News Impact Score and Stock Market Quality: Information Content in Business News

The weak form of market efficiency states that stock prices reflect all available past information about a company and hence the only occasion where such prices move is when any new price-sensitive information arrives in the market. Finance literature documents that any new information is indirectly captured in the order book in terms of movements in net order flows. In the past decade and half, researchers and analysts have started looking at a direct measure of new information and observed that business news from reliable sources can be used as a direct measure of new information. Use of news from business dailies has several advantages: (a) structured language with minimum grammatical errors; (b) written by experts (financial reporters and analysts) who have reasonable knowledge of finance and economics; and (c) mostly free from biases as a newspaper has to disclose conflict of interests in case there is any news of such kind. Additionally, due to technological advancements, newspapers can be crawled on real time basis and thus any price-sensitive information can be captured almost immediately.

While it has been empirically established that news carry vital information, it is also demonstrated that quality of news (called news sentiments) matter more than mere volume of news. We have developed a measure of news sentiment, called News Impact Score, which attempts to capture the tonal quality and thus the sentiments of a news. The News Impact Score (NIS) evaluates sentiments of any news article at sentence level and then aggregates such sentence-level scores into document (article) level using a decay function. In other words, sentiments of the initial sentences in any news piece carry more weights than sentiments of later sentences in the same piece. Sentence-level sentiment scores are generated using machine learning techniques, such as support vector machine algorithm. Therefore, NIS for any particular news piece represents quality of information contained in the article. Such information could be in terms of business growth, higher profitability which have positive impact or financial distress, deceleration of growth, losses, factory lockdown which have negative impact. The dominant objective of NIS is to capture corporate vulnerability- financial and otherwise. Hence, a higher (positive) NIS indicates greater vulnerability and a lower (negative) NIS indicates good news.

NIS about a company is captured at three different frequencies- daily, weekly, and monthly. The first challenge in identifying appropriate corporate-level news articles is to check relevance of a news- whether a news really 'belongs' to a specific company. We use a proprietary algorithm to identify news relevance. Higher the relevance score, greater is the certainty that the news actually talks about a particular company. Since news is crawled from multiple sources, it is also important to identify novelty (i.e., freshness) of a news. A repeat news carries weak signal. Novelty can be captured using news topic and time stamp. One can filter any news article based on novelty and thereby look at NIS of only breaking news. Thus, one cans use 'relevance' and 'novelty' filters to fine tune search for NIS.

We use monthly-level NIS to examine (a) whether it explains stock returns; and (b) whether it impacts volatility of stock returns. We also check the spill over effects of past news on stock return and volatility- two important indicators of stock market quality. We find that NIS affects stock returns negatively- higher the NIS, lesser the stock returns. We also find that NIS affects volatility of stock returns positively. Since NIS measures corporate vulnerability, it is expected that higher value of NIS will send negative signals to the market about a particular stock resulting in greater volatility. It is observed that volatility in equity markets is asymmetric[1]. In other words, equity market volatility is observed to be higher in declining markets than in rising markets. Therefore, during periods of higher corporate stress (high values of NIS), stock prices should fall and volatility rise.

News Frequency and NIS: Basic Features

We look at firm-level news in this paper and hence ignore any news on economy in general. Since news is sourced from business dailies (e.g Economic Times, Mint, Money Control etc.), NIS does not capture sentiments of any general, value-neutral news. We have looked at top 500 NS-listed companies (NSE 500 Index) and analysed the NIS data for these companies over 11 years (2010 to 2020). We have removed companies for which stock price data were not

[1] Geert Bekaert, and Guojun Wu. Asymmetric Volatility and Risk in Financial Markets. The Review of Financial Studies, Spring 2000. Vol 13. No. 1, pp 1-42

available, leaving us with 399 of the Nifty 500 companies. These companies are sorted on the basis of aggregate news frequency in a particular year and the top decile represents top ten percent companies which are heavily reported in the media in that year. We observe (Table 1) that news frequency has increased over the years for almost all the companies. Looking at the top decile, it can be seen that the news frequency for the year 2020 is approximately 3.85 times of the number for the year 2010.

Table 1: Aggregate news frequency (deciles) year-wise from 2010-2020

	YEAR										
AGGREGATE NEWS FREQUENCY (DECILE)	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
	435	522	803	621	624	660	796	830	929	772	1677
	175	225	292	205	206	238	258	254	202	216	458
	102	138	178	133	120	143	169	162	126	124	253
	78	95	113	82	78	113	117	110	80	81	168
	43	78	78	54	44	78	89	78	50	55	103
	39	45	55	39	39	49	65	44	39	39	78
	39	39	39	33	36	39	39	39	35	33	40
	6	39	39	0	0	30	31	16	0	0	39
	0	5	9	0	0	0	0	0	0	0	11
	0	0	0	0	0	0	0	0	0	0	0

One may find significant difference in the breadth of media coverage for the top Indian listed companies. Though the NIFTY 500 Index represents more than 90% of the free float market capitalization of the stocks listed on NSE, each stock in index does not generate enough newsworthy stories. Companies in the bottom two deciles (about 80 out of 399 companies) do

not feature in the media at all. On the other hand, the top two deciles contributed 75% of the news reports in 2019 and 2020. Therefore, we have done further analysis on the NIS scores of companies in the top two deciles- about 80 companies.

Return and Volatility

Investors in stock markets are concerned with three broad measures of market quality- asset prices, volatility, and liquidity. Stock returns (measured by percentage change in traded price over a period) and volatility of stock returns are two important parameters that every portfolio manager and traders closely follow. So, if stock returns and risks (i.e., volatility) can be modelled using news sentiments, that would greatly benefit long-term investors in shares and also the short-term traders in cash and options markets.

We focus, in this paper, on whether NIS can explain these two indicators of market quality- asset returns and historical volatility.

Data and Methodology

We have looked at the effect of NIS on stock returns and volatility. Since, NIS are used at monthly frequency, stock prices and hence returns are estimated at monthly rest. Stock prices for firms are collected from Yahoo Finance. Stock returns are estimated as (natural)log of price relatives. In other words, we have defined stock returns as monthly change in prices on a logarithmic scale. Historical volatility (standard deviation) of stock returns is estimated using past 30 months returns on a rolling basis.

News flow was much lower during 2010 through 2013. Also, a large number of companies had very low news volume even after 2013. We have only considered, for our analysis, those companies which had news for at least 80% of the months since July 2014. This filter has reduced our sample size to 86 companies out of about 400 companies. Thus, we have monthly data for the variables since July 2014 till December 2020.

Since our data consist of cross-section variables changing across time; panel regression is used[2]. With weakly balanced data (each panel contains the same number of observations but not the same time points); a panel is set, using the 'Company' (that takes value from 1 to 86 for 86 companies) as the cross-section variable and 'Month' (that runs from 1 to 6708 for 86 companies across 78 months) as the time variable.

To calculate the effect of NIS on Return and Volatility, we run two separate sets of panel regressions.

Panel A consists of two regressions where Return is the dependent variable.

- $Return_{it} = \alpha + \beta_1 NIS_{it} + \beta_2 NIS_{-1(it)} + \beta_3 Volatility_{-1(it)} + \varepsilon_{it}$..... [eq 1]
- $Return_{it} = \alpha + \beta_1 NIS_{it} + \beta_2 NIS_{-1(it)} + \varepsilon_{it}$..... [eq 2]

We have used Lag of Volatility as an explanatory variable instead of the lag of Return since Returns are calculated using stock prices which themselves are independent of their past values.

Panel B includes regression with Volatility as a dependent variable.

- $Volatility_{it} = \alpha + \beta_1 NIS_{it} + \beta_2 NIS_{-1(it)} + \beta_3 Volatility_{-1(it)} + \varepsilon_{it}$.....[eq 3]

Lag of Volatility is used as an independent variable since expected volatility is calculated using past volatility (hence, the high correlation between them) and also explains why there are periods of high and low volatilities.

Results and Analysis

We observe (Table 2) that returns are negatively correlated with NIS, lag of NIS, and lag of volatility whereas volatility is positively correlated with NIS, which is ideally correct since higher NIS indicates bad news and thus lower returns, but high volatility). Also, return and

[2] Panel analysis is a statistical method to analyse two-dimensional (cross-section and time) panel data. The data are usually collected over time and over the same individuals and then a regression is run over these two dimensions

volatility are positively correlated as higher risk demands greater rewards. There is a negative relationship between current and past returns suggesting negative autocorrelations in returns which is a well-established feature of stock returns.

Table 2: Correlation between Return, NIS, Lag of NIS, Lag of Return, Volatility, Lag of Volatility.

	Return	Volatility	NIS	Lag.Volatility	Lag.NIS	Lag.Return
Return	1					
Volatility	0.0045	1				
NIS	-0.1100	0.0332	1			
Lag Volatility	-0.0016	0.9895	0.0239	1		
Lag NIS	-0.0269	0.0477	0.4653	0.0414	1	
Lag Return	-0.2326	0.0042	-0.0635	-0.0082	-0.1140	1

There are two types of effects in a panel model: Fixed or Random effect. A model is a fixed effect model if the variables are constant across individuals, random otherwise. To check which effect is best suited for the given data, we have conducted the Hausman test. Results show that it is better to use random effect models.

Also, to correct for Autocorrelation or Heteroscedasticity, if any, the Feasible Generalized Least Square (FGLS) Panel model is used. But since the coefficients of FGLS and normal panel model are the same, we can conclude that there is no problem of Autocorrelation or Heteroscedasticity and the normal random effect model is used (Table 3).

Results show that NIS has a negative and significant effect on stock returns. Lag NIS has no effect. This is due to the fact that news has a latency effect. News which are more than one month old fail to influence present stock prices- a sign of market efficiency. Past volatility has

no effect on present returns. Thus, NIS has a better explanatory power for returns that historical volatility of stocks.

Table 3: Effect of NIS on Stock Returns

Return	Coefficient	P>|z|
Random Effect Panel Regression of Return on NIS and lag of NIS		
NIS	-0.0001156	0.000
Lag NIS	0.0000205	0.388
Constant	.0074776	0.017
Random Effect Panel Regression of Return on NIS, lag of NIS, and lag of Volatility.		
NIS	-0.0001582	0.000
Lag NIS	0.0000389	0.090
Lag Volatility	0.0001521	0.995
Constant	0.006722	0.166
Wald chi2(2) =25.71; Prob>chi2=0.0000. Panel A(i) Wald chi2(2) =49.24;Prob>chi2=0.0000. Panel A(ii)		

We now turn to the efficacy of NIS in explaining stock volatility (Table 4). The coefficient of NIS is positive and significant implying that an increase in NIS leads to an increase in volatility while the lag of NIS is insignificant. Lag volatility is used as an explanatory variable to control for autocorrelations, if any. Even after correcting for autocorrelations, NIS has a significant and positive effect on volatility of returns.

Table 4: Effect of NIS on Volatility

Panel B		
Volatility	Coefficient	P>|z|
NIS	0.00000699	0.004
Lag NIS	0.00000211	0.382
Lag Volatility	0.97499	0.000
Constant	0.00282	0.000
Wald chi2(2) =144649.2;Prob>chi2=0.0000.		

Results in table 4 show that a change in volatility is partially explained by past volatility and present news sentiments. One may wonder why past NIS (lag NIS) did have no effect on present volatility. One possible explanation is that the variable on past volatility (lag volatility) has already captured the effect of past information (lag NIS) on volatility and hence lag NIS has no residual power to explain variations in returns volatility.

The above results fairly establish that it is not the quantity, rather quality of news that carry price sensitive information. Therefore, if one can carefully capture the news sentiments, one may be able to use that information to explain stock market behaviour and devise any long-term investment strategy using news impact score (NIS) services of Textplor.

NIS-Returns-Volatility: Examples from a select Companies

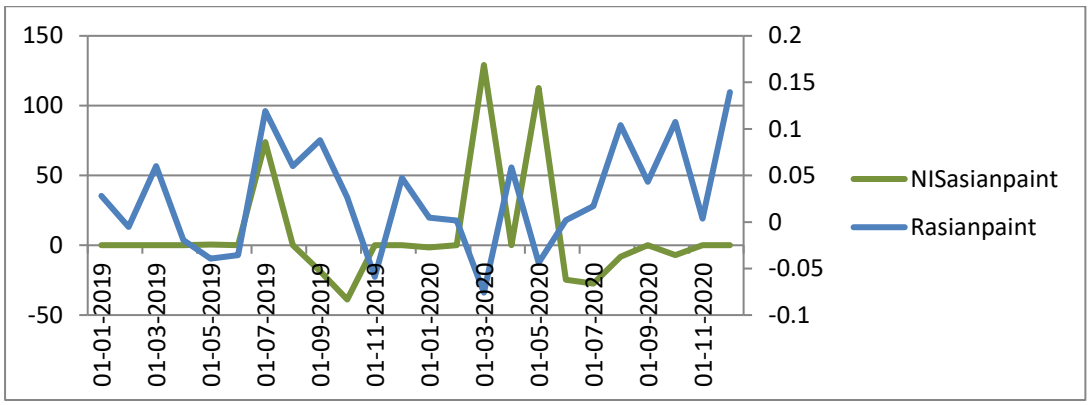
The above analysis shows the aggregate relationship between NIS, return, and volatilities. But data on individual companies also show a similar pattern (see Graphs in the Annexure). Five companies are selected for the analysis based on news frequency for the past two years 2019 and 2020. It is observed that there is a strong negative correlation between NIS and stock returns. The relationship is relatively weaker between volatility and NIS. Yet, we observe that higher NIS is followed by greater volatility of stock returns.

The news impact scores of companies capture price-sensitive information that may help portfolio managers, lenders, and even rating agencies in making economic decisions. NIS is

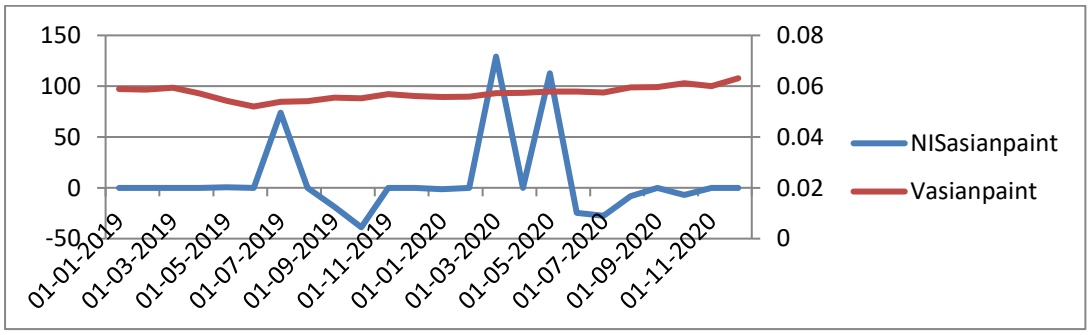
drawn from a news corpus that focus on fundamental performance of any firm, and not on stock-related news. Since, share price and their volatilities are in the long-run influenced by fundamental factors, mutual funds which hold assets for a longer tenure will be able to use NIS for identifying stocks to buy/sell. Similarly, NIS also provides insights to the lenders on possibility of changes in credit quality of borrowers. The only limitation of NIS is if the newspapers do not carry news about a company, NIS will be of no help.

APPENDIX

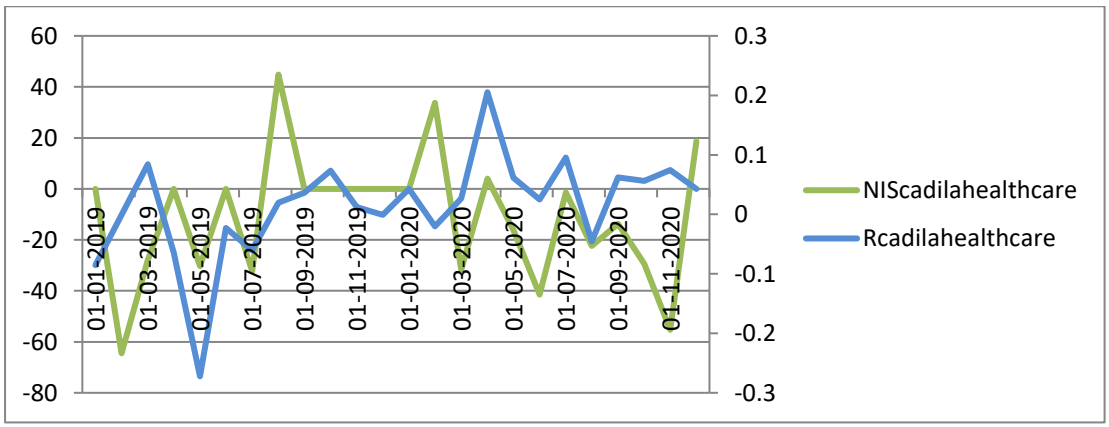
Asian Paints: NIS and Stock Returns



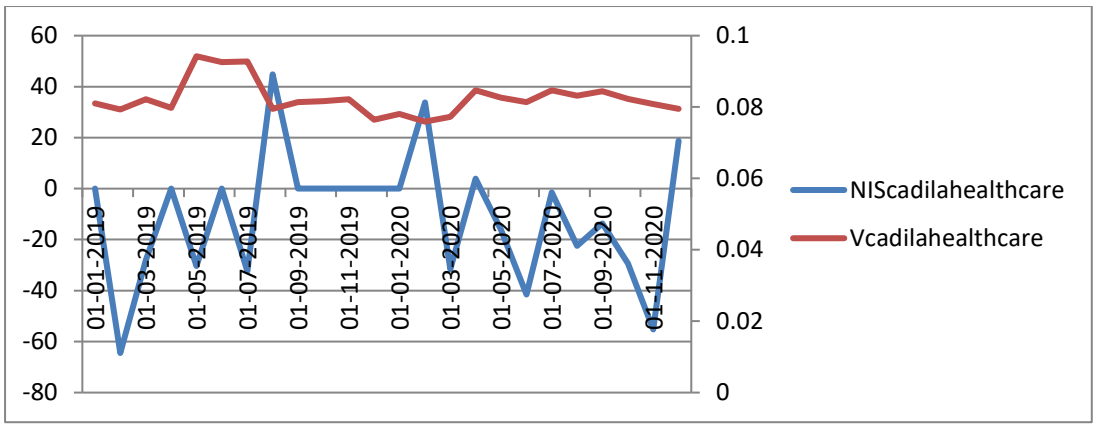
Asian Paints: NIS and Volatility of Stock returns



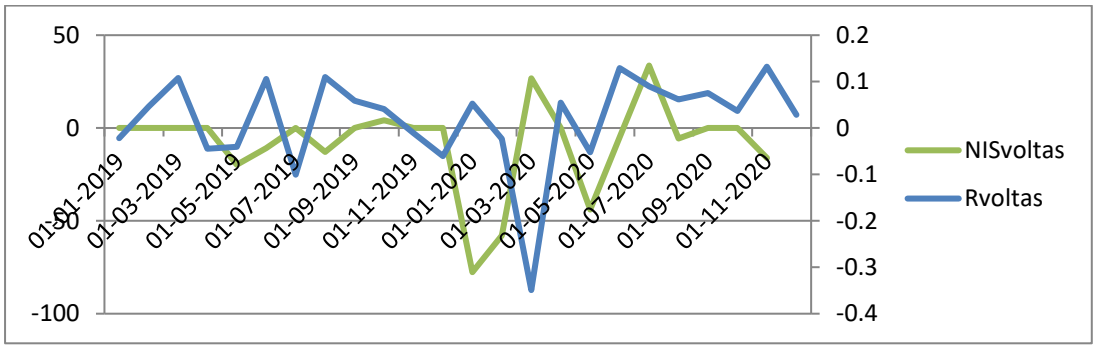
Cadila Healthcare: NIS and Stock Returns



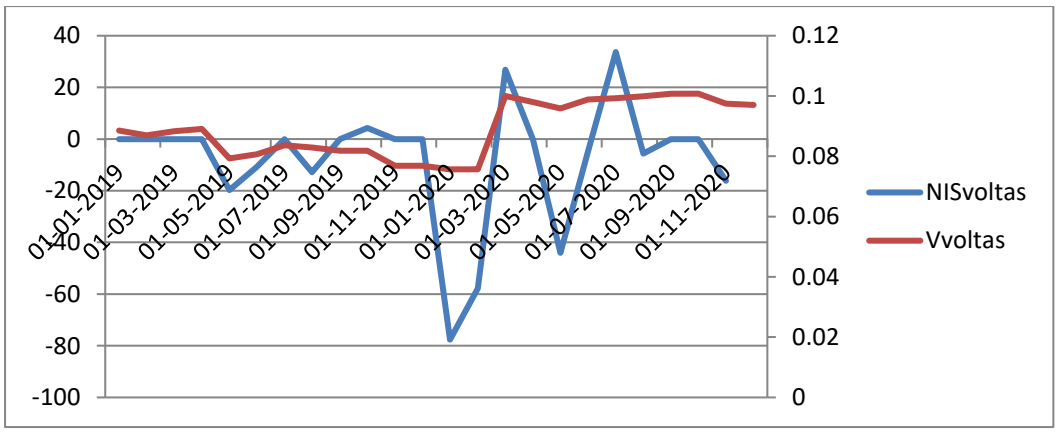
Cadila Healthcare: NIS and Volatility of Stock returns



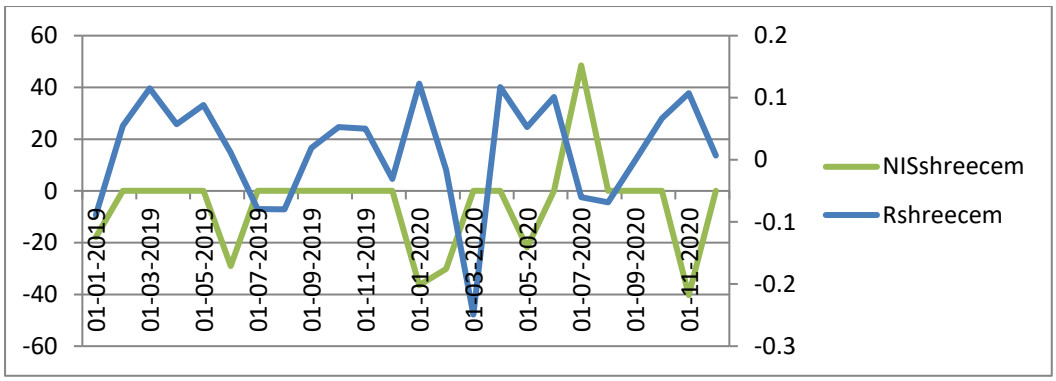
Voltas: NIS and Stock Returns



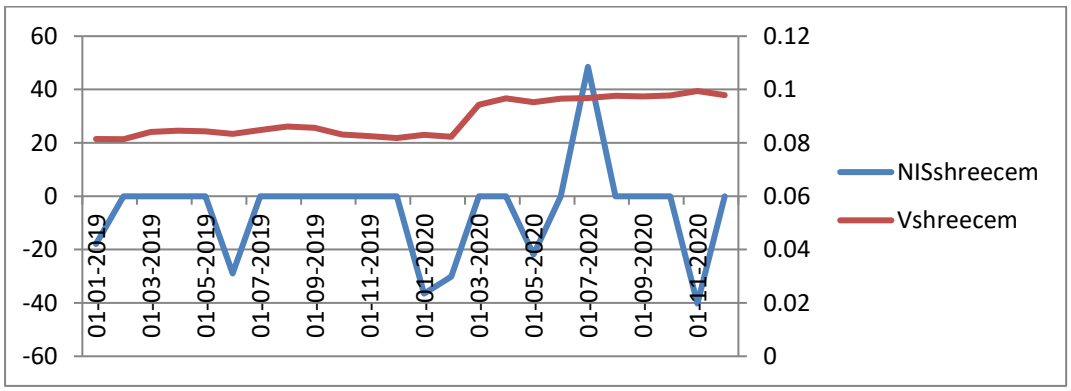
Voltas: NIS and Volatility of Stock returns



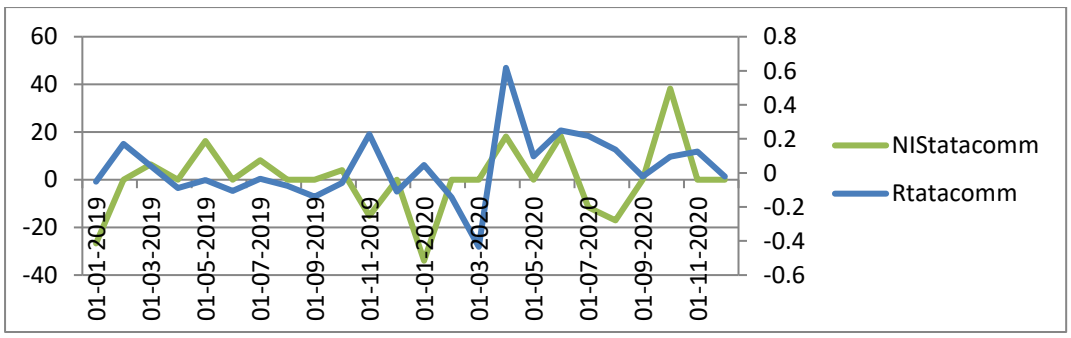
Shree Cement: NIS and Stock Returns



Shree Cement: NIS and Volatility of Stock returns



Tata Communications: NIS and Stock Returns



Tata communications: NIS and Volatility of Stock returns