# Stock Market Analysis

## Team Members

**Vaibhav Choudhary 01FB14EEC225 ECE 8th SEM**

**Ritik Kumar 01FB16ECS304 CSE 4th SEM**

## Mentor: Mr. Karthik S Joshi

# Report Submitted to Hitachi India Pvt. Ltd. as part of the

# Hitachi – Pentaho Promotional Activity conducted

# During June – July 2017

# At PES University

PESU-logo.png

# Introduction

We have been working on two problem statements, as follows:

**CESI**-

The stock market is a rendezvous point for buying and selling securities. It’s performance is measured using indices like Nifty-50.It is affected by many factors making it effectively random.

**Autocorrelation-**

For short time scales (~mins to ~day), the share price affects itself.

**Share price at t(i+1) = W(t) Share price at t(i-1).**

This phenomenon occurs because the market sentiments drive the share prices which in turn drive the market sentiment in reciprocation.

# Analytics Techniques Required:

* Cleaning and preparing data by filtering columns with missing values or possessing NULL.
* Validation and computation on the prepared data.
* Data visualization against different parameters.

# Problem Statement

**CESI-**

Our focus is on-

* How significant global events affect the market.
* Analysis of market trend depending on the magnitude of CESI.

**Autocorrelation-**

We propose to-

* Explore the nature of W(t).
* Approximate the time of persistence of this cycle.

# Methodology-

**1.Impact of Global Significant events:**

**Implementation:**

**CESI-**

We propose to quantify and study the impact of global significant events.

For instance - Lehman Brothers filing for bankruptcy.

To quantify their impact, we define-

Critical Event Sensitive Index (CESI) = (Previous Day SI) / (D-Day SI)

Source for data is NSE

We calculate the CESI. Pentaho in association with R is used for analytic and visualization purpose. Data is analyzed for various categories of events, in different timeframes.

The value of CESI implies *how much* the market went up or down. If the value of CESI>1, it implies that the market went down and the magnitude determines *how much.* Similarly, if the value of CESI<1, that implies that the market went up and the magnitude determines *how much*.

**Analytics Used:**

* Kettle (Pentaho Data Integration) features like Sort rows, Stream Lookup, filter rows was used to clean, sort and prepare the data.
* Computation of CESI is done using the calculator model provided in Kettle which takes 2 columns as input.
* Data visualization (Scatterplot) is implemented against the parameter year using R as plug-ins installed failed to work in the community version of Kettle.

# Results and Discussion

**CESI-**

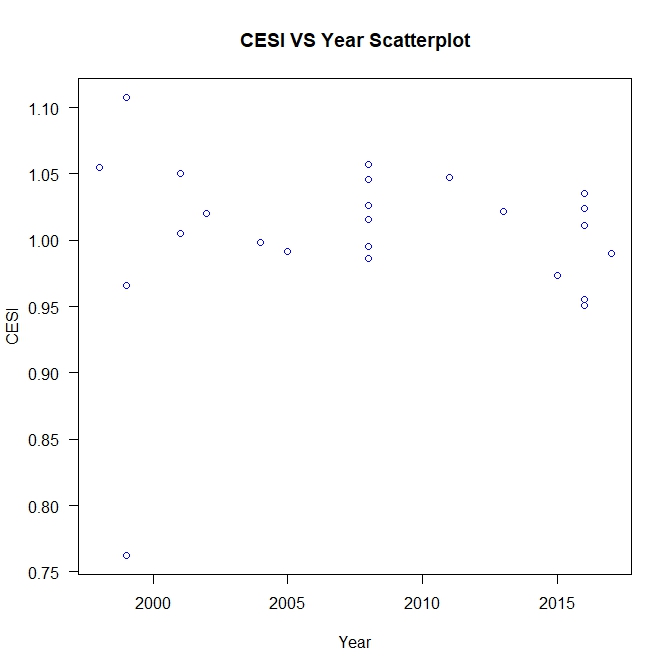
EVENTS TABLE AND SCATTERPLOT

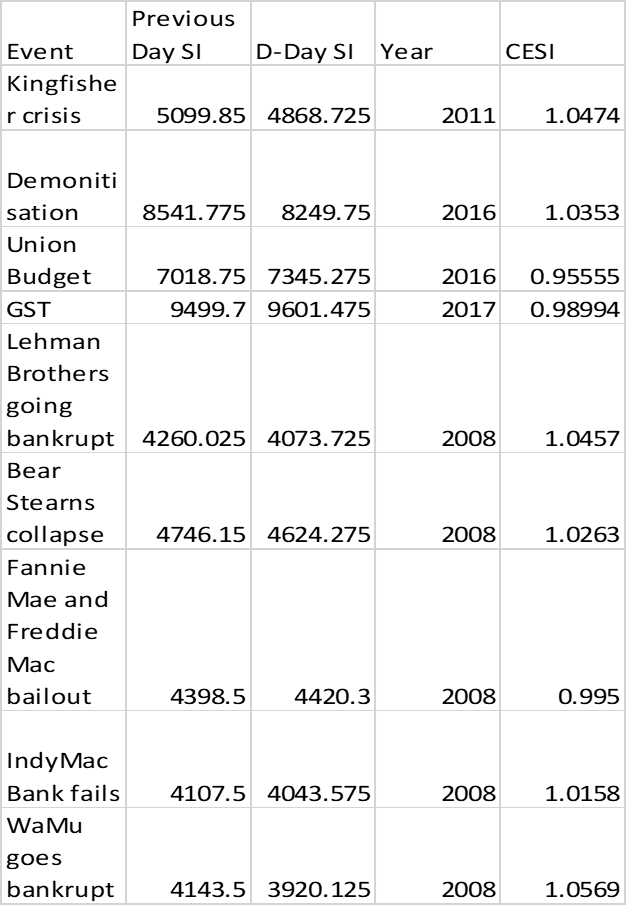
CESI for various events.

CESI= (Previous Day SI) / (D-Day SI)

CESI < 1 implies market went up.

CESI > 1 implies market went down

****

****

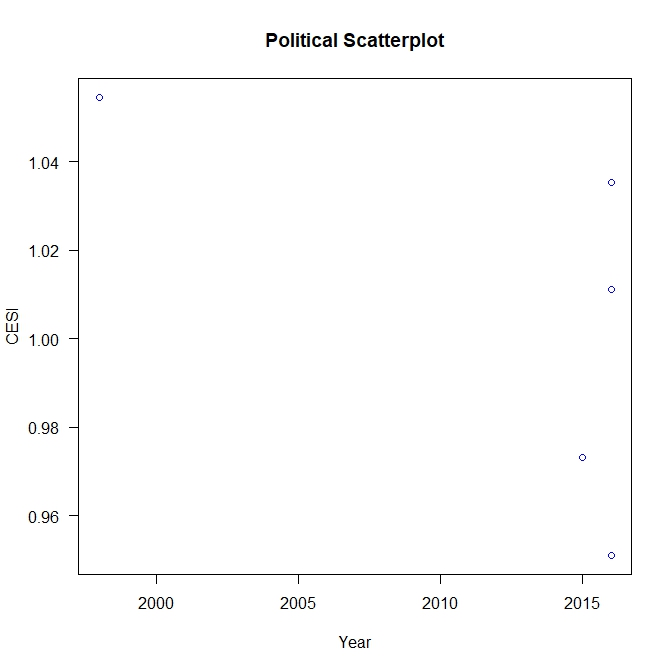
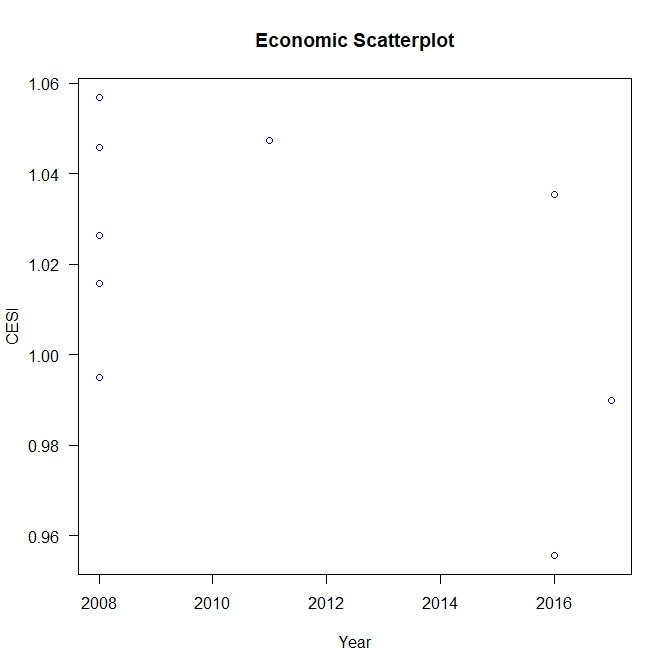


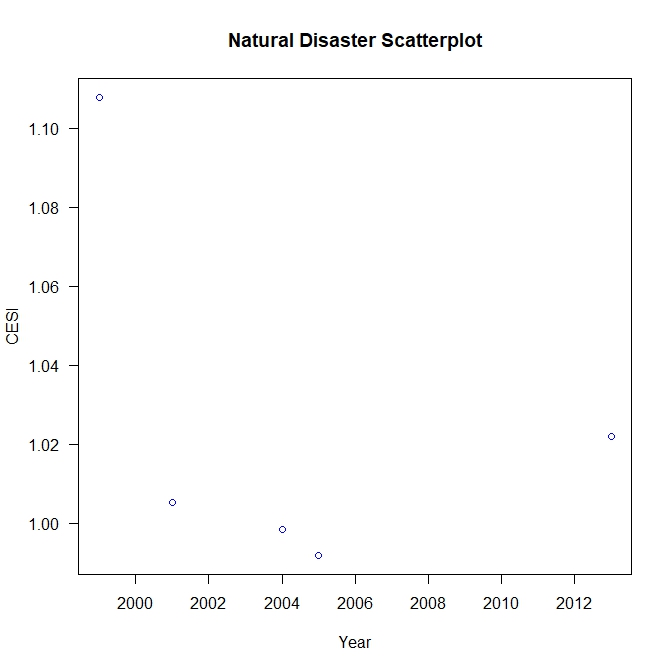
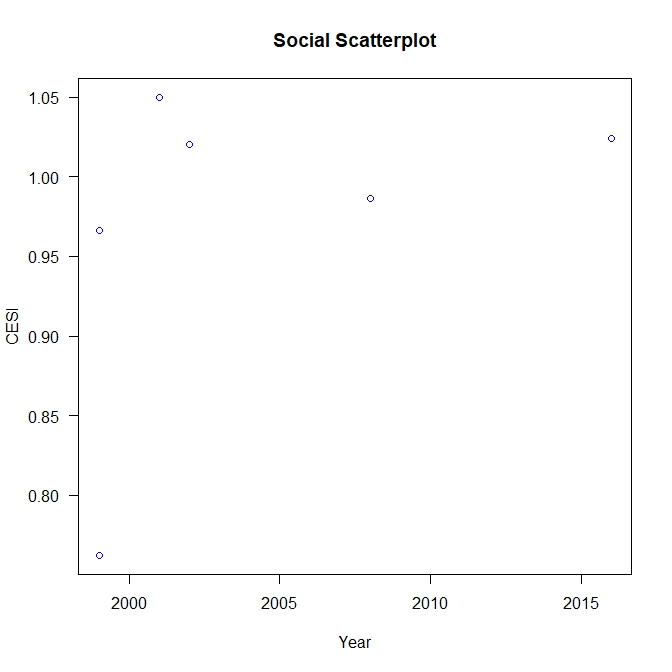
**Economic Political**





**Social Natural Disasters**

****



**Outcomes and Prospects:**

Based on our study we can infer that the Stock market was **not** significantly affected.

We propose to further the study-

* By taking more events into consideration.
* By attempting to explain the apparent indifference to significant events.
* Estimate the “time of influence” of significant events.

**2. Auto-Correlation**:

For short time scales (~mins to ~day), the share price affects itself.

**Share price at t(i+1) = W(t) Share price at t(i-1).**

This phenomenon occurs because the market sentiments drive the share prices which in turn drive the market sentiment in reciprocation.

We propose to-

* Explore the nature of W(t).
* Approximate the time of persistence of this cycle.

**Implementation:**

The first step requires the collection of data for all the Nifty-50 companies.

We start by assuming W(t) to be the auto-correlation function.

Further we calculate auto-correlation. Pentaho in association with R is used for the aggregation and plotting of graphs.

Data is analyzed for various categories of events, in different timeframes.

The value of autocorrelation in the range .95-.99 implies that the opening price and the closing price are strongly correlated.

# Results and Discussion:

Correlation between Open Price and Closed Price.

Autocorrelation between Closed Price.

A correlation of .95-.99 shows that open and closed price are strongly correlated.

# An instance of the Tabulated Data-

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Serial No. | Symbol | Correlation(Open Price-Close Price) | Autocorrelation(Close - Close) |  |
| 1 | INFY | 0.9714588 | 0.9724882 |  |
| 2 | AUROPHARMA | 0.9821403 | 0.9813279 |  |
| 3 | CIPLA | 0.9588861 | 0.955918 |  |
| 4 | DRREDDY | 0.989273 | 0.9859707 |  |
| 5 | HCLTECH | 0.9637116 | 0.96044 |  |
| 6 | ITC | 0.9897194 | 0.9393181 |  |
| 7 | LUPIN | 0.9875661 | 0.986928 |  |
| 8 | SUNPHARMA | 0.9907277 | 0.9893139 |  |
| 9 | ACC | 0.9866345 | 0.9848909 |  |
| 10 | AMBUJACEM | 0.986338 | 0.9837879 |  |
| 11 | BAJAJ\_AUTO | 0.953261 | 0.9496748 |  |
| 12 | BOSCHLTD | 0.9726461 | 0.9751798 |  |
| 13 | EICHERMOT | 0.9870206 | 0.9856032 |  |
| 14 | GRASIM | 0.9995426 | 0.988368 |  |

**Analytics Used:**

* Features like filter rows, Null If, Merge rows among others from Kettle (Pentaho Data Integration) is used to clean and prepare the data.
* Computation of auto-correlation and correlation is done via R as there is no such provision in the community version of Kettle.
* Data Visualization (Scatterplot) using R as the installed plug-ins failed to work in the community version of Kettle.

**Outcomes and Prospects:**

Better estimation of W (t).

We propose to use our study to-

* Understanding of the dynamics of stock market at different time frames.
* Prediction of trend reversal and possible fluctuations of stock prices.

**Overview of Kettle:**

* Kettle (Pentaho Data Integration) has a lot of features for data collection, data cleansing and preparation as it uses a graphical view called spoon. This makes the user interface very clean compared to other Data Analytical tools/languages.
* Kettle is easy to use for computation of general statistical modes among others.
* However, certain components like correlation coefficient cannot be calculated in the community version so that is a drawback.
* Since R can be integrated with Kettle so a lot of features can be directly accessed in R for basic purposes as integration process can prove to be a hassle.
* Data visualization plug-ins failed to work with the community version of Kettle so R was instead used which was easy to use.
* The online user forum is not interactive enough for an amateur user to solve his problems.

# Conclusion

**CESI-**

Based on our study we can infer that the Stock market was **not** significantly affected.

We arrived to the above conclusion with the help of the following :

* By taking more events into consideration.
* By attempting to explain the apparent indifference to significant events.
* Estimate the “time of influence” of significant events.

**Autocorrelation-**

With the study and implementation we achieved a better estimation of W(t). This in turn led to the following:

* Understanding the dynamics of stock market at different time frames.
* Prediction of trend reversal and possible fluctuations of stock prices.

# Bibliography-

For data extraction we used –

<https://www.nseindia.com/products/content/equities/equities/eq_security.htm>

<https://www.nseindia.com/products/content/equities/indices/historical_index_data.htm>

For details regarding global events ( e.g. Dates)-

[www.wikipedia.com](http://www.wikipedia.com)

<http://www.global-events.com/na/>

<https://www.infoplease.com/world-history-and-timelines>