Google Search Volume and Stock Market Liquidity

* Sumit Kumar Jha
** Sharad Nath Bhattacharya
*** Mousumi Bhattacharya

Abstract

Our research showed that search volume on Google serves or GSV as an intuitive proxy for overall stock market recognition. We proposed a predictive parsimonious model TFARM (two factor auto - regressive methodology) on stock market liquidity measures (bid - ask spread, market efficiency coefficient, trading probability, turnover ratio (TR), and total volume (TV)) and employed public and free information such as GSV (Google search volume) on a dataset from NSE (National Stock Exchange) for period between 2004 - 2016 divided into pre, during, and post subprime crisis of 2007-2008. We found that an increase in Google search queries was linked to a rise in stock liquidity and trading activity. We characterized the improved liquidity to a decrease in asymmetric information costs and thus, concluded that GSV mainly measured attention from uninformed investors. Moreover, we found evidence that an increase in search volume was associated with temporarily higher future returns, which reinforced the previous findings. Impact of GSV on both TV and TR in terms of direction was similar in nature and consistent with the findings of Preis, Moat, and Stanley (2013).

Keywords : **Google Insights, GSV, stock liquidity, trading activity, stock returns**

JEL Codes : C13, G12, G14, G17

Paper Submission Date : August 10, 2018 ; Paper sent back for Revision : July 16, 2019 ; Paper Acceptance Date : July 29, 2019

Predicting conducts of the stock market has been the "holy grail" of finance. Many researchers (Ang & Bekaert, 2007 ; Campbell & Yogo, 2006) investigated it in the backdrop of EMP - the efficient market hypothesis. The vast literature of recent work is devoted to effecting of investor sentiments, and they utilized available data on Wikipedia (Preis, Moat, & Stanley, 2013) and Twitter (Bollen, Mao, & Zeng, 2011). Recently, utilization of Google trends monthly data has seen significant traction in the research community (Challet & Ayed, 2013 ; Preis, Reith, & Stanley, 2010 ; Preis et al., 2013). While the above - mentioned studies along with the majority of others focused on stock return, our objective is to analyze the impact of public information available freely in general, and Google trends or Google search volume (GSV) in particular, on the liquidity of the Indian stock market. Various studies based on different dimensions[1] of market liquidity showed the

[1] Kindly refer to Table 1 for brief discussion on various dimensions of market liquidity.

* *Doctoral Scholar,* Indian Institute of Management Shillong, Shillong - 793 014, Meghalaya.
(Email : sumitsony@gmail.com) ; ORCID : https://orcid.org/0000-0002-5930-3095
** *Assistant Professor,* Indian Institute of Management Shillong - 793 014, Meghalaya.
(Email : sharadbhattacharya@gmail.com) ; ORCID: https://orcid.org/0000-0002-5568-234X
*** *Assistant Professor,* Indian Institute of Management Shillong, Shillong - 793 014, Meghalaya.
(Email : mbhattacharya9@gmail.com)

DOI : 10.17010/ijf/2019/v13i8/146304

benefits of the presence of liquidity in the financial market (Chordia, Roll, & Subrahmanyam, 2008). The most notable episode is when market liquidity vanishes. The global financial crisis of 2008 has shown researchers that it is liquidity resilience which breaks the financial market system (Duarte & Eisenbach, 2018). In the flash crash of 2010 (Kirilenko, Kyle, Samadi, & Tuzun, 2017) and the flash rally of 2014 (Bouveret, Breuer, Chen, Jones, & Sasaki, 2015), market liquidity looked to vanish for 36 minutes and 12 minutes, respectively. These events raise suspicion about a fundamental flaw in the microstructure of the market. However, such brief periods of vanishing liquidity do not threaten the fundamentals of the financial system as a whole.

In extant literature, very few attempts have been made to study stock markets (market microstructure in general and liquidity in particular) based on Google search volume or trends data. Studies have investigated the correlation between Google search volume and returns, but they found that Google search data can be used to predict trading volume - one of the widely used indirect measures of stock market liquidity. Specific term search volume on Google not only serves as an intuitive proxy for overall recognition at the exchange level, but it also captures the attention of stock market investors. Research work about the stock market microstructure has overlooked the fact that Google search volume could primarily measure attention from uninformed investors and consequently, affect the liquidity in the stock market (Bank, Larch, & Peter, 2011). Our work attempts to address the above-mentioned research gap in the unique setting of India's stock market, which is order - driven ; whereas, developed countries mostly have quote - driven stock markets. Therefore, a predictive model for stock market liquidity based on public and free information such as GSV (Google search volume) could be immensely useful for investors in the Indian stock market.

Data and Methodology

Our study focuses on a major stock exchange of India – National Stock Exchange (NSE) and considers composite NSE500 - a well-diversified index consisting of companies of different market capitalization and categories - for the period from August 2004 to February 2016. Our data covers pre and post subprime crisis period, which is apt for studying the impact on market liquidity during that period caused by some exogenous factors like Google search volume (GSV) as in our case.

In this paper, to gauge the robustness of the effect of GSV on multiple dimensions of liquidity, we consider popular measures : natural log of bid-ask spread (Spread), turnover rate (TR), and natural log of trading value (TV) along with the current prevalent trading probability (TP), market efficiency coefficient (MEC) as liquidity measures. We consider the trading probability (TP) as an additional measure of liquidity, which is calculated as probability equals 1/ (1 + the number of non-trading days in a month) following Narayan and Zheng (2011). Along with the ease of trade, this measure also captures the speed dimension of liquidity and avoids the bias effects from the noise in the market as a noisy market has more risks of serial correlation effects. We also consider MEC that measures the impact of execution costs on price volatility over short horizons and compare the long term variance with the short-term variance. MEC is calculated as MEC = Long Term Variance/(T × Short Term Variance) where T be the number of sub - periods into which a longer period can be divided. We considered five days as short period and 30 days long period, that is, T = 6. When MEC is less than but closer to 1, it suggests that the market is resilient and minimum price volatility is expected.

The Table 1 describes the central tendency of all liquidity measures for the three periods separately. Except for TP, all other measures appear to be increased during the crisis on an average. We also observe that MEC and TR show the highest skewness and kurtosis post the global financial crisis.

The Figure 1 visualizes natural log of NSE 500, monthly return from NSE 500, and the natural log of moving average. Moving average graph shows flatness for approximately 2 years between 2008 and 2010.

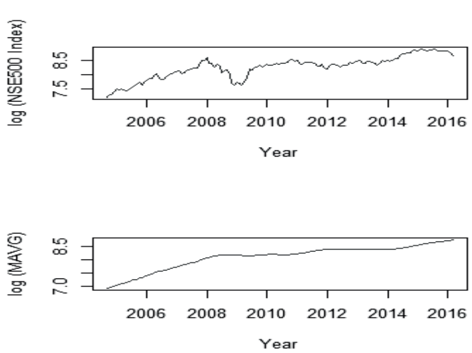
The Figure 2 shows the monthly google search volume (standardized) for the term “NSE”. Internet-savvy

Table 1. Descriptive Statistics for All Liquidity Measures

	Period	Mean	*SD*	Median	Min	Max	Range	Skewness	Kurtosis
MEC	1	0.48	0.41	0.35	0.11	1.72	1.61	1.63	1.95
	2	0.51	0.42	0.33	0.09	1.75	1.67	1.46	1.35
	3	0.62	0.6	0.36	0.11	3.42	3.31	2.19	5.37
Spread	1	5.24	0.55	5.13	4.23	6.66	2.43	0.38	0.19
	2	6.19	0.51	6.17	5.36	7.51	2.14	0.56	0.25
	3	5.86	0.44	5.9	4.65	6.81	2.16	-0.24	-0.16
TP	1	0.1	0.01	0.1	0.08	0.12	0.05	0.33	-0.47
	2	0.09	0.01	0.09	0.07	0.11	0.04	-0.06	-1
	3	0.09	0.01	0.09	0.07	0.11	0.04	0.06	-0.81
TV	1	22.16	0.37	22.2	21.22	22.63	1.41	-1.04	0.37
	2	22.7	0.29	22.71	22.2	23.23	1.03	-0.17	-1.02
	3	23.24	0.22	23.22	22.85	23.88	1.03	0.61	0.2
TR	1	0.05	0.01	0.05	0.02	0.07	0.05	-0.56	-0.52
	2	0.06	0.01	0.05	0.04	0.08	0.04	0.51	-1.29
	3	0.04	0.01	0.04	0.03	0.11	0.08	2.15	5.05

Note. The values are shown in three periods comprising of pre (1), during (2), and post (3) global financial crisis.

Figure 1. Evolution of NSE 500 Index, Returns, and MAVG from August 2004 to February 2016



investors seem to be particularly interested in the Indian stock market. One reason could be the initial forecast about the effects of the global financial crisis on the Indian economy was less severe (Ghosh & Chandrasekhar, 2009 ; Rakshit, 2009).

The Figure 3 depicts the development of liquidity measures for the period of study in the Indian stock market. MEC shows a volatile and non-stationary (by variance) pattern. It takes the maximum range in recent time. Spread stays within a range after the global crisis. TP is not a measure concern in the Indian market, which remains well above 7%. Like spread or bid-ask spread, TV does not show an increasing trend. In the same line, TR after 2010 appears steady.

Figure 2. Google Search Volume for NSE for the Period of August 2004 - February 2016

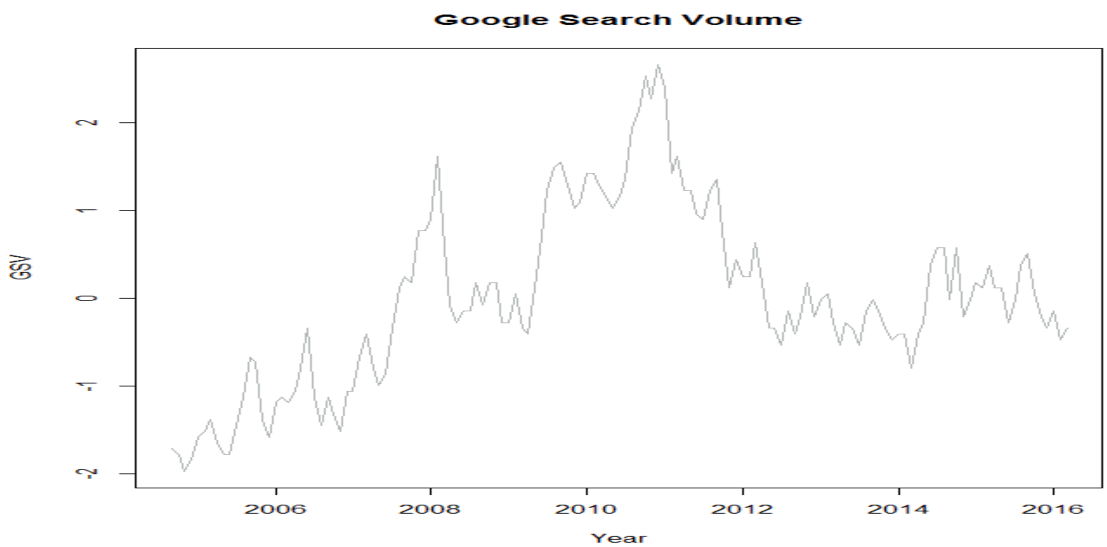
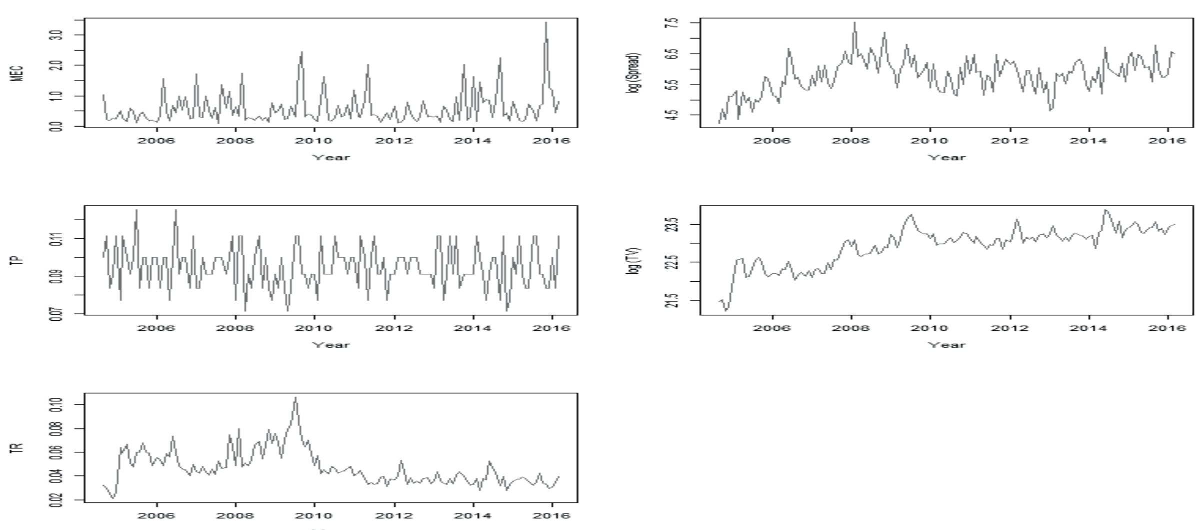


Figure 3. Evolution of Market Liquidity Dynamics for NSE 500 from August 2004 to February 2016



➪ **Two Factor Auto-Regressive Model (TFARM) :** Our analysis follows Tetlock (2007) and Bijl, Kringhaug, Molnár, and Sandvik (2016). The independent variables in the model have three lags (ACF and PACF shown in Figure A1, Appendix A and Figure A2, Appendix A). The explanatory variables in the model are three lags each of GSV (standardized) and individual liquidity measures.

Therefore, our TFARMA model can be specified as :

$$Liq_t = \alpha + (\textstyle\sum_{i=1}^{3}\beta_i L^i)\ Liq_t + \sum_{i=1}^{3}\Upsilon_i L^i) GSV_t + \varepsilon_t \qquad \text{....................(1)}$$

The solution to the equation (1) can associate GSV with a rise in trading activity and stock liquidity. Such knowledge can help investors - especially those with a lack of private information - in formulating their trading strategy.

Results

In this section, we discuss the TFARMA model estimation of equation (1). We present only those results that are significant. Any insignificant coefficient estimates or results related to post-estimation diagnostics are reported in Appendix A and Appendix B.

From the Table 2, we note that TFARM is not suitable for predicting MEC with GSV. Therefore, we will not make any comment on MEC. We also observe that even though the *F*-stat is significant for Spread and TP, because of low adjusted *R*-squared values of 10% and 30% (shown in Table B1 and Table B2 of Appendix B), we will not make any comment on predicting the bid-ask spread and TP using GSV. Therefore, we begin to examine the impact of GSV on liquidity.

From Table 3, we note that GSV shows significant impact on TV. Current GSV has a positive effect, and 1 lag GSV has a negative impact on market liquidity. A 100 b.p.s positive change in current month GSV could increase TV more than 32 b.p.s. However, a change of 100 b.p.s in GSV of more than one-month-old could decrease TV by

Table 2. TFARM Evaluation

Model	*F*- Statistic	*p*-value	Significant
MEC	1.239	0.2862	
Spread	9.438	2.08E-09	***
TP	2.924	7.15E-03	***
TV	163.8	2.20E-16	***
TR	48.41	2.20E-16	***

Significance. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 3. *TV* Model (Adjusted *R* - Squared : 0.8941)

Coefficients	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	2.35105	0.818352	2.873	0.00476**
Google Search Volume (Standardized over each month)	0.328109	0.0391	8.392	7.63E-14***
Google Search Volume (Standardized monthy with lag 1)	-0.359029	0.060814	-5.904	3.00E-08***
Google Search Volume (Standardized monthy with lag 2)	0.020788	0.068779	0.302	0.76296
Google Search Volume (Standardized monthy with lag 3)	0.024285	0.049095	0.495	0.6217
Trading volume (lag 1)	0.875548	0.0865	10.122	2.00E-16***
Trading volume (lag 2)	0.02904	0.115477	0.251	0.80185
Trading volume (lag 3)	-0.006502	0.085616	-0.076	0.93958

Significance. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 4. *TR* Model (Adjusted *R* - Squared : 0.7332)

Coefficients	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	0.007543	0.002409	3.131	0.002156**
Google Search Volume (Standardized over each month)	0.0129	0.001928	6.691	6.25E-10***
Google Search Volume (Standardized monthy with lag 1)	-0.010274	0.002987	-3.439	0.000787***
Google Search Volume (Standardized monthy with lag 2)	-0.001219	0.003143	-0.388	0.69883
Google Search Volume (Standardized monthy with lag 3)	-0.002164	0.002266	-0.955	0.341364
Turnover Rate (lag 1)	0.572305	0.087991	6.504	1.60E-09***
Turnover Rate (lag 2)	0.141997	0.100933	1.407	0.161894
Turnover Rate (lag 3)	0.124323	0.08503	1.462	0.14616

Significance. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 5. Predictability of TFARM

Model	*N*-Step Ahead	Model	RMSE
TV	5	2 Factor AR	9.21%
TR	5	2 Factor AR	0.005%

Table 6. *TV* Model Results for Pre, During, and Post Global Crisis Period

TV Model	Pre-Crisis	During-Crisis	Post-Crisis
(Intercept)	10.57041**	3.06131	1.90086***
Google Search Volume (Standardized over each month)	0.28211*	0.23027*	0.31663***
Google Search Volume (Standardized monthly with lag 1)	-0.15199	-0.26773·	.-0.32013***
Google Search Volume (Standardized monthly with lag 2)	-0.15713	-0.02708	0.01764
Google Search Volume (Standardized monthly with lag 3)	0.07728	0.08766	-0.04658
Trading volume (lag 1)	0.83902***	0.66088*	0.61818***
Trading volume (lag 2)	-0.16875	0.36077	0.04641
Trading volume (lag 3)	-0.14282	-0.15462	0.0393

Significance. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

as much as 35 b.p.s. The predictive power of GSV in case of TV is a higher autoregressive term (lag more than 1).

From Table 4, we note that Google search volume (GSV) shows significant impact on TR. Current GSV has a positive effect, and 1 lag GSV has a negative impact on market liquidity. A 100 b.p.s positive change in current month GSV could increase TR more than 1 b.p.s. However, a change of 100 b.p.s in GSV of more than one-month-

Table 7. ***TR*** **Model Results for Pre, During, and Post Global Crisis Period**

TR Model	Pre-Crisis	During-Crisis	Post-Crisis
(Intercept)	0.03133	0.01227	6.96E-03**
Google Search Volume (Standardized over each month)	0.01474·	1.87E-02**	1.13E-02***
Google Search Volume (Standardized monthly with lag 1)	-0.0009	-0.019528	-9.31E-03**
Google Search Volume (Standardized monthly with lag 2)	-0.0047	0.006721	4.24E-05
Google Search Volume (Standardized monthly with lag 3)	-0.0045	-0.001365	-2.65E-03
Turnover Rate (lag 1)	0.40306·	4.63E-01	6.31E-01***
Turnover Rate (lag 2)	0.32423	0.032532	1.87E-03
Turnover Rate (lag 3)	-0.2142	0.303689	2.02E-01·

old could decrease TR by as much as 1 b.p.s. Impact of GSV on both TV and TR is similar and consistent with the findings of Preis et al. (2013).

We also analyze 5 months ahead prediction of liquidity measures based on their respective models, and Table 5 shows that TFARM on TR appears to result in the least RMSE.

From Table 6 and Table 7, we note that GSV predictability decreased during the financial crisis for both TV and TR and increased for TV in the post-crisis period and decreased for TR.

Overall, GSV seems to be predicting TV and TR, but it affects TV more significantly in magnitude as well as in direction.

Robustness : The robustness of our results can be visualized through the diagnostic plots (as shown in Figure A3 and Figure A4 of Appendix A). We check the square of residuals (presented in Figure A5, Appendix A) for stationary and Q - Q plot for normality. Our Durbin - Watson test (reported in Table B3, Appendix B) could not reject the null hypothesis of non-auto correlated errors. CERES plots (residual histograms shown in Figure A6 Appendix A ; component and residual plots shown in Figure A7, Figure A8, Figure A9 Appendix A) take care of possible non-linearity for GSV, which appears to be not present.

Conclusion

In this paper, we attempt to propose a parsimonious two - factor autoregressive model to predict liquidity measures based on monthly Google search volume data available freely in the public domain. Equation (1) specifies our model. We worked on monthly data of five liquidity measures of NSE 500 for a period August 2004 - February 2016. Our model is not suitable for the prediction of MEC, bid - ask spread, and TP. However, we provide empirical evidence to show GSV's explanatory and predictability power for TV in both directions as well as magnitude. A 1% increase in current month GSV could increase TV by more than 0.32%. However, an increase of 1% in GSV of more than one-month-old could decrease TV by 0.35%. Impact of GSV on both TV and TR in terms of direction is similar and consistent with the findings of Preis et al. (2013). Though, compared to TV, GSV is weak in explaining TR. We also analyze the next five months prediction of liquidity measures based on their respective models, and our model with TR data appears to result in the least RMSE. Overall, we conclude that

Google search volume primarily captures the attention of uninformed investors resulting in reduced information asymmetry, improved liquidity, and short - term buying pressure.

Research Implications, Limitations of the Study, and Scope for Further Research

Our results that associate GSV with a rise in trading activity and stock liquidity have implications for investors - especially those who lack private information in formulating their trading strategy. Since our work extends the studies of both Tetlock (2007) and Bijl et al. (2016) in terms of methodology and scope, therefore, an empirical validation of our model based on panel data of all listed companies in the Indian stock exchange where the impact of GSV on market liquidity, incorporating the fixed or random effects, is left for future research.

References

Ang, A., & Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies, 20*(3), 651- 707.

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*(3), 239 - 264.

Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis, 45*, 150 - 156.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1 - 8.

Bouveret, A., Breuer, M. P., Chen, M. Y., Jones, D., & Sasaki, T. (2015). *Fragilities in the U.S. Treasury Market: Lessons from the "Flash Rally" of October 15, 2014*. Retrieved from https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Fragilities-in-the-U-S-43338

Campbell, J. Y., & Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics, 81*(1), 27 - 60.

Challet, D., & Ayed, A. B. H. (2013). *Predicting financial markets with Google Trends and not so random keywords.* Retrieved from https://arxiv.org/abs/1307.4643

Chordia, T., Roll, R., & Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics, 87*(2), 249 - 268.

Duarte, F., & Eisenbach, T. M. (2018). *Fire - sale spillovers and systemic risk*. DOI : https://dx.doi.org/10.2139/ssrn.2340669

Ghosh, J., & Chandrasekhar, C. P. (2009). The costs of 'coupling': The global crisis and the Indian economy. *Cambridge Journal of Economics, 33*(4), 725 - 739.

Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash : High frequency trading in an electronic market. *The Journal of Finance, 72*(3), 967 - 998.

Narayan, P. K., & Zheng, X. (2011). The relationship between liquidity and returns on the Chinese stock market. *Journal of Asian Economics, 22*(3), 259 - 266.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports, 3*, Article Number : 1684.

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.

Rakshit, M. (2009). India amidst the global crisis. *Economic and Political Weekly, 44*(13), 94 - 106.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance, 62*(3), 1139 - 1168.

Appendix A

Figure A1. ACF for All Five Liquidity Measures

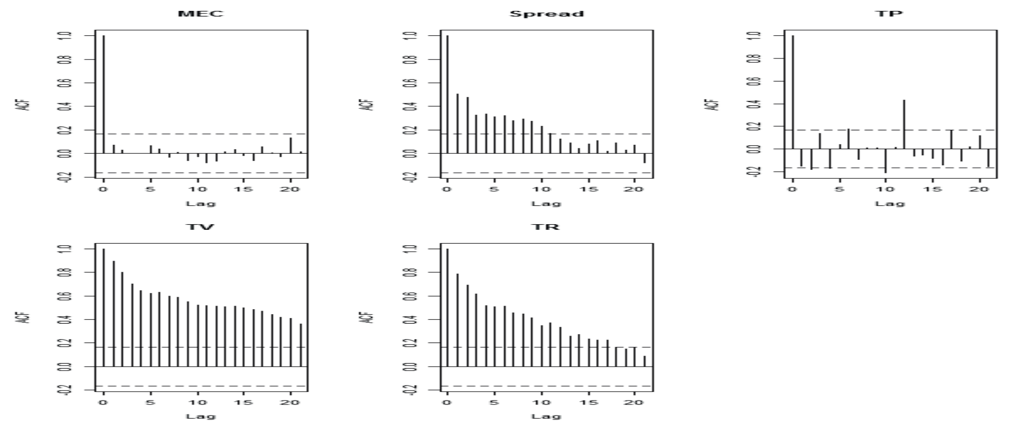


Figure A2. PACF for All Five Liquidity Measures

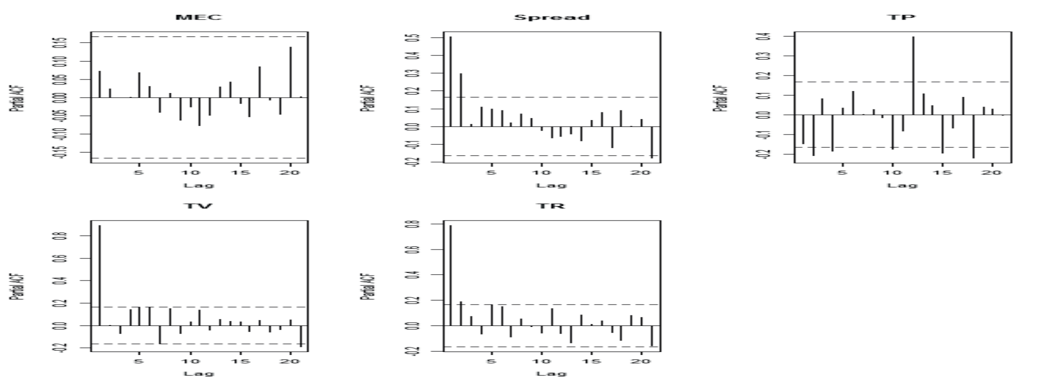


Figure A3. Diagnostic Plots for the Model on *TV*

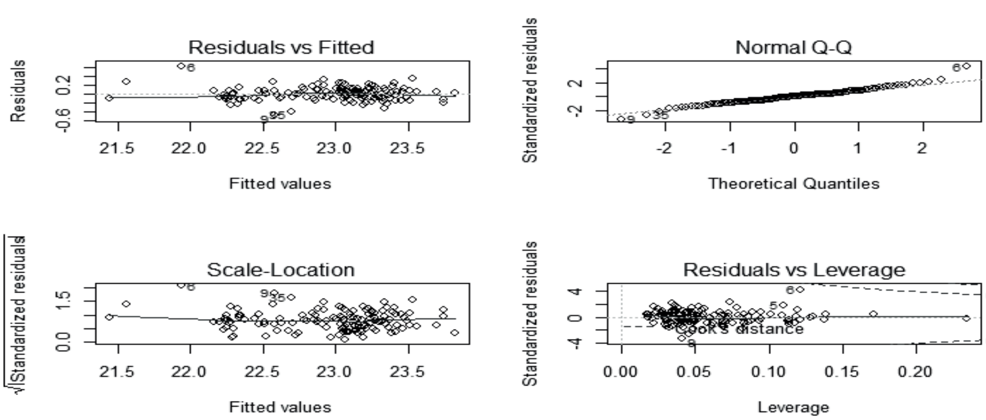


Figure A4. Diagnostic Plots for the Model on *TR*

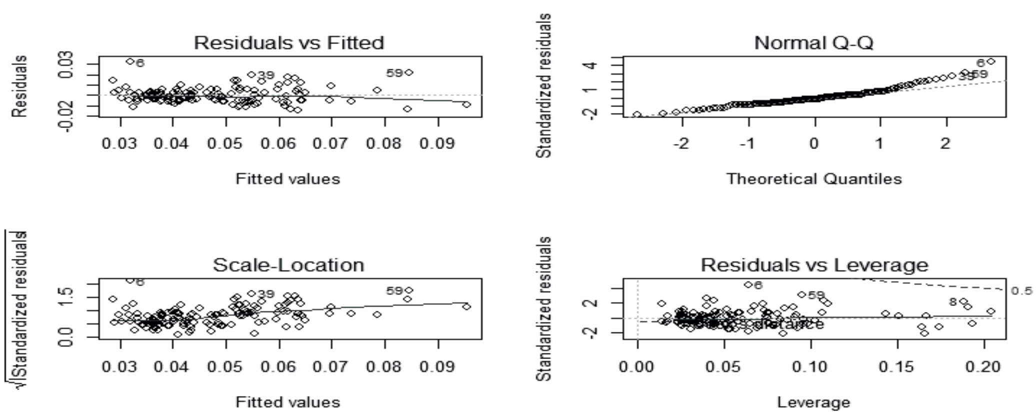


Figure A5. Residual Squared Plots

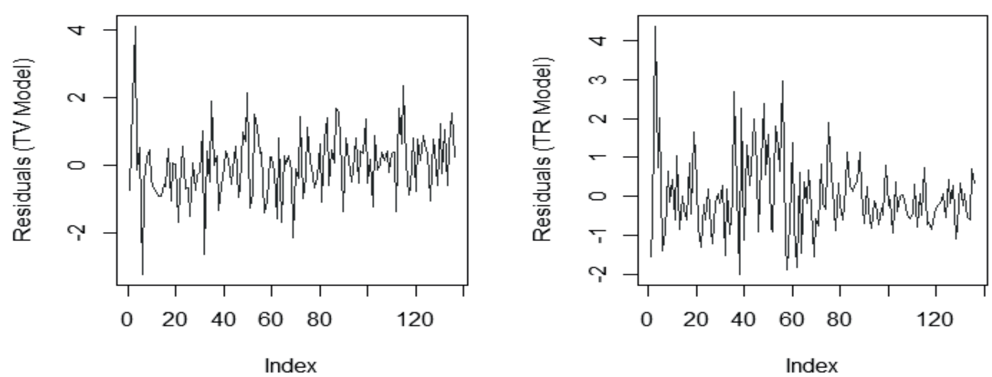


Figure A6. Residuals Histograms

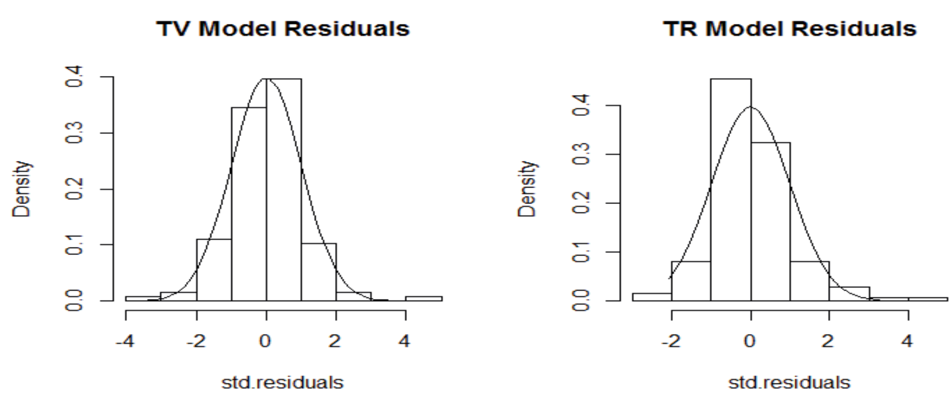


Figure A7. Spread Plots for TFAR Models on *TV* and *TR*

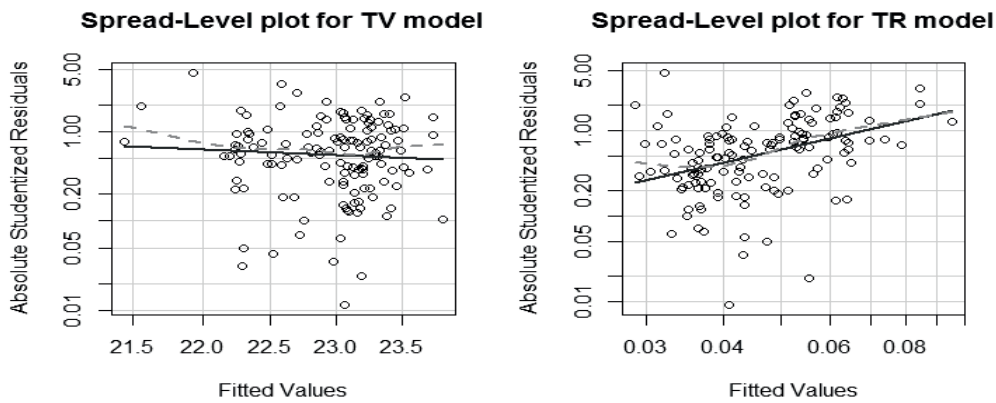


Figure A8. Component and Residual Plot for *TV*

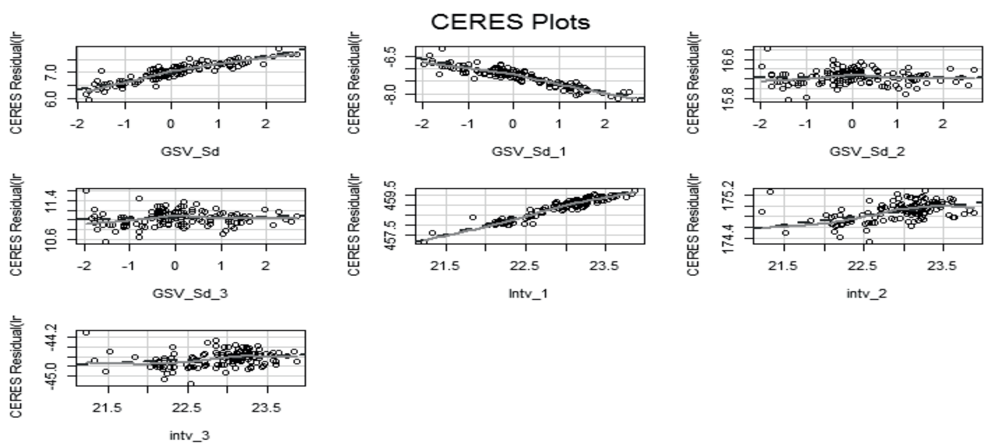
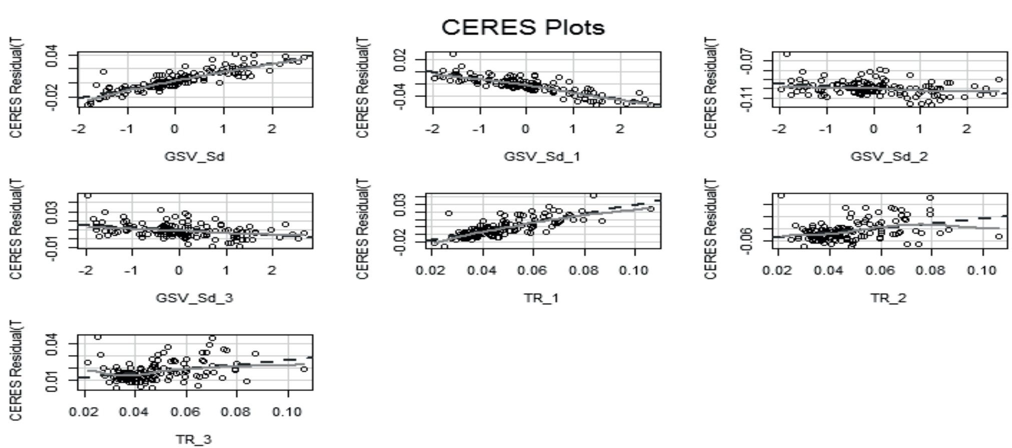


Figure A9. Component and Residual Plot for *TR*



Appendix B

Table B1. Spread Model (Adjusted *R* - Squared: 0.3044)

Coefficients	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	2.33691	0.56215	4.157	5.86E-05***
Google Search Volume (Standardized over each month)	0.15296	0.11489	1.331	0.18544
Google Search Volume (Standardized monthly with lag 1)	0.01504	0.15992	0.094	0.9252
Google Search Volume (Standardized monthly with lag 2)	-0.0291	0.16048	-0.181	0.85637
Google Search Volume (Standardized monthly with lag 3)	-0.0658	0.1164	-0.565	0.57287
Bid-Ask spread (lag 1)	0.28248	0.08872	3.184	0.00182**
Bid-Ask spread (lag 2)	0.28318	0.08874	3.191	0.00178**
Bid-Ask spread (lag 3)	0.03412	0.08779	0.389	0.69814

Significance. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table B2. *TP* Model (Adjusted *R*-squared: 0.09073)

Coefficients	Estimate	Std. Error	*t* - value	Pr(>|*t*|)
(Intercept)	0.1135827	0.016686	6.807	3.46E-10***
Google Search Volume (Standardized over each month)	0.0076705	0.0029715	2.581	0.011*
Google Search Volume (Standardized monthly with lag 1)	-0.004185	0.0044159	-0.948	0.3451
Google Search Volume (Standardized monthly with lag 2)	-0.0006154	0.0043854	-0.14	0.8886
Google Search Volume (Standardized monthly with lag 3)	-0.0028001	0.0029917	-0.936	0.3511
Trading probability (lag 1)	-0.1218381	0.0931913	-1.307	0.1934
Trading probability (lag 2)	-0.1983481	0.0904254	-2.194	0.0301*
Trading probability (lag 3)	0.1180402	0.0896094	1.317	0.1901

Significance. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table B3. D - W Statistics

Model	Lag	Autocorrelation	D-W Statistics	*p* - value
TV	1	-0.01457804	2.024528	0.98
TR	1	-0.01073995	2.002395	0.87

Significance. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

About the Authors

Sumit Kumar Jha is a Doctoral Fellow at IIM Shillong. With a background of software engineering, his research interest is fintech and allied topics in the Indian financial market. He specializes in time varying nature of stock market liquidity.

Sharad Nath Bhattacharya is a Ph.D. in finance from University of Calcutta, India and is currently an Assistant Professor at Indian Institute of Management Shillong. His research interest is in exploring the time-varying dynamics of financial markets.

Mousumi Bhattacharya is a Ph.D. in finance from the University of Calcutta, currently working as an Assistant Professor at Indian Institute of Management Shillong, and her areas of interest are financial services and FDI.