several keywords from “Google Trends” using the same gtrendsR package is still functional.

---

[1] https://wrds-www.wharton.upenn.edu/

Weekly data in ‘‘Google Trends’’ mean index is computed by summing the search volume during that week and reported on every Sunday. No trading data are reported on weekend. To merge the query data with trading data, we forward the date of ‘‘Google Trends’’ data by one day. Specifically, how the query for company name made in ‘‘Google Trends’’ from Sunday to Saturday in every week influences the closing bid-ask spread quoted on Monday of following week will be examined. After necessary transformation, finally we use 112,319 observations of 497 companies for range of T= 9-236.

Weekly frequency of data is used to cover wider period as daily frequency data are available for shorter period. ‘‘Google Trends’’ computes “Search Volume Index (SVI)” in scale of 0-100 for specific terms by computing total searches which have been done for specific term relative to the total number of searches done on Google for specific time period – here we use the 497 company names. Either company name or stock ticker is usually used to explore information. The reasons for using company name as keyword are it generates large quantity of firm related information rather than stock related information and the problem of generic meaning of several ticker names (e.g. ticker for Ford and Caterpillar are “F” and “CAT” correspondingly) can be avoided. Since exact company names e.g. Apple Inc. – may return with poor SVI, we optimize the name of company by setting keyword to common abbreviations - e.g. Apple. Besides, if we compare the search index for Apple as technology company and Apple as generic name, no significant difference in SVI is seen (Appendix: Figure i). List of a portion of companies along with corresponding ticker and keyword has been provided in appendix.

2.2 Variables

In this study bid-ask spread is calculated by taking difference between ask and bid price relative to midpoint of bid and ask price. To follow the convention, this measure is expressed in basis point (BP) (0.01=1%=100BP). we use the SVI provided by “Google Trends” to capture the query data for company name. This index is already normalized with respect to total searches for 497 company names. To separate the incremental effect of query data reflected by SVI on bid-ask spread, we control for market capitalization, trading volume, and return volatility. Market capitalization is measured by multiplying outstanding shares with price and trading volume reflects total number of shares sold on a day. Return volatility is measured based on squared return. Stocks have been categorized based on the type of market capitalization since it is an important characteristic which causes variation of spread among stocks. A table for category of market capitalization can be found in appendix. Descriptive statistics of the spread and SVI can be found in Table 1 and Table 2 respectively.

Table 1. Descriptive Statistics of Bid-ask Spread (in bp)

Category	Mean	Median	Q1	Q3	Minimum	Maximum
Mega-cap stocks	2.2591	1.8187	1.1065	2.7089	0.0553	37.7905
Large-cap stocks	2.2230	1.8168	1.2244	2.7045	0.0274	113.2909
Mid-cap stocks	4.0144	2.6835	1.7891	4.0891	0.0363	1030.71
Small-cap stocks	21.447	7.681	4.090	43.011	2.908	61.538

The central tendency and dispersion of bid-ask spread varies significantly across the type of stocks with mean spread ranging between 2.26bp and 21.45bp. Small-cap stocks are traded at higher bid-ask spread compared to large-cap stocks because of higher illiquidity and information asymmetry associated with small-cap stocks.

Table 2. Descriptive Statistics of SVI (in 0-100 scale)

Type	Mean	Median	Q1	Q3	Minimum	Maximum
Mega-cap stocks	56.35	57.00	45.00	67.00	7.00	100.00
Large-cap stocks	51.87	53.00	35.00	68.00	1.00	100.00
Mid-cap stocks	52.37	54.00	35.00	69.00	1.00	100.00
Small-cap stocks	54.57	55.00	42.00	66.25	12.00	100.00

Central tendency of SVI reflects that on an average company name of mega-cap and small-cap stocks are searched more compared to other stocks which implies that large and very small companies are searched more frequently. However, high dispersion of SVI can also be observed from the difference between minimum and maximum value.

2.3 Methods

To investigate how query data influence, the bid-ask spread, we estimate the following equation where bid-ask spread (Spread) is regressed against the search volume index (SVI), market capitalization (MarketCap), trading volume (Volume), and return volatility (Volatility). All variables are expressed in natural logarithm as it enhances statistical properties of the residuals and enables direct interpretation.

$$Spread_{it} = \beta_0 + \beta_1 SVI + \beta_2 MarketCap + \beta_3 Volume + \beta_3 Volatility + \eta_i + \upsilon_{it}, \quad \text{..................(1)}$$

β_0 is intercept, β_1, $\beta_2, and\ \beta_3$ are estimated coefficients of search volume index, market capitalization, trading volume, and return volatility respectively. η_i captures unobservable heterogeneity and υ_{it} captures disturbance. we apply correlations and panel regression along with graphical analysis. The stationarity of the variables is assessed using Augmented Dickey–Fuller test (lag order 7 for weekly data) and the results shown in Table 3 confirm stationarity of the variables. Density distribution of variables can be observed from Figure 2. Normal distribution of the residuals is assumed.

Table 3. ADF Test Statistics

	Spread	SVI	MarketCap	Volume	Volatility
Test Statistic	-38.17	-24.678	-17.03	-22.325	-94.37
p-value	0.01	0.01	0.01	0.01	0.01

Panel specific regressions along with pooled regression is conducted for recognizing heterogeneity across the companies and omitted variable bias. Panel regressions include fixed-effect, random-effect, and first-differencing estimate. As it is difficult to assume absence of relationship between firms and explanatory variables, fixed-effect estimate is preferred over random-effect. Moreover, Hausman test ($Chi^2 = 363$, df = 4, p-value <0.001) confirms that fixed effect is more efficient. Additionally, first differencing estimate, which effectively deals with serial correlation problem, has been used. Correlations presented in figure 2 confirm that econometric models are free from multicollinearity problem as correlations between pair of explanatory variables are less than 0.50. Breusch-

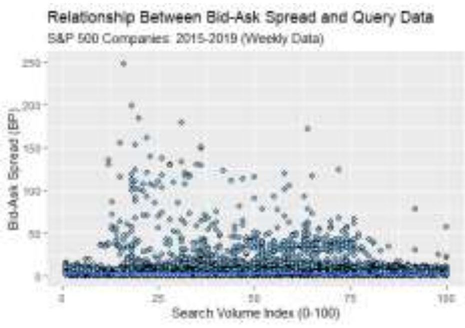
Pagan test and Breusch-Godfrey/Wooldridge tests are used to test heteroskedasticity and serial correlation. Newey-West HAC standard errors and covariances are employed in the estimation as these standard errors are robust for both heteroskedasticity and serial correlation problem.

2.4 Results

Figure 1 presents visualization of names of the 100 companies which were searched most (SVI>=70) in Google during the study period. The font size of the words reflects the average of SVI. It can be immediately observed that companies name (i.e. mastercard and yum) were searched more relative to the companies (i.e. flir and ametek).

Figure 1: Word cloud for the company names of which weekly mean of SVI is at least 70. The font size of the company name is proportionate to the mean value of SVI. Larger font size indicates the company name, which is searched more often.

Figure 2: Scatterplot shows the interaction of search volume index (x-axis), and bid-ask spread in basis point (y-axis). Negative trend line indicates higher query data are associated with lower bid-ask spread.



Scatterplot in Figure-2 can be observed to approximate the relationship between bid-ask spread and SVI visually. Bid-ask spread tends to decrease as the relative number of query data for company name increases. Slightly negative trend line indicates that stocks of the more frequently searched companies are likely to be traded in lower bid-ask spread. Though this negative trend line may not appear clear visually, it can be confirmed from Figure 3 by looking at the value of correlation between SVI and Spread (Pearson correlation coefficient $r= -0.0282$, $df=112{,}317$, $p< 0.001$). This correlation coefficient has been found statistically significant which implies that this relationship measured by sample data can be generalized for other samples. Positive correlation between SVI and volume implies higher attention reflected by SVI raises volume. Negative correlation between SVI and volatility suggests more searches generates more information which lessens volatility.

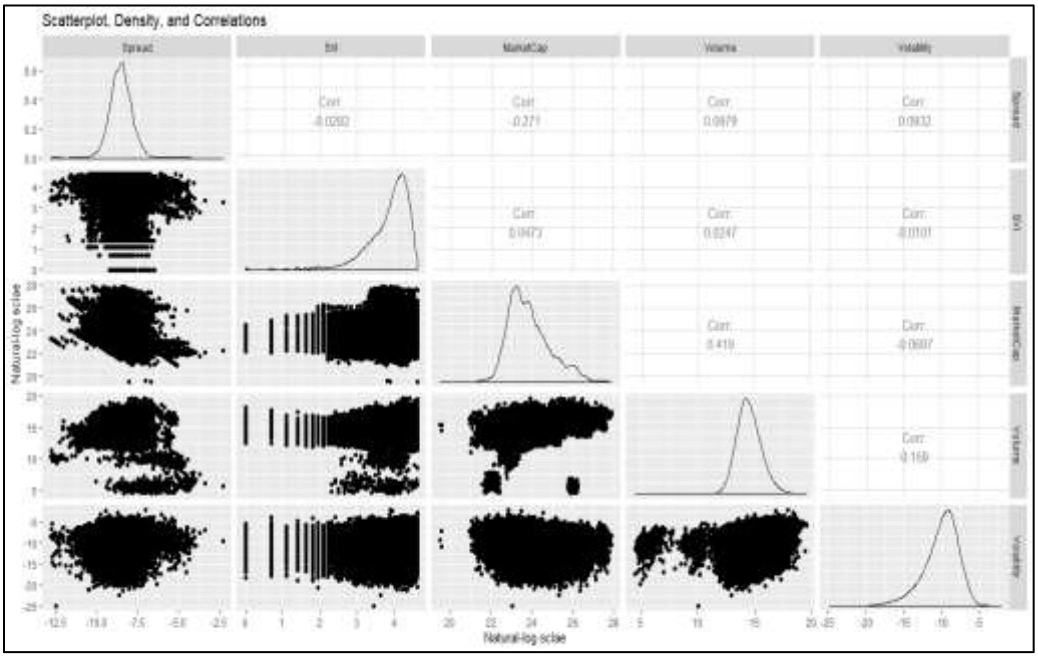


Figure-3 Left part exhibits the scatterplots of each pair of variables with log scale in both x-axis and y-axis. Right part displays the Pearson correlation coefficient between bid-ask spread and query data and other control variables. Density distribution of the variables is drawn on the diagonal which reflect left-skewed distribution of SVI. Negatively skewed distribution of SVI indicates that most of the companies are highly searched. All correlation coefficients are significant at 1% significance level.

The linkage between query data and bid-ask spread for one proxy of S&P 500 stocks - Microsoft Corporation can be analyzed from the Figure 4 where both variables have been plotted in correspondence with the date. In ending of the 2015, when the search volume index of Microsoft peaked, the bid-ask spread dropped. Conversely, when search volume index dropped before beginning of 2019, significant spiral of bid-ask spread happened. However, the significant correlation coefficient of 0.3645 between SVI and bid-ask spread testifies that two variables often move in same direction.

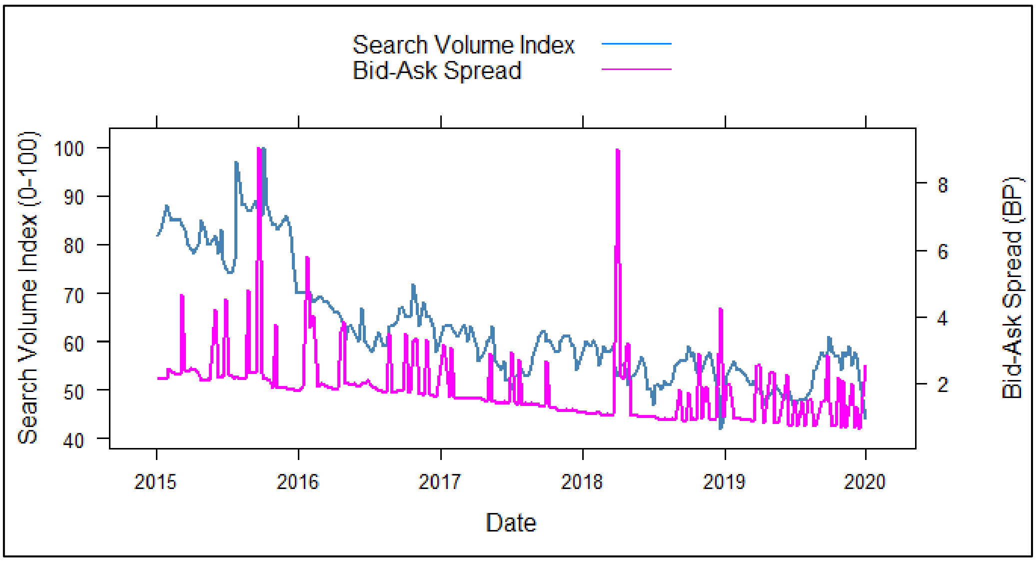


Figure-4. The relative number of queries for 'Microsoft' (Left Y axis) is plotted for the time period from 2015 to 2019. The bid-ask spread for Microsoft (Right Y axis) is shown for the same period. Noticeable co-movement of query data and spread over time can be observed.

Table-4 offers the coefficients along with standard error, estimated from the different models of panel regression. In addition, Table-5 exhibits the coefficients from fixed effect models based on the type of stocks.

Table 4. Bid-ask Spread and Query Data Panel Regression

Dependent variable: Bid-ask spread (Spread)	Pooling estimate	Fixed-effect estimate	Random-effect estimate	First-difference estimate
Search volume index (SVI)	-0.0185 *** (0.0046)	0.01345** (0.0064)	0.0077 (0.0064)	0.0141* (0.0084)
Market capitalization (MarketCap)	-0.2502*** (0.0066)	-0.5135*** (0.0086)	-0.4832*** (0.0081)	-0.9730*** (0.0426)
Trading volume (Volume)	0.1413*** (0.0096)	0.0221*** (0.0044)	0.0340*** (0.0045)	-0.0214*** (0.0041)
Return volatility (Volatility)	0.0081*** (0.0012)	0.0058*** (0.0007)	0.0053*** (0.0007)	0.0021*** (0.0006)
Intercept	-4.4544 *** (0.0902)	N/A	2.5466*** (0.2099)	0.0011 (0.0019)
Observations	112,319	112,319	112,319	112,319
R2	0.1290	0.0758	0.0895	0.0049
Adjusted R^2	0.1290	0.0717	0.0894	0.0048

The number inside the parenthesis represents the Newey-West standard error. ***, **, and * indicate significance at 1%, 5%, and 10% significance level respectively.

3. Discussion and Conclusion

This study attempts to quantify the relationship between bid-ask spread of stocks and the query data for company name. The results in Table 4 and 5 suggest that regression model can explain considerable amount of variability in bid-ask spread with adjusted R^2 ranging between 7.17% and 12.90% except first-differencing estimate. The coefficient of query data measured by SVI has been found statistically significant in all regression models but random-effect estimate. Besides, when fixed-effect estimates are observed separately based on type of stocks in Table 5, it is evident that coefficient of query data is significant in determination of bid-ask spread of both large-cap and mid-cap stocks which in fact constitute the largest portion of whole sample.

The coefficient of the SVI is significantly negative in pooling regression which implies that higher level of attention reflected by SVI increases liquidity in the market which reduces bid-ask spread. These findings are consistent with findings of Ding and Hou (2015). In fixed-effect regression for large-cap stock, similar findings suggest that low spread is caused by higher attention of investors measured by SVI. More search of company information reflected by SVI reduces information asymmetry and causes lower volatility. Bid-ask spread declines when volatility shrinks. Although coefficient of SVI shows negative sign across all type of stocks except mid-cap stocks, it is not statistically significant in case of mega-cap and small-cap stocks.

Table 5. Bid-ask Spread and Query Data Fixed-effect Regression by Stock Type

Dependent variable: Bid-ask spread (Spread)	**Mega-cap stocks**	**Large-cap stocks**	**Mid-cap stocks**	**Small-cap stocks**
Search volume index (SVI)	-0.0292	-0.0160**	0.0442***	-0.3557
	(0.0592)	(0.0080)	(0.0121)	(0.2820)
Market capitalization (MarketCap)	-0.3403***	-0.4700***	-0.5902***	-1.3292***
	(0.0669)	(0.0110)	(0.0235)	(0.4091)
Trading volume (Volume)	0.1942***	0.0193***	-0.0093	-0.0060
	(0.0297)	(0.0050)	(0.0092)	(0.0663)
Return volatility (Volatility)	-0.0022	0.0064***	0.0040***	0.0149
	(0.0042)	(0.0008)	(0.0015)	(0.0155)
Observations	4,533	84,652	23,038	96
R^2	0.0331	0.0450	0.0626	0.2100
Adjusted R^2	0.0256	0.0396	0.0534	0.1370

The number inside the parenthesis represents the Newey-West standard error. ***, **, and * indicate significance at 1%, 5%, and 10% significance level, respectively.

However, the positive coefficient of SVI in fixed-effect and first-difference regression for overall sample can be interpreted in following way. People search for more information of the companies when their sentiment - perception about future risks of companies, is high which can cause reduction in liquidity and spiral of volatility for the stock in market. This in turn results in higher bid-ask spread. This finding is further revealed in separate estimate for mid-cap stock. Apart from the SVI, coefficients of other control variables have been found statistically significant with expected sign. Stocks with higher market capitalization trade at lower spread for higher liquidity.

These findings are consistent with Preis, Reith and Stanley (2010) who assert that present attractiveness for trading stock is reflected by search volume but it does not reflect preference either for buying or selling transactions. If query data reflect optimism of people about the prospect and performance of the companies, higher search volume will reduce bid-ask spread. However, if query data rises for concern over outlook of company, bid-ask spread will become higher for higher search volume.

Given the recent restrictions imposed by Google on retrieving more recent data using the R-based methodology (via the gtrendsR package), there is a compelling opportunity for future research in the same area. Such research can encompass analysis of texts that effectively indicate positive or negative interest for the company. Analysis of query data facilitates prediction for bid-ask spread movement in the financial market. The implications of this study extend beyond the academic interest. For market intermediaries, understanding the reflection of investor attention on bid-ask spread dynamics can suggest optimal pricing and risk management strategies. For regulators, the causal relation between information seeking behavior and market liquidity can guide policy decisions related to market structure and transparency requirements. For retail and institutional investors, awareness of how search behavior affects transaction costs can help in improving trading strategies.

References

Ahmed, M. F., & Ali, M. B. (2013). Efficiency in Bangladesh stock market behavior: empirical evidence. *Journal of Business Studies*, *34*(3), 79-103.

Aitken, M., & Frino, A. (1996). The determinants of market bid ask spreads on the australian stock exchange: Cross-sectional analysis. *Accounting & Finance*, *36*(1), 51-63. https://doi.org/10.1111/j.1467-629X.1996.tb00298.x

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of economic perspectives*, *21*(2), 129-151. https://doi.org/ 10.1257/ jep.21.2.129

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. https://doi.org/10.1007/s11408-011-0165-y

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. https://doi.org/ 10.1093/ rfs/hhm079

Bijl, Laurens, Glenn Kringhaug, Peter Molnár, and Eirik Sandvik. "Google searches and stock returns." *International Review of Financial Analysis* 45 (2016): 150-156. https://doi.org/10.1016/j.irfa.2016.03.015

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, *2*(1), 1-8. https://doi.org/ 10.1016/j.jocs.2010.12.007

Brochado, A. (2020). Google search based sentiment indexes. IIMB Management Review, 32(3), 325-335. https://doi.org/10.1016/j.iimb.2019.10.015

Copeland, T. E., & Galai, D. (1983). Information effects on the bid-ask spread. *the Journal of Finance*, *38*(5), 1457-1469. https://doi.org/ 10.1111/j.1540-6261.1983.tb03834.x

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The journal of finance*, *66*(5), 1461-1499. https://doi.org/10.1111/j.1540-6261.2011. 01679.x

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies*, *28*(1), 1-32. https://doi.org/10.1093/rfs/hhu072

Ding, R., & Hou, W. (2015). Retail investor attention and stock liquidity. *Journal of international financial markets, institutions and money*, *37*, 12-26. https://doi.org/10.1016/j.intfin.2015.04.001

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The journal of finance*, *64*(5), 2023-2052. https://doi.org/ 10.1111/ j.1540-6261.2009.01493.x

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The journal of finance*, *64*(5), 2023-2052. https://doi.org/10.1111/ j.1540-6261.2009.01493.x

Griffin, J. M., Hirschey, N. H., & Kelly, P. J. (2011). How important is the financial media in global markets?. *The Review of Financial Studies*, *24*(12), 3941-3992. http://dx.doi.org/10.2307/41302005

He, Z., He, L., & Wen, F. (2019). Risk compensation and market returns: The role of investor sentiment in the stock market. *Emerging Markets Finance and Trade, 55*(3), 704-718. https://doi.org/10.1080/1540496X.2018.1460724

Huang, M. Y., Rojas, R. R., & Convery, P. D. (2020). Forecasting stock market movements using Google Trend searches. *Empirical Economics*, *59*, 2821-2839. https://doi.org/10.1007/s00181-019-01725-1

Joseph, K., Wintoki, M. B., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, *27*(4), 1116-1127. https://doi.org/10.1016/j.ijforecast.2010.11.001

Joseph, K., Wintoki, M. B., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, *27*(4), 1116-1127. https://doi.org/10.1016/j.ijforecast.2010.11.001

Joseph, K., Wintoki, M. B., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, *27*(4), 1116-1127. https://doi.org/10.1016/j.ijforecast.2010.11.001

Kostopoulos, D., Meyer, S., & Uhr, C. (2020). Google search volume and individual investor trading. Journal of Financial Markets, 49, 100544. https://doi.org/10.1016/j.finmar.2020.100544

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., & Chen, Y. (2014). The effect of news and public mood on stock movements. *Information Sciences*, *278*, 826-840. https://doi.org/10.1016/j.ins.2014.03.096

Menyah, K., & Paudyal, K. (1996). The determinants and dynamics of bid-ask spreads on the London Stock Exchange. *Journal of Financial Research*, *19*(3), 377-394. https://doi.org/10.1111/j.1475-6803.1996. tb00220.x

Mian, G. M., & Sankaraguruswamy, S. (2012). Investor sentiment and stock market response to earnings news. *The Accounting Review, 87*(4), 1357-1384. https://doi.org/10.2308/accr-50158

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T. (2013). Quantifying Wikipedia usage patterns before stock market moves. *Scientific reports*, *3*(1), 1801. https://doi.org/10.1038/srep01801

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific reports*, *3*(1), 1-6. https://doi.org/10.1038/srep03578

Preis, T., Reith, D., & Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1933), 5707-5719. https://doi.org/10.1098/rsta.2010.0284

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of economic perspectives*, *17*(1), 83-104. https://doi.org/10.1257/089533003321164967

Stoll, H. R. (2003). Market microstructure. In *Handbook of the Economics of Finance* (Vol. 1, pp. 553-604). Elsevier. https://doi.org/10.1016/S1574-0102(03)01013-6

Swamy, V., & Dharani, M. (2019). Investor attention using the Google search volume index–impact on stock returns. *Review of Behavioral Finance*, *11*(1), 56-70. https://doi.org/10.1108/RBF-04-2018-0033

Yang, S., Santillana, M., & Kou, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. *Proceedings of the National Academy of Sciences*, *112*(47), 14473-14478. https://doi.org/10.1073/pnas.1515373112

Yang, S., Santillana, M., & Kou, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. *Proceedings of the National Academy of Sciences*, *112*(47), 14473-14478. https://doi.org/10.1073/pnas.1515373112

Appendices

Appendix: Figure i

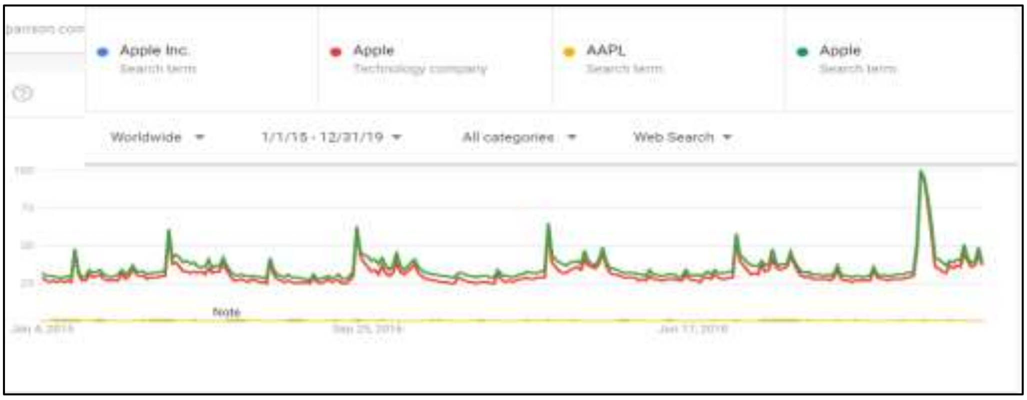


Figure i. The line chart for the relative number of queries for "Apple Inc.", "Apple", "AAPL", and "Apple" keywords for the time period from January 1, 2015, to December 31, 2019. It has been taken from Google Trends. Very poor search volume for exact company name and ticker name can be observed. Besides strong co-movement of query data for optimized keyword – Apple and query data for generic name – Apple can be noticed.

Appendix Table i. Type of stock based on the category of market capitalization

Type of Stock	Market Capitalization Range
Mega-cap	More than $200 billion
Large-cap	$10 billion to $200 billion
Mid-cap	$2 billion to $10 billion
Small-cap	$300 million to $2 billion
Micro-cap	$50 million to $300 million

Appendix Table ii. List of a portion of companies along with their ticker and keyword

Company	Ticker	Keyword
Microsoft Corporation	MSFT	"microsoft"
Apple Inc.	AAPL	"apple"
Amazon.com Inc.	AMZN	"amazon"
Facebook Inc. Class A	FB	"facebook"
Berkshire Hathaway Inc. Class B	BRK.B	"berkshire hathaway"
Alphabet Inc. Class A	GOOGL	"alphabet"
Alphabet Inc. Class C	GOOG	"google"
JPMorgan Chase & Co.	JPM	"jp morgan"
Johnson & Johnson	JNJ	"johnson and johnson"
Visa Inc. Class A	V	"visa"
Procter & Gamble Company	PG	"procter gamble"
Mastercard Incorporated Class A	MA	"mastercard"
AT&T Inc.	T	"at&t"
UnitedHealth Group Incorporated	UNH	"unitedhealth"
Intel Corporation	INTC	"intel"
Home Depot Inc.	HD	"home depot"
Bank of America Corp	BAC	"bank of america"
Verizon Communications Inc.	VZ	"verizon"
Exxon Mobil Corporation	XOM	"exxon mobil"
Walt Disney Company	DIS	"disney"
Coca-Cola Company	KO	"coca cola"
Merck & Co. Inc.	MRK	"merck"
Comcast Corporation Class A	CMCSA	"comcast"
Pfizer Inc.	PFE	"pfizer"
PepsiCo Inc.	PEP	"pepsico"
Chevron Corporation	CVX	"chevron"
Cisco Systems Inc.	CSCO	"cisco systems"
Adobe Inc.	ADBE	"adobe systems"
NVIDIA Corporation	NVDA	"nvidia"
Netflix Inc.	NFLX	"netflix"
Walmart Inc.	WMT	"wal mart"
Wells Fargo & Company	WFC	"wells fargo"
salesforce.com inc.	CRM	"salesforce"
Boeing Company	BA	"boeing"

Company	Ticker	Keyword
McDonald's Corporation	MCD	"mcdonalds"
Citigroup Inc.	C	"citigroup"
Abbott Laboratories	ABT	"abbott laboratories"
Bristol-Myers Squibb Company	BMY	"bristol myers"
Medtronic Plc	MDT	"medtronic"
Costco Wholesale Corporation	COST	"costco"
PayPal Holdings Inc	PYPL	"paypal"
Philip Morris International Inc.	PM	"philip morris"
AbbVie Inc.	ABBV	"abbvie"
NextEra Energy Inc.	NEE	"nextera energy"
Amgen Inc.	AMGN	"amgen"
Thermo Fisher Scientific Inc.	TMO	"thermo fisher"
Accenture Plc Class A	ACN	"accenture"
International Business Machines Corporation	IBM	"ibm"
Honeywell International Inc.	HON	"honeywell"
NIKE Inc. Class B	NKE	"nike"