<u>Impact of Technology on Market Surveillance</u>

Report is completed in partial fulfilment of the course

Practice School – 1

By Abhishek V Joshi,

2015A7PS116P

Under the guidance of Dr. Vardhana Pawaskar

Table of Contents

Acknowledgements

I am grateful to the organisation for giving me the opportunity to work with them. Special Thanks to the mentor, Dr. Vardhana Pawaskar (General Manager and Research Head) for providing her valuable inputs and insights for my work. A special mention is also made for the accompanying instructor, Dr. Monika Gupta for helping me out in every possible way. I acknowledge the team of fellow interns, Vaibhav Jain, Ashutosh Goyal, Parth Thacker and Raghav Goyal for helping me learn and explore the field of financial markets, which was previously unknown to me. I would also like to thank Viraj Bambulkar and Prasad Deshmukh from the BSE Research to help me identify some important features to further my work.

Abstract

Fraud Detection and Market Surveillance are of great importance to financial institutions like Stock Exchanges and Regulations Boards. Not only are these institutions interested in detecting the malpractices but also in imposing various regulations and limits to curb them. This report deals with the study that was conducted in BSE during an internship on the various Implications of a plethora of Technological Advances on Stock Market Surveillance. It includes the knowledge gained after literature reviews of the various technologies to detect and curb frauds and malpractices in Securities Market. In the end it also includes a newly proposed model employed to set Periodic Price Bands on Stocks.

Introduction

As the transactions have shifted from materialized paper ledgers to D-Mat accounts over the decades, the transactions have become much faster than earlier and have been safe-guarded against the traditional security threats. However new types of security issues are bound to occur with every new type of system; never can a system be truly safe.

With the expanded use of internet and shift of financial transactions online, new threats and malpractices linger around more than ever. These may range from a D-Dos attacks to Cross Trading to price Manipulation by strong arming all the transactions.

Computers have revolutionized the trading of securities in the secondary market and it is in the midst of a dynamic transformation. It is clear that the market of the future will not resemble the market of the past.[1] And hence the related Market Surveillance tools and techniques must also keep changing and adopting to the new types of threats and malpractices.

The following were the areas of study during the course of the internship:-

1) Data Mining Techniques for Fraud Detection in Real Time.
2) Block-Chain based products used in Market Surveillance.
3) Penny Stocks and Periodic Price Bands

Data Mining Techniques for Fraud Detection in Real Time

Data Mining very informally can be defined as the extraction of some knowledge (in the form of some structures) which was previously unapparent or unknown. These structures can be of the form of relations, hyper-geometric surfaces or rules in some regard. These structures are essentially the knowledge that one gains out of the vast reserves of data in any dataset. Deriving these structures is not trivial and hence comes the need of sophisticated algorithms and well defined problem statements.

The kind of knowledge structure that one needs to achieve not only depends upon the problem itself but also it largely depends upon the kind of data that has to be worked upon. The data itself may be available in a very raw form or in a form of a numerical representation. This data has to be cleaned and pre-processed into a suitable representation before it is fed to an algorithm for the derivation of any knowledge structure.

Only the derivation of the knowledge structure doesn't end the job. The structures themselves are very raw in form and are only representations of the result. It is the interpretations of these structures that truly develops knowledge.

Hence all in all a Data Mining task includes defining a well framed problem, conversion of the raw data into a process-able form, processing it to obtain some structure, then reverting back the structures to a higher level of understanding.

Why Data Mining?

There is no proper definition of a Fraud in Securities Market, so a programmer simply cannot write a deterministic algorithm for detecting Frauds. Generally speaking Frauds are just transactions which are unlike normal ones. If a hypothesis is put that the Normal Transactions form a specific pattern, then Fraudulent Transactions don't fall very well into these patterns.

There cannot be any deterministic algorithm which simply identifies an unknown pattern, but intuitively a structure of some sort may be able to capture these patterns mathematically. Further the structure has to be learned over time (here time signifies the number of examples), and hence a Learning Algorithm might be suitable for such a task.

A well framed problem statement would mean classifying a given transaction as fraudulent or in-fraudulent depending upon a captured pattern or behaviour from the past. Hence it is clear that Fraud Detection can be modelled as a binary classification task.

What kind of Data Mining Techniques can be employed?

A Classification task is a Supervised Learning Task and hence requires examples of fraud and non-fraud transactions. Due to unavailability of such labels in abundance, the classification task seems impossible. Although since there are far more number of normal transactions as compared to frauds, a pattern can still be captured using Unsupervised Techniques.

We may then infer that all those points which don't seem to lie in these captured patterns may be fraudulent, and hence require further investigation. Furthermore, all of this should be done in real time for quick and efficient investigation to take place. Real Time, here means that the model needs to be learned incrementally and so the algorithm must be a one-pass type (i.e. a single Data Point cannot be visited outside a short time frame).

Researchers have developed such IT systems that can detect fraud in financial markets in real time. The systems assess historical trading activity of traders and capture patterns. The determined patterns are used to make decisions on new trading activity of those traders to check if the activities fit the pattern or if there is some deviation. The system raises an alarm on finding differences and sends it to a business analyst for further analysis to determine the validity of the alert.[2]

Fraud Detection is also done by finding outliers in time series financial data using Peer Group Analysis, which is an unsupervised technique for fraud detection. PGA can detect those brokers who suddenly start selling a stock in a different way as compared to other brokers to whom they were previously similar.[3]

Most Stream Mining methods make an unrealistic assumption that 'labelled' data stream is readily available and can be mined at any time. However, in most real-world problems labelled data streams are rarely immediately available. A new model is proposed by Wei Fan et al.[4] which estimates the error of the model on the new data stream without knowing the true class labels. When significantly higher error is suspected, it investigates the true class labels of a selected number of examples in the most recent data stream to verify the suspected higher error.

Block-Chain based products used in Market Surveillance

Block-Chain is a Publicly Distributed Digital World Wide Ledger that keeps an account of all the transaction done in a Crypto-Currency and uses strong encryption algorithms for security. Bitcoin is the category creator of Block-Chain, which in analogy is similar to being able to send gold via email.[5]

It is the first decentralised P2P (peer to peer) payment network that is that is powered by its users with no central authority or middlemen. It is this unique property which has increased confidence amongst its users and attracted new people to invest in it, but it is this very property that raises doubts in the minds of people regarding the security of their assets.

Block-Chain keeps the record of all the crypto-currency based transactions taking place worldwide, it stores the asymmetric cryptographic hashes of users involved in the transactions, making the system pseudonymous. These records are completely unalterable due to its strong cryptographic security.

How does it work?

All the transactions are verified before they are added to the public ledger. This Transaction Verification process is computationally expensive and requires a high amount of area expertise. Special people known as Miners devote their computational resources and expertise to verify these transaction paths and balances across the huge network. Miners get newly generated Bitcoin in proportion to the work they do. They are the agents that make sure that the system works perfectly.

Latest Trends

Bitcoin is at its highest exchange rate with other materialized currencies as compared to any period earlier. This aptly shows the high amount of confidence investors have in this crypto-currency.

As of July 9, 2017; 1 Bitcoin was equivalent to 165127.29 Indian Rupee. [Source: https://google.com/]

NASDAQ has recently come up with a new Block-Chain based product of theirs for improving efficiency of their platform.[6]

LINQ by NASDAQ

In October 2016, NASDAQ unveiled LINQ, a solution enabling private companies to digitally represent their share ownership using Block-Chain based technology. On the surface it looks like a secure and complex database but within it is also an improved information distribution system.

The first participants to use LINQ included Chain.com, ChangeTip, PeerNova, Synack, Tango and Vera, all of which are involved in Block-Chain technology. Just as the underlying attribute for the success of Block-Chain technology has been trust as it is a publicly distributed ledger, NASDAQ is expecting its efficiency and transparency to be the foremost virtues of LINQ as it is a privately distributed ledger as opposed to Block-Chain.

NASDAQ manager Voss also believes that LINQ will solve yet-to-be-defined problems of the future too. LINQ is still in its early stages of development and experimental phase. As to what impact it will have on Stock Market Surveillance is still very uncertain and vague.

Penny Stocks and Periodic Price Bands

A penny stocks typically trades outside of the major market exchanges at a relatively low price and has a small market capitalization. These stocks are generally considered highly speculative and high risk because of their lack of liquidity, large bid-ask spreads, small capitalization and limited following, disclosure. Penny Stocks are not very well defined in the context of Indian markets as opposed to the Western markets.

Factors that make a penny stock risky:

1) Lack of information available to the public.
2) No minimum reporting standards
3) Lack of historical information
4) Illiquidity[7]

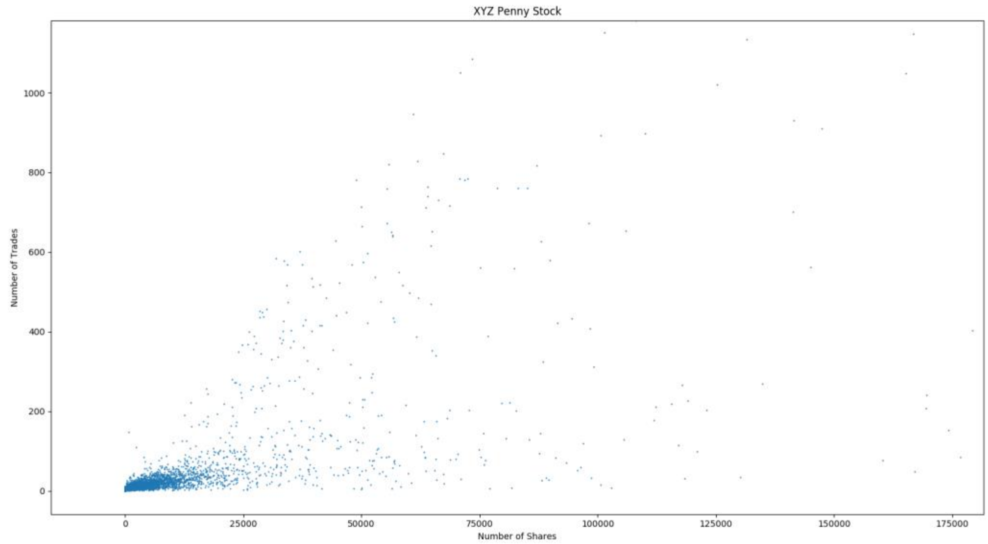
Why are these of interest?

These stocks can be easily manipulated by a fairly strong trader of a group of traders to their advantage. To avoid such malpractices from happening BSE uses price bands on daily fluctuation, different limits for different stocks. Although these daily price bands are not sufficient to control the overall fluctuation that occurs, hence a larger periodic price band is also used for every stock.

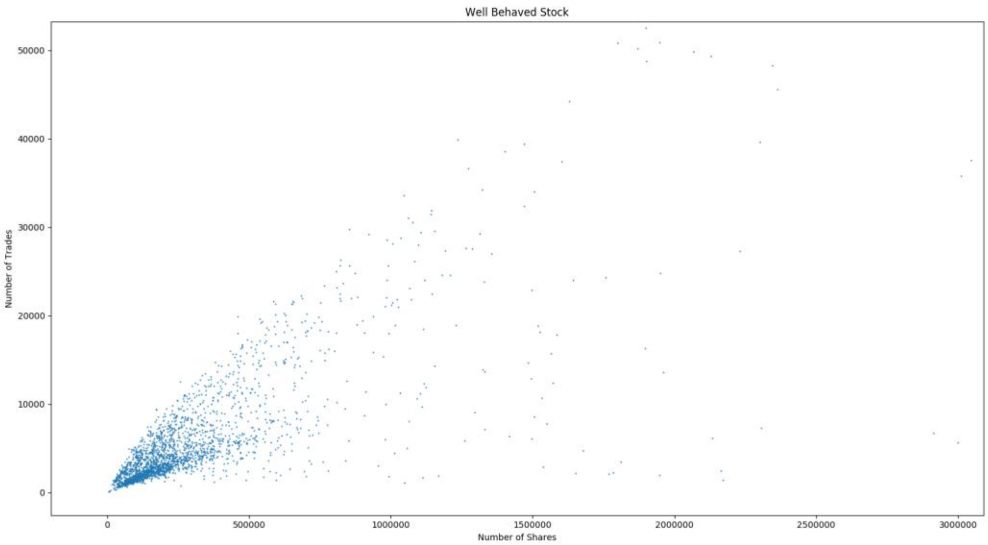
For example, say there is a daily price band of 10%, so even 8-10% positive fluctuation everyday will lead to a very large percentage increase at the end of a month or an year. Hence there is a necessity to control the stock movement over a period, and thus Periodic Price Bands. Thus Top Gainers are of main interest.

Initial Work to Identify features of Interest

Since trade frequency of the Penny Stocks is very low, very intuitively comparison of the graphs of Number of Shares Traded vs The Number of Trades per day for the Penny Stocks and the well-behaved stock ought to reveal interesting results. Shockingly when compared, the shape of the scatter plots of both the types of stocks came out to be very similar. Below are two examples in plots (1) and (2).



P1:Number of Shares Traded vs Number of Trades per day for XYZ Penny Stock



Plot 2: Number of Shares Traded vs Number of Trades per day for a Well Behaved Stock

Periodic Price Bands

Exchange had introduced periodic price bands for the securities that are exclusively traded at the Exchange.[8]

Daily	Weekly	Monthly	Quarterly	Yearly
20%	60%	100%	200%	400%
10%	30%	60%	100%	200%
5%	20%	30%	60%	100%
2%	10%	20%	30%	50%

Table – 1 : Periodic Price Bands

In addition to the above, for quarterly and yearly price bands, a factor of 200 % of **higher** of "S & P BSE Midcap Index" and "Sectoral Index" movement (subject to minimum 2% movement) shall be applied to the applicable periodic limits to take into account the overall movement in the market. This factor shall be added in the direction of index movement (rounded off to the nearest number) and threshold of the opposite direction will remain unchanged.

(Further details can be found on BSE's Website [8])

Need for a new Model of Evaluation

- The price bands are revised only on those days when there is an increase in 200 basis points of the respective index. So if there is a substantial movement of say, 190 basis points, but it still doesn't meet the set condition, the market movement is not captured and there is no revision of price bands for that particular day. Hence there seems to be a discontinuity in capturing the trends.
- The current system doesn't take into account the behaviour of the stock from the past.
- The price bands today are very static and discrete, in the sense that if a stock is imposed with a 10% band, then it will remain in that band only; it won't be revised to 9% or 12% depending on the new kinds of movements of the stock and its behaviour.

What needs to be improved?

- The overall market movement must be captured continuously along with the stock prices itself for better modelling of the behavioural trends.
- The Price Bands need to be changed daily on the basis of newer trends and not remain static and discrete. These Bands should adopt to the past behaviours incrementally.
- A common model should be built such that it effectively captures the trends of normal or so called 'well-behaved' stocks as well as the highly speculative penny stocks.

Some Important Identifications

- To judge the overall market movement corresponding to any stock, its respective closest Sectoral Index can be considered as a good choice.
- For varying time intervals only the relevant period of past shall be considered for proper behavioural modelling.
    - For example : For setting the annual price bands, the past 3 - 5 years may be considered, but for setting the quarterly price bands, only past 4 - 6 quarters will be suffice.
- Proper weightage must be given to the behaviour of the stock as well as its respective sectoral index. Hence, if the fluctuations in the stock are comparatively higher, then a higher weightage

The Proposed Model

A new model is proposed that takes into account above identified attributes and needs.

New Term Definitions

a_1 = average percentage change in price of stock over a period window in the positive direction.

b_1 = average percentage change in price of the sectoral index over a period window in the positive direction.

a_2 = average percentage change in price of stock over a period window in the negative direction.

b_2 = average percentage change in price of the sectoral index over a period window in the negative direction.

Formulas

- $a_1 = avg_{stock}[\frac{(\text{daily high price of the latest day in the window} - \text{closing price of the day before the window})}{(\text{closing price of the day before the window})} \times 100]$
- $b_1 = avg_{sector}[\frac{(\text{daily high price of the latest day in the period} - \text{closing price of the day before the window})}{(\text{closing price of the day before the window})} \times 100]$
- $a_2 = avg_{stock}[\frac{(\text{daily low price of the latest day in the window} - \text{closing price of the day before the window})}{(\text{closing price of the day before the window})} \times 100]$
- $b_2 = avg_{sector}[\frac{(\text{daily low price of the latest day in the period} - \text{closing price of the day before the window})}{(\text{closing price of the day before the window})} \times 100]$

New Parameter Definitions

$$\alpha_1 = \frac{a_1}{a_1 + b_1}$$

$$\beta_1 = \frac{b_1}{a_1 + b_1}$$

$$\alpha_2 = \frac{a_2}{a_2 + b_2}$$

$$\beta_2 = \frac{b_2}{a_2 + b_2}$$

New Metrics

$$p_{high} = \alpha_1 \times current\ percentage\ positive\ movement\ of\ the\ stock + \beta_1 \times current\ percentage\ positive\ movement\ of\ the\ sector$$

$$p_{low} = \alpha_2 \times current\ percentage\ negative\ movement\ of\ the\ stock + \beta_2 \times current\ percentage\ negative\ movement\ of\ the\ sector$$

These newly defined metrics shall be used to set the Periodic Price Bands on the price of stocks instead of the price itself.

Usage of the Metrics

- A list of such p_{high} and p_{low} shall be computed for the required period of the past over a given time window, by simply using the daily highs for p_{high} and daily lows for p_{low} of the last day of the time window, instead of the current percentage movements.
- A band shall be set upon the computed list such that about 90-95% of points lie in that band. [**NOTE:** The band must be central, meaning it should be equally distributed about the mean of the list. In the implementation of the actual model, k standard deviations are used to contain the points centrally (where k is any suitable quantity).]
- If the margin on the lower side of the positive band or the higher side of the negative band happens to cross the zero barrier, it should be neglected.
- To make the model computationally cheaper, the required quantities of mean and standard deviations can be computed incrementally.

Hypothetical Example

Consider a Stock with a_1 = 12.5% and its respective Sectoral Index with b_1 = 2.5%

For the simplicity of understanding only the positive movement of market is considered for now, extension on the negative side is pretty trivial.

Here, $\alpha_1 = 0.833$ and $\beta_1 = 0.167$

Hence it is clearly evident that the weightage of the stock with respect to its sectoral index is proportional to its average movement from the past.

Table – 2 : Example

Time Instance	Percentage Positive movement of stock	Percentage Positive movement of the respective Sectoral Index	p_{high}
1	10%	10%	10%
2	0%	10%	1.67%
3	10%	0%	8.33%
4	20%	5%	17.5%
5	5%	20%	7.5%

Hence it is the α and β quantities that actually capture the trends/behaviour from the past.

Analysis of the Time Instances

1. Whenever there is an equal movement in the stock as well as its respective sectoral index, the metric will also have the exact same movement. One can infer that the rise in stock is not just due to investment in their shares but actually due to investment in its sector at large.
2. The small number is indicative of the fact that the movement of the sectoral index is captured even though there's no positive movement of stock. Such points need to be worried about as they show a comparatively abnormal behaviour from the overall market.
3. The loss of 1.67% can be attributed to the zero movement of the sector, and thus all the movement can be attributed to the stock itself.
4. The loss of 2.5% as compared to the original price is due to the 5% movement of the sector, meaning that the actual movement contribution of the original price from the stock is just 17.5%.
5. Again a rise of 2.5% can be inferred as the actual movement that should've been there but is not present.

Validation of the Model on actual data

Two Top Gainers Penny Stocks were identified by the organisation (BSE) for studying and experimenting; let's call them ABC and XYZ. The above Proposed Model was applied on ABC and XYZ independently and results were noted down.

ABC

Table – 3 : The average percentage movements of ABC stock and its respective sectoral index on the given number of days as the time window. (The 0.0 or similarly low percentages are indicative that the stock or the index moved in the opposite direction as that considered while calculating the percentages)

	1	5	21	63	252
Positive movement of Stock	3.09%	7.61%	21.72%	48.72%	322.54%
Positive movement of Sectoral Index	0.85%	1.60%	3.65%	8.25%	25.03%
Negative movement of Stock	-2.37%	-4.04%	-6.31%	-3.96%	0.0%
Negative movement of Sectoral Index	-0.67%	-1.27%	-2.08%	-2.51%	-0.07%

The p_{high} and p_{low} were computed for 311 days as the data was available, margins were set such that about 90% of the points lie within it, and then these margins were tested on a further of 65 points. The Percentage of points out of the last 65 lying within the set margin are shown below.

Table – 4 : below shows the percentage of test points lying within the tightly set margin that captures the behaviour from the past.

	1	5	21	63
On the positive side of movements	92.19%	90.63%	96.88%	100%
On the negative side of movements	95.31%	93.75%	100%	100%

The model here does a pretty good job of restricting points that may be too out of normal behaviour on the first three Time Windows. The 100% inclusion of the last time Window was initially of some suspicion, but once the returns plot for that particular time window were plotted, it was observed that the stock had some movement initially but then its fluctuations tapered down; so the set margins ought to include all the latest points. Hence the results are consistent.

XYZ

Table – 5 : The average percentage movements of XYZ stock and its respective sectoral index on the given number of days as the time window. (The 0.0 or similarly low percentages are indicative that the stock or the index moved in the opposite direction as that considered while calculating the percentages)

	1	5	21	63	252
Positive movement of Stock	5.04%	9.67%	26.46%	84.79%	978.90%
Positive movement of Sectoral Index	0.77%	1.4%	2.73%	5.22%	13.03%
Negative movement of Stock	-2.96%	-3.41%	-2.70%	-1.13%	0.0%
Negative movement of Sectoral Index	-0.62%	-1.01%	-1.62%	-1.88%	-0.08%

The p_{high} and p_{low} were computed for 311 days as the data was available, margins were set such that about 90% of the points lie within it, and then these margins were tested on a further of 165 points to increase the granularity of percentages to gain more information. The Percentage of points out of the last 165 lying within the set margin are shown below. This could be done due to availability of comparatively more data than ABC stock.

Table – 6 : below shows the percentage of test points lying within the tightly set margin that captures the behaviour from the past.

	1	5	21	63	126
On the positive side of movements	93.33%	93.94%	96.97%	98.79%	34.55%
On the negative side of movements	92.12%	90.91%	92.73%	94.55%	63.64%

The model again does a pretty good job of setting margins such that most of normal points lie within the margin for the first four Time Windows. The absurdly low percentages for the half yearly plot is also consistent as there was too much fluctuation over that time window.

Conclusions and Further Work

- A Literature Survey was done on Data Mining Techniques for Fraud Detection in Real Time.
- A Literature Survey was done for Finding out the Advantages of Block-Chain based products for Stock Market Surveillance.
- The newly proposed model is consistent with the data on which it was tested. It is very simplistic but the results are encouraging.
- Further Work would include Back-Testing on various Time Horizons and testing the consistency on over various such horizons.

References

[1] Solomon, L. D., & Corso, L. (1990). The Impact of Technology on the Trading of Securities: The Emerging Global Market and the Implications for Regulation. *J. Marshall L. Rev.*, *24*, 299.

[2] Subramanyam, J. B. V. (2014). *The Proceedings of the International Conference on Information Engineering, Management and Security 2014: ICIEMS 2014* (Vol. 88). Association of Scientists, Developers and Faculties.

[3] Ferdousi, Z., & Maeda, A. (2006). Unsupervised Fraud Detection in Time Series data. In *Proceedings. 22nd International Conference on Data Engineering Workshops*.

[4] Fan, W., Huang, Y. A., Wang, H., & Yu, P. S. (2004, April). Active mining of data streams. In *Proceedings of the 2004 SIAM International Conference on Data Mining* (pp. 457-461). Society for Industrial and Applied Mathematics.

[5] https://weusecoins.com/

[6] https://business.nasdaq.com/

[7] www.investopedia.com/terms/p/pennystock.asp

[8]http://www.bseindia.com/investors/periodic_price_bands.aspx?expandable=5