DATA ANALYSIS OF COMMODITY PRICES

A thesis submitted in partial fulfillment
of the requirements for the degree of

MASTER OF TECHNOLOGY

in

Computer Science & Engineering

by

KAPIL THAKKAR

Entry No. 2014MCS2124

Under the guidance of

Dr. Aaditeshwar Seth

Department of Computer Science and Engineering,
Indian Institute of Technology Delhi.

Jan 2016.

Certificate

This is to certify that the thesis titled **DATA ANALYSIS OF COMMODITY PRICES** being submitted by **KAPIL THAKKAR** for the award of **Master of Technology** in **Computer Science & Engineering** is a record of bona fide work carried out by him under my guidance and supervision at the **Department of Computer Science & Engineering**. The work presented in this thesis has not been submitted elsewhere either in part or full, for the award of any other degree or diploma.

Dr. Aaditeshwar Seth
Department of Computer Science and Engineering
Indian Institute of Technology, Delhi

Abstract

Supply demand imbalance, natural calamities etc. may not always be the reason behind the rise in the price of a commodity. It may be a result of artificial supply deficit planned intelligently by traders nexus to earn more profits through manipulation of supply of commodity and hence indirectly controlling their prices. Our attempt is to locate such hikes in prices which seem suspicious (we call them anomalies). We try to detect these anomalies by characterizing them and forming hypothesis based on their characteristics and then algorithmically try to find the days in time series during which it violates the stated hypothesis.

Acknowledgments

I would like to express my heartiest gratitude to our supervisors Dr. Aaditeshwar Seth for guiding this work with utmost interest, patience, care and scientific rigor. We thank him for setting high standards, giving us freedom to explore multiple facets of the problem and teaching us value of analytical thinking and hard work.

I would also like to thank Dipanjan Chakraborty who have helped a great deal by providing technical guidance and support whenever needed.

KAPIL THAKKAR

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Supply demand imbalance, natural calamities etc. may not always be the reason behind the rise in the price of a commodity. It may be a consequence of artificial supply deficit planned intelligently by traders nexus for profiteering through manipulation of supply of commodity and hence indirectly controlling their prices. Our attempt is to locate such hikes in prices which seem suspicious (we call them anomalies). To detect and analyse the characteristics of anomalies in the prices of commodities. Currently we have considered the case of onion and based on that we are looking forward to develop library for any time-series.

1.2 Objective

Our objective is that we need to highlight anomalies which may be an indicator of illegal market manipulation act by traders nexus in the provided time series.For this purpose we need to create library with some set of functionalities to detect anomalies in the give input time series. Anomalies will be based on the hypothesis stated by the user. For particular scenario stated by user, execute appropriate algorithms and report the anomalies in the time series.

1.3 Relavence of Project

Anomaly detection techniques help to explore situations which might be different from the expected behavior and could reveal interesting facts. It is used many areas which are explained in detail in next chapter. The project

aims to raise potential red flags for days which are suspect of illicit market manipulation activities by set of traders. This type of monitoring system may help people monitor and hence control these illicit market manipulations better. For example, unnecessary hike in the prices may be due to wrong government policies, loopholes in the supply chain of commodity, intention of profiteering by traders etc. So our project will help journalist or end user interested in detecting such abnormal behavior.

Chapter 2

Literature Survey

2.1 What is Anomaly Detection?

According to wikipedia, anomaly detection (or outliers detection) is the identification of items, events or observations which do not conform to an expected pattern or other items in a dataset.

Here, our major focus is on detecting anomalies in the time series data. Time series usually considers data about price of some commodity, production, sell, etc. Usually, these time series follow some normal pattern. Some time series may be independent or behavior of some may depend on other time series. It may also consist of seasonality and trend along with some noise. So, considering all these factors our aim will be to detect some points on time line during which these time series does not follow a normal pattern.

Reasons for presence of anomaly may be different depending upon the type of time series. We have considered time series of onion data as a use case. In case of onion presence of anomalies could be because of unseasonal rainfall, hoarding, price manipulation by traders' nexus, effect of import/export of onions, variation in production, etc. Being a seasonal crop, some part of onions are stored during harvest, so that demand can be met in lean season. But some traders hoard huge stocks for the purpose of profiteering. According to wikipedia, in economics, hoarding is the practice of obtaining and holding scarce resources, possibly so that they can be sold later to customers for more profit. So during this time also, we are able to see anomalies.

2.2 Onion Case

Onion is a staple ingredient for almost every Indian kitchen and hence its demand is almost constant throughout the year but not the supply. In order

to supply onions throughout the year, they are stored during harvest and released into markets in lean seasons. Its importance can be well estimated by the fact that it is one among few essential commodities and often rise in its price has resulted into downfall of state and central government.

One major tragedy in onion market occurred in end of year 2010. The prices of the onion increased so much that it was out of reach from poor people. There was a study conducted by the CCI (Competitive Commission of India) for this case and they created report on that [18]. In this [18], they tried to find out the reasons behind this scenario. They came with the following things in their study:

- Large wholesalers/traders mainly operates in metropolitan city markets and large number of farmers dispose their bulk of produce in nearby markets because of absence of storage facility, immediate cash need for loans, family expenses, purchase of inputs of next season, etc.
- Concentration of large storage capacities with traders,Vertical Integration of various market functions by onion traders(one name, many roles), Existence of established traders and barrier to new entry
- On December 23 of 2010, The Times of India published in an article that on Tuesday alone, wholesale traders in Delhi bought onion at about Rs.34 per kg while it was sold in retail at Rs. 80 per kg, the margin of Rs. 46 per kg or 135 %.
- In the weeks of November and December, wholesale price remains high, so retailers do not get much profit, but even after that when wholesale price go down, retailers particularly in metro cities, show strong rigidity in holding price and earn margin from 60 to 110 %. This clearly shows that along with traders, retailers also exploit the situation of crisis for their own benefits.
- If we take this forward, then government policies also had a great role in the December 2010 high price episode (export of 1.33 lakh tones onion in October 2010).

So if we consider its overall picture, then there was unseasonal rainfall in the month of September and October 2010, but after that also government policy regarding export of onion was unexplainable. The news article published in Times of India also questions why there is so much difference in the wholesale and retail price of onions. Study also suggests that all the traders operating in the market have experience of many years (20 years on an average) and this is sort of family business for them. Due to limited entries, there is also barrier for new traders in entering to the market. So no new person enters and because of existing traders monopoly, they operate in the market together by forming the nexus. So many times they can alter the prices in many regions so that farmer has to pay money they decide. Traders have monopoly in onion markets and due that prices do not follow normal behavior of demand-supply and goes out of the way.

2.3 Other Cases

2.3.1 Sugarcane Case

This [21] study on sugarcane shows the connections between Politicians and Sugar Mills in Maharashtra and explains how such connection may benefit to firms and politicians. Sugar mills are cooperatives and regions are formed according to mills present. Each sugarcane farmer has to sell his produce to the mill present in his region, he can not sell it to some other mill. Each mill pays its farmers a single price per metric tonne of cane every year, based on weight (not on the quality).

This study investigates how price of the sugarcane, paid to farmers, changes in the election year. Usually, chairman of the sugar mills are politicians who stands in election. They need funding for elections, so here author explains how sugarcane mills and election funding may be related. And if some chair person wins the election then what is effect on prices offered to farmers. Some findings are:

- Prices are lowered by about Rs. 20 a ton in politically controlled mills during election years

- The results were robust to including rainfall and mill capacity as controls, as well as including mill-specific outcomes such as the recovery rate - sugar produced per unit cane, a measure of productivity - as well as various other mill level shocks such as mill breakdowns and cane shortages
- Price fall may be due to mill closure, i.e. mill is not operating for profitably, but politicians has kept mill open as a way to gain votes. But analysis shows that mill closure is not affected by political control
- Paying farmers Rs. 20 per ton less for their cane amounts to a total of Rs. 6 million
- Mills whose chairmen won national elections pay Rs. 80 per ton more in the year after elections
- Author finds that when the party affiliated with the mill chairmen is in power in Maharashtra, the mill pays Rs. 23 more in cane price and also Chairmen who win national elections seem to be able to keep their mills open far more successfully than chairmen who lose

So here author has strong belief that all facts indicates that funding for election campaign comes from these sugarcane mills if mill is politically controlled. Reason why farmers supports this may be that, with average probability 1/3rd of winning election, on an average farmer gets Rs. 27 on their principal of Rs. 20, so still in profit. The overall effect on farmer welfare is difficult to determine. On average, cane prices and recovery rates in politically connected mills are no different than those in non-politically connected mills, and the levels of public goods are no different either.

So this example explains how time series may go out of their normal behavior though there is no supply-demand crisis or any other case.

2.3.2 Builder-Politician Case

This [20] study shows in developing countries like India, where elections are costly and accountability mechanisms are poor, politicians often turn

to private firms for illicit election finance. Land is one of the highly regularized sector in India which provides discretionary power in the hands of state(indirectly in the hands of politicians). It is easy for politicians to accumulate resources than to hide them. To hide these assets from scrutiny politicians often use real estates as medium because of its features which are absorptive capacity,liquid asset and Contract enforcement. As elections approach, however, builders are often compelled to provide politicians with money with which to contest elections; the mechanism can be a simple under-the-table transfer or an in-kind contribution. Although builders have to transfer funds back to politicians around elections, the transaction brings long-term benefits in terms of future goodwill. Based on this author have quoted following hypothesis:

- Cement consumption should exhibit a significant contraction during the month of the state election.Because builders are a leading source of election finance, one would expect activity in the sector to slow down during the month-long campaign period prior to Election Day.
- Contraction in cement consumption will be significant in national elections, though of a smaller magnitude than in state-level elections
- The magnitude of the contraction in cement consumption to be larger for dual elections than if only a state or national election is being held.
- cement consumption should exhibit a larger contraction in urban versus rural states.
- The contraction in cement consumption should be comparatively larger in more competitive elections

The paper was able to bring the quid pro quo relation between builders and politicians quantitatively.

Chapter 3

Study of Onion Data: Collection and Analysis

3.1 System

For now, we are only working over onion data. We have three actors in model: Farmers who are producers of onion, traders who are collectively responsible for supply of onions across country and consumers who purchase onions.

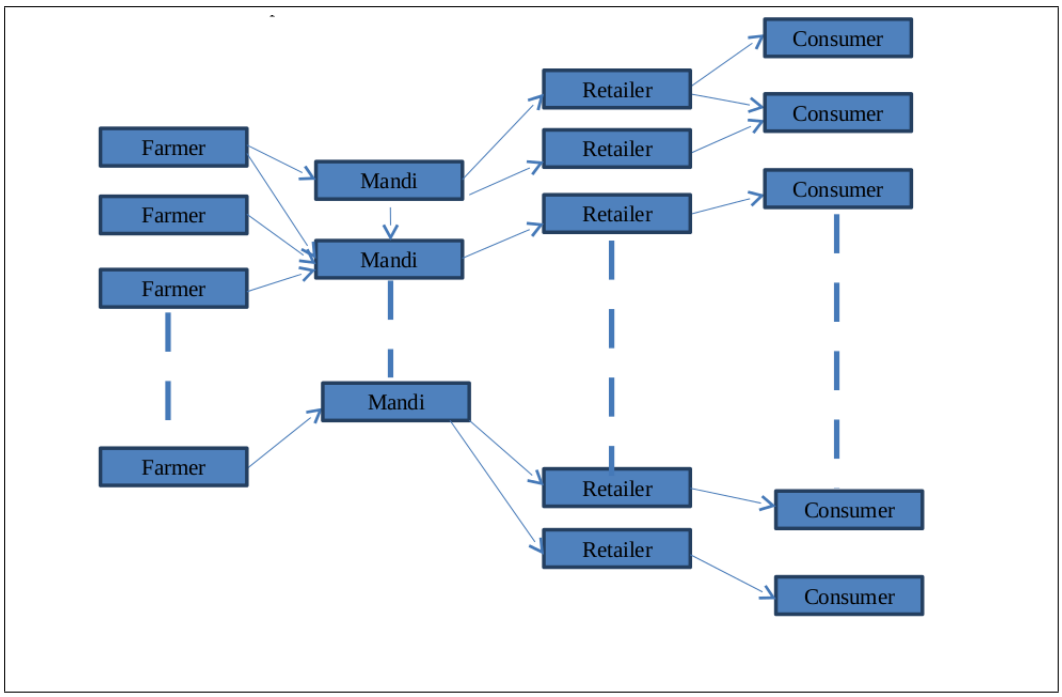

Figure 3.1: Normal Supply Chain

Farmers sell their produce to traders in nearest mandi offering better price. These traders sell these commodities to traders in other mandis or to retailers. Consumers purchase commodities for consumption from one of the retail stores. This way commodity reaches consumers from farmers following a huge chain of traders and retailers.

Under APMC act, mandis were established at different places across country so that farmers can sell their produce directly in mandi and get good returns (wholesale price). There are around 1500 mandis located in different places across country which log their daily arrival of onion, minimum, maximum, modal selling price per quintal of onion data to AGMARKNET. Retailers purchase from these mandis and sell to end customers at retail price. There are around 70+ centres across country which maintains retail price of onion on Ministry of consumer affairs website.

3.2 Data we have

We have following data:

1. Daily wholesale price of onion for 1514 mandis
2. Daily arrival of onion information for 1514 mandis
3. Daily retail price of onion for 76 centres
4. Dates and location for hoarding reports from news articles
5. Longitude and latitude of mandis and centres

Onion Data was collected from the the government websites,[1] for arrival and wholesale price data and [2] (Department of Consumer Affairs) for retail price data. Crawlers were written to collect the data from these datewise for the period of approximately 9.5 years, starting from Jan 2006 to Jun 2015. More data can be added simply by running crawlers again.

3.3 Normal market behavior

1. Wholesale price is inversely proportional to arrival of commodity higher production of crop will lead to more and more crop hitting market for sell. Hence more arrival which will result in surplus supply leading to drop in wholesale prices.

2. Retail price is directly proportional to wholesale price commodities reach customers through a long chain of traders and retailers, adding value at every stage of chain. So, retail price at which customers purchase commodities are more than wholesale.

Any divergence from these characteristics of normal market leads to suspicious price hike situations/anomalies.

3.4 What are the reasons for anomaly?

Primarily there are 3 main reasons of anomaly.

1. **Government Policies:** When the production is low in the country, still government allows the export of onion in large amount, or supports it by keeping low minimum price then the prices can rise up drastically.

2. **Unseasonal Rainfall:** Due to insufficient, heavy or unseasonal rainfall, onion crop may get affected and the produce is low and wholesale price may rise up. But, this reason still is validating that wholesale price is inversely proportional to arrival, it may be just prices will be little higher than what was supposed to be.

3. **Hoarding:** When traders/wholesalers store the onion and does not release the stock in the market in the expectation of the good prices in the future, it will create the artificial deficit in the market and will shoot up the onion prices in the retail market due to low arrival in the retail market. The reason people do this is to expect the higher prices in the time of low production or may be for security. For example, if it is expected that in some year the rainfall is not good, then people may predict production to be low in the future and so they will start storing onion so that they can gain more profit. It will also create deficit in the market and price will go up.

So our study will focus on detection of anomalies in data and if possible comment on the possible reason for the anomaly.

3.5 Mapping of wholesale price to retail price

Voronoi Diagram is used to map every mandi to nearest possible centre. The centres with retail data were considered fixed points and country was divided into 76 regions. All the mandis falling in that region are mapped to the respective centre.

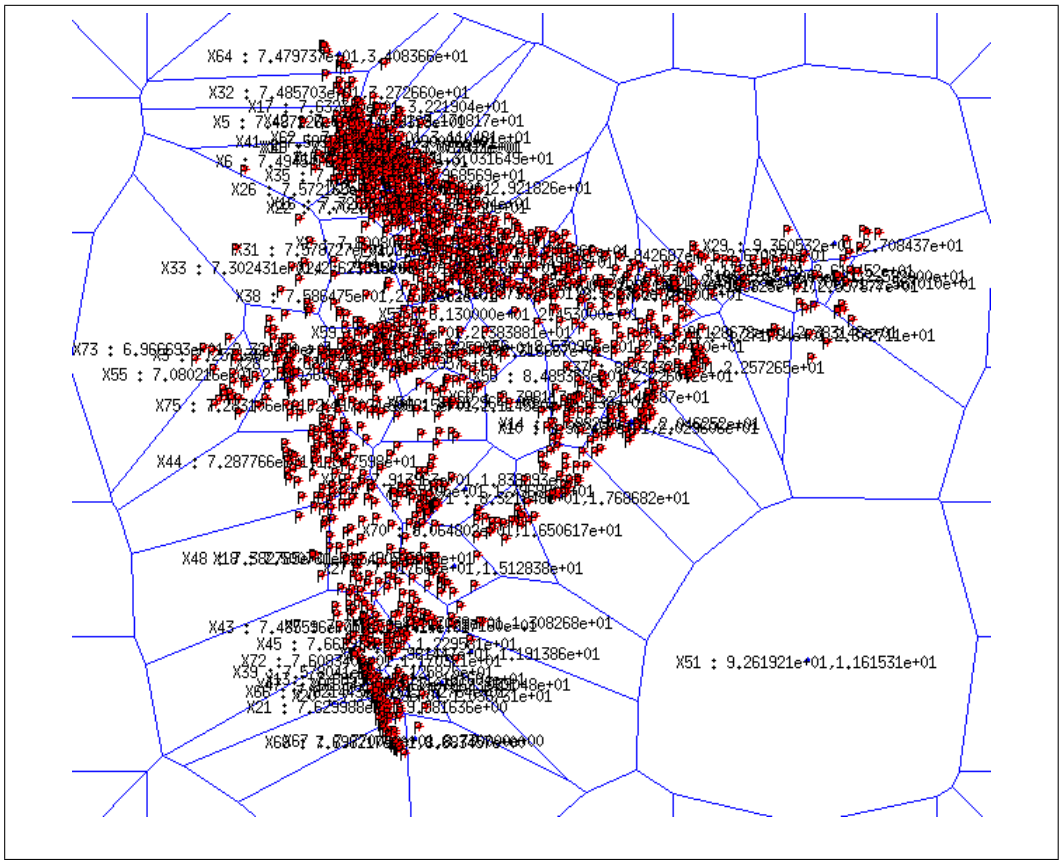

Figure 3.2: Voronoi Diagram

After all the mandis are mapped to their nearest centre, wholesale and arrival at every centre is computed. Wholesale price at centre is average of modal price of all mandis in its region and arrival was computed as the sum of the arrival at the mandis in its region. While calculating the wholesale price, the distance between centre and the mandi was not considered.

Corresponding to every date and location of hoarding news report, values like current year arrival, last year arrival, Percentage difference in wholesale-retail etc. were computed. Following table resulted from these computations.

3.6 How to Define Anomaly?

To answer this question, we went through the series of the news articles when the hoarding is in the news. Then looking over those articles, we try to see that why they are reporting in news, what happened so that people are giving it name of hoarding and how reporters are making conclusion that it may be the class of the hoarding.

3.6.1 Summary Of News Atricles

First such incident was reported in the 1998. The article [14] dated on 21 st September 1999 states as follows:

Onions were retailed at Rs 6 a kilo two weeks back. Today, the price was almost 100 per cent up, hovering in the Rs 10-12 band in different parts of the country.

So this article is comparing the retail price of today with the price before 2 weeks. The rise upto 100% is what has come to notice. Also article says that,

There is talk in the market that the government is likely to lift the ban on onion exports. Apparently, some traders are resorting to hoarding in anticipation of demand from markets abroad.

As stated previously also, government policy also plays a major role in this. After that Onion was in news in 2010. NDTV [15], TOI [5] and many more reported the incident. In 2010, unseasonal rainfall and the government policy on export price were also the reason for hike in the price. The report dated Dec 23, 2010, TOI states the follows (in Delhi):

On Tuesday alone, wholesale traders in Delhi bought onions at about Rs 34 per kg while it was sold in retail at Rs 80 per kg. That's a margin of Rs 46 per kg or 135%!

Here, they have compared the difference between the wholesale price and the retail price. The margin of 135% is reported. When we looked into data we have, we got the following results for Dec, 2010. (See figures 3.3 and 3.4)

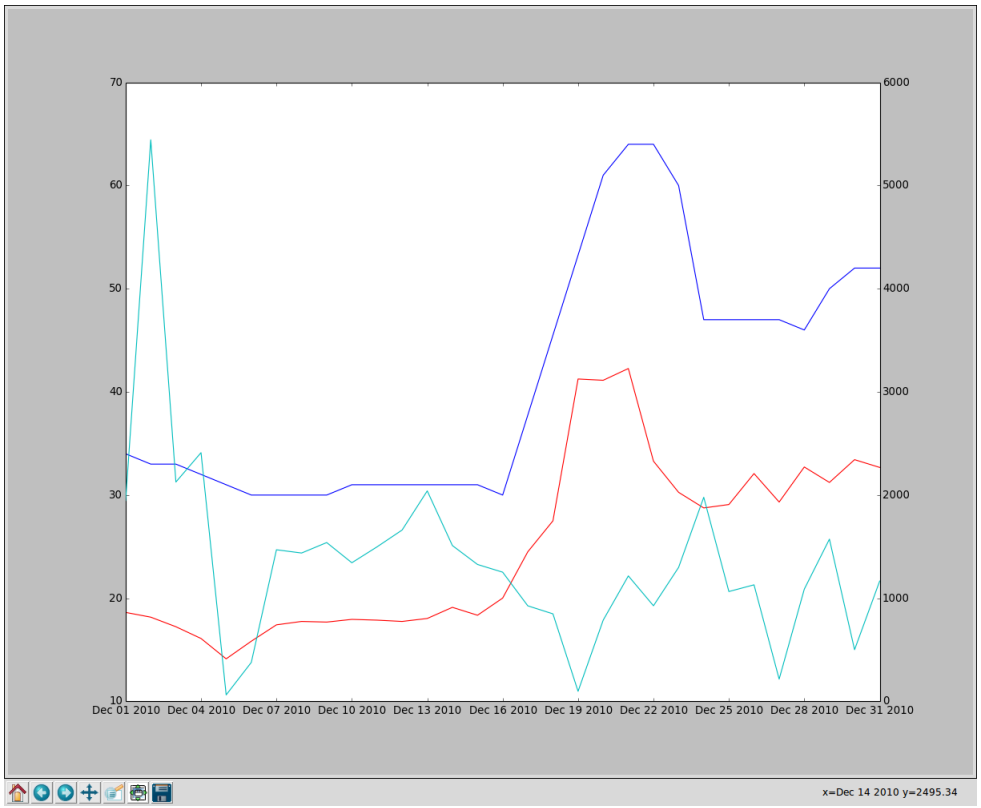

Figure 3.3: Delhi, Dec 2010. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival)

So, as per our data, the maximum price difference observed was of $\tilde{1}00\%$. Note that the retail prices we have is the minimum price observed in the market.

NDTV on Dec 22, 2010 reported the following (NASIK):

"The average purchase rate of a trader here is about Rs. 3,000, but they go to the cities and claim it's nearly Rs. 8,000 and that's how the rates go up. It's all the fault of the traders. They loot the people," said Sangdeorao Holkar, Director, National Agricultural Cooperative Marketing Federation of India.

Here also, as we see they have mentioned price difference between retail and wholesale in the market. It is approximately $\tilde{1}66\%$.

Report also adds the following:

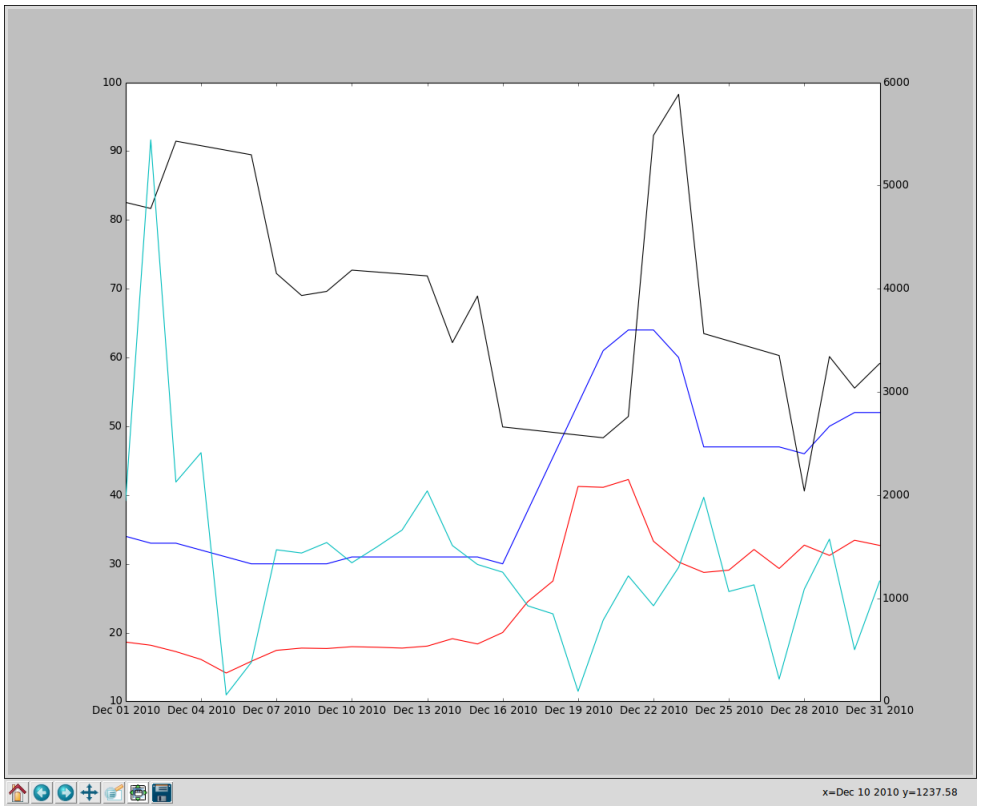

Figure 3.4: Delhi, Dec 2010. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival,Black - Relative difference %)

The government, on a back foot, banned exports last evening. In less than 24 hours, prices in Lasalgaon crashed by 25 per cent.

Navi Mumbai: Wholesale price
Tuesday: Rs. 60 per kg
Wednesday: Rs. 45 per kg

And this is finally what the consumer in Mumbai is paying:

Mumbai: Retail price
Tuesday: Rs. 60 to Rs. 70 per kg
Wednesday: Rs. 60 to Rs. 70 per kg

So as we can see, there is significant drop in the wholesale price, but no drop in the retail price, so that has been reported. Also difference in the hike in retail price, from 40 to 60 was reported in on week. What our data says:

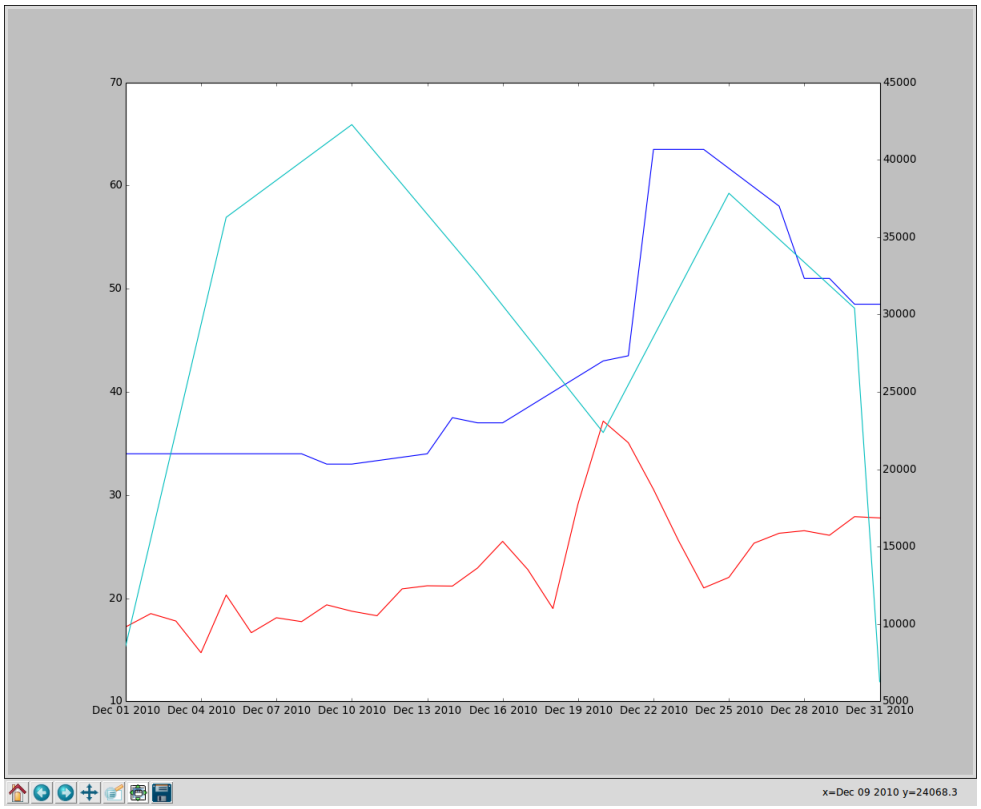

Figure 3.5: Maharashtra, Dec 2010. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival)

As we can see from the Figures 3.5, 3.6, 3.7 and 3.8, the difference between retail and wholesale went as much high as 200% in both overall Maharashtra as well as in the Mumbai. Also, as per report [12], this trend was also continued in the next month, i.e. January 2011, which can be seen in the following graph. (See figures 3.9 and 3.10)

Next hoarding news was reported in the last week July 2013. As per the Business Standard Report [3],

Onion prices in Nashik, Pune and Ahmednagar have increased to Rs 2,400 a quintal as on July 21, compared to Rs 1,500-1,800 a quintal during the corresponding period of last year. Arrival of onions in Nashik, which contributes

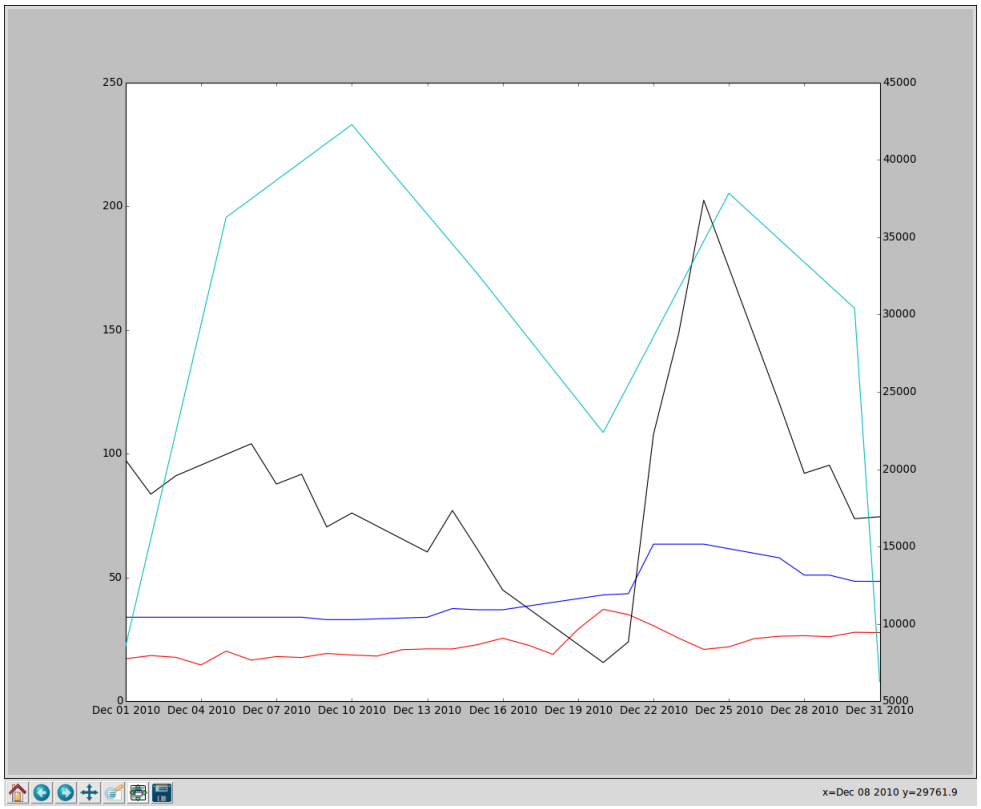

Figure 3.6: Maharashtra Dec 2010. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival, Black - Relative difference %)

35-40 per cent to the state production, has been 83,000 quintals compared with 82,000 quintals last year

Also, In the month of August, September and October of 2013, the price rise of the Onion in various parts of countries like Banglore ([9]), New Delhi ([6],[10],[8]) and Gandhinagar ([[4]) was in the news. Reports of Banglore and New Delhi has just reported the hike in the retail price. DNA report on Gandhinagar says,

Retail onion prices in the city have increased from around Rs40/kg, a month ago, to Rs70/kg on Saturday. In the wholesale market, onion prices have increased around Rs35 per kg till last week to Rs45 per kg on Saturday. The wholesale price is Rs.40 to 45 per kg. Ideally, in the retail market the price should not be more than Rs.60 per kg

Let's look at the data we have. As from the figures 3.11 and 3.12, the

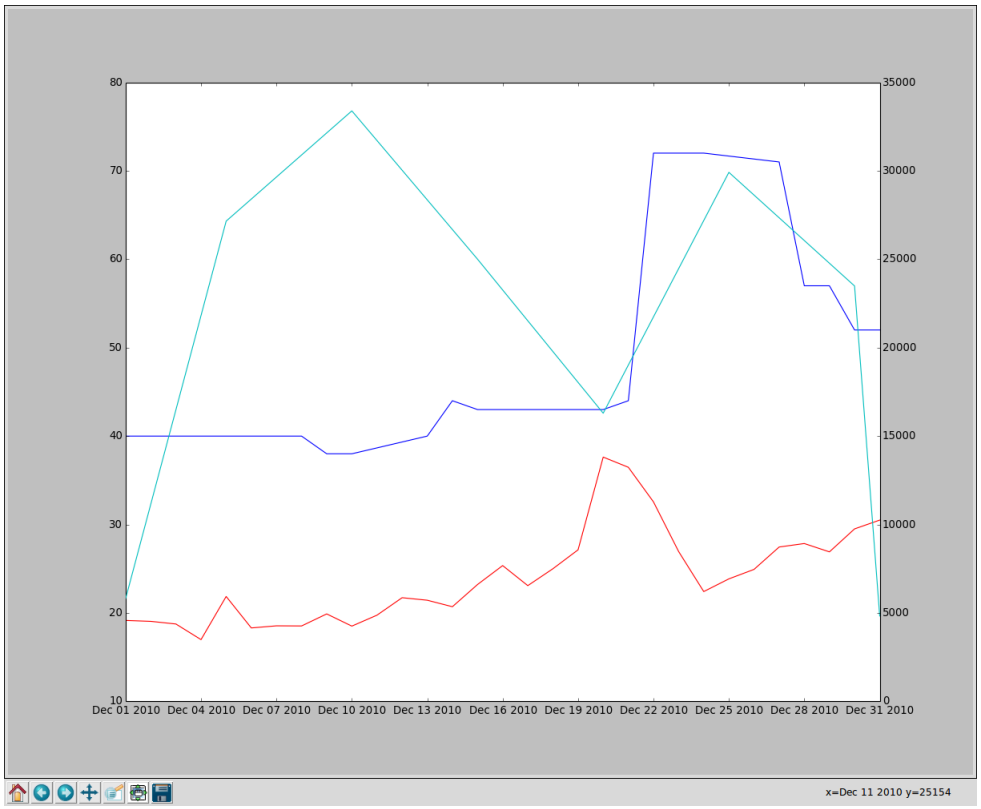

Figure 3.7: Mumbai , Dec 2010. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival)

wholesale rates around the mandis present around the Mumbai in the month of July, 2012 was about Rs. 5/Kg, but during the same time period in the 2013, the wholesale prices went from Rs. 15/Kg to Rs. 25/Kg. Although, there was decrement in the arrival by 7% as compared to July 2012, but the rise in the wholesale price is very much high.

Figures 3.13 and 3.14, states the scenario of Gujarat. There also the wholesale as well as the retail prices became suddenly almost double in the month of August 2013.

Also, in year 2014, price hike of Onions was in the news. There were reports from Mumbai [11], Kerala [7] and Hydrabad [13] which stated the hike in the prices of the onion. Let's look at the graph of the Mumbai for the year of 2014. As we can see from the graph (Figure 3.15) that for the period of July to October the wholesale price were decreasing, but still that was not

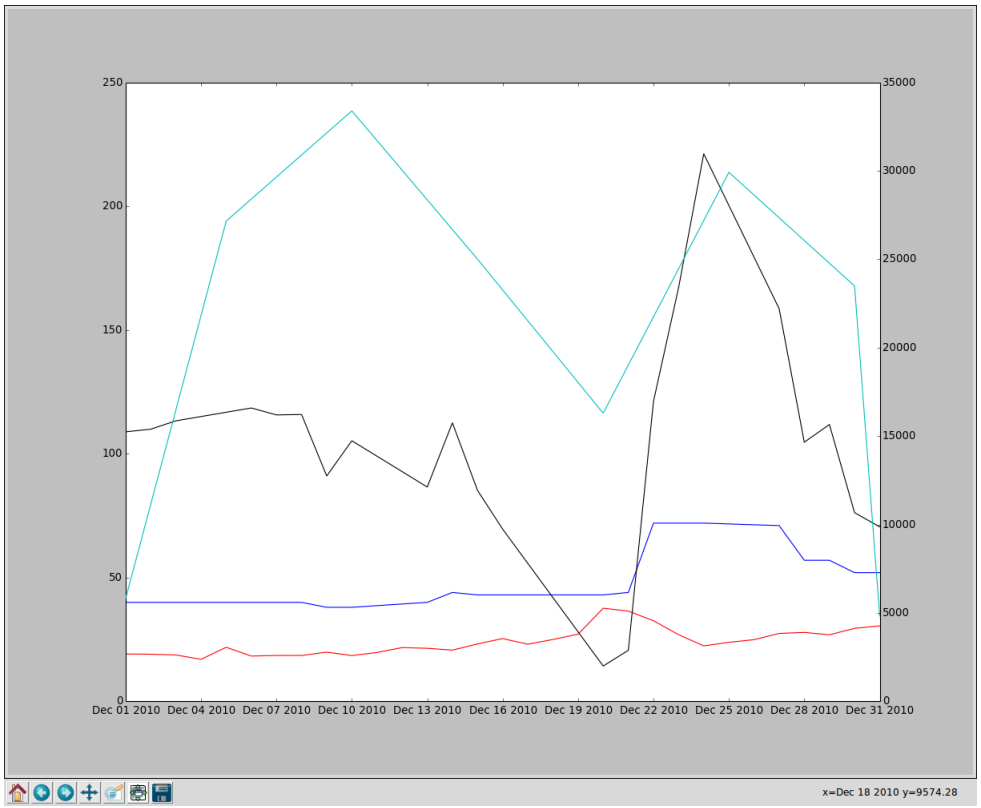

Figure 3.8: Mumbai, Dec 2010. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival Black - Relative difference %)

reflected in the retail price and instead of decreasing, the retail price kept on increasing.

3.7 Characteristics of anomaly

Hoarding of commodities in excess, results in anomaly. We segregated news articles on hoarding of onion and tried to spot some characteristics of data for anomalies.

So, major characteristics of hoarding spotted in newspapers are following:

1. Huge difference in wholesale and retail prices
2. Sudden rise in wholesale or retail price

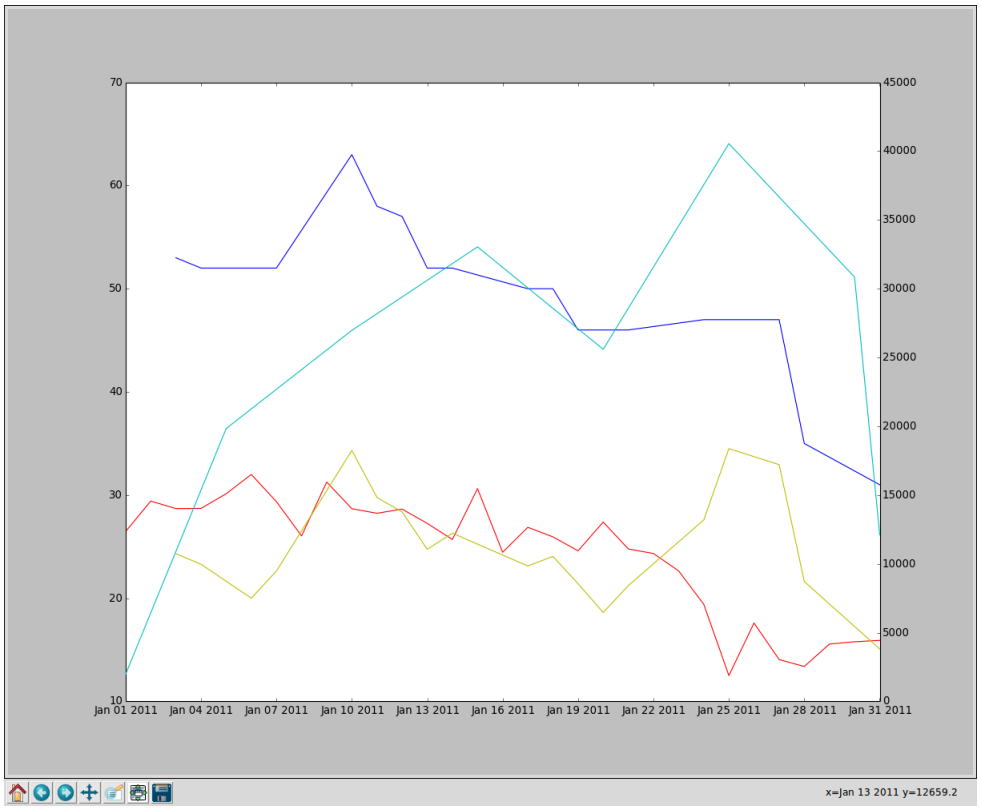

Figure 3.9: Mumbai , Jan 2011. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival)

3. Rise in wholesale prices when arrival is enough/high

We can also generalize anomaly cases as shown in table 3.1 and 3.2. The estimated arrival, wholesale and retail price data is labeled up, down and constant based on if it exceeds actual data value. Red flags are raised based on the following tables if the condition falls under the category marked with As.

W\A	↑	↔	↓
↑	-	A	A
↔	-	-	A
↓	-	-	-

Table 3.1: Anomaly Scenarios - 1

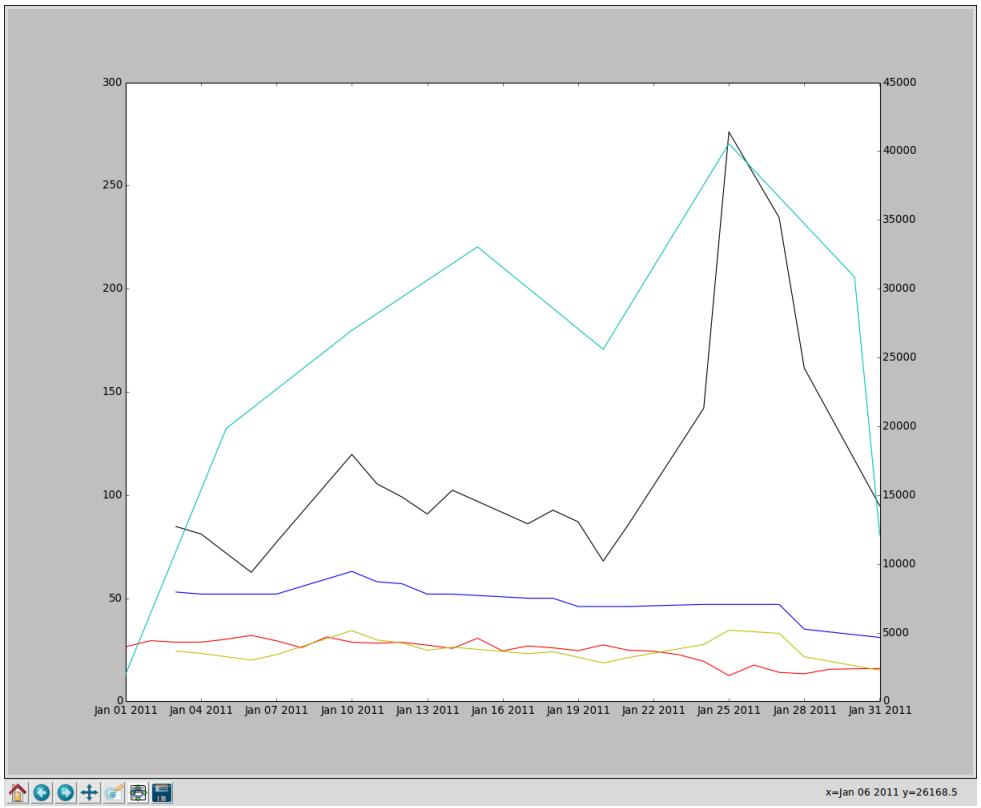

Figure 3.10: Mumbai, Jan 2011. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival Black - Relative difference %)

3.8 Hypothesis

So from the above analysis of all news reports, we conclude four hypothesis.

Wholesale price is inversely proportional to arrival in the market and we assume it is the only factor on which wholesale price is dependent. We also noticed in our literature review that most of the farmers dont hoard in India. So, from this relation, we have the following hypothesis:

W\A	↑	↔	↓
↑	A	A	-
↔	A	-	-
↓	-	-	-

Table 3.2: Anomaly Scenarios - 2

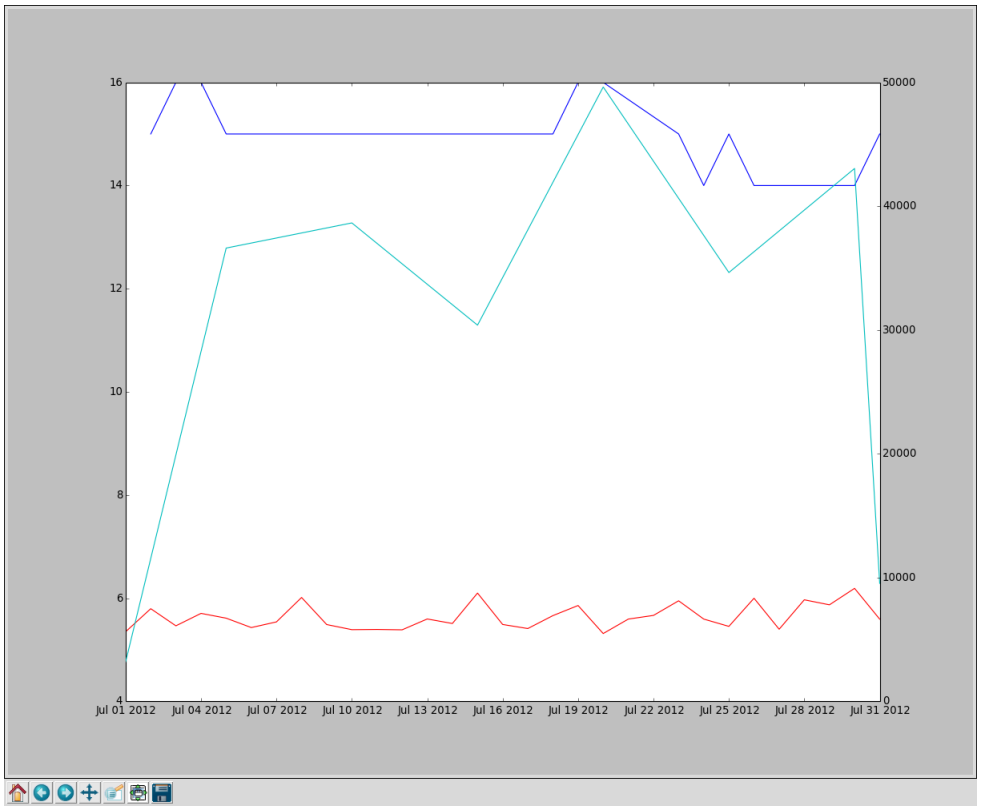

Figure 3.11: Mumbai , July 2012. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival)

H1. If there is increase in arrival pattern, there should be decrease in the wholesale price and if there is decrease in arrival pattern, there should be increase in the wholesale price considering a lag factor of 15 days.

Retailers purchase onions from the wholesale markets, mandis, etc. So, rate at retail level should be directly proportional to wholesale price in that region. We assume here that demand remains constant and there is no supply shock created because of excessive export of onion So from here we get the following hypothesis.

H2. If there is increase in the wholesale price of onion, then there will be corresponding increase in the retail price and vice-versa assuming demand remains constant and there is no supply shock created because of excessive export of onion.

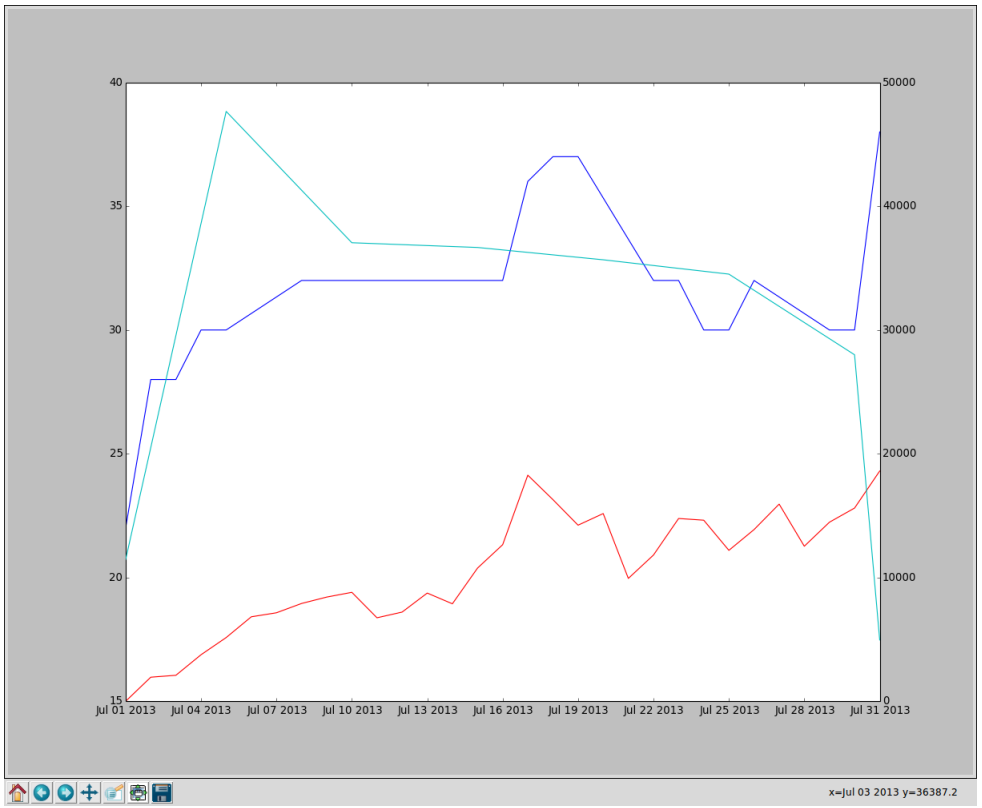

Figure 3.12: Mumbai , July 2013. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival)

Similarly, we can state that if there is no change in the arrival for some period, then wholesale may remain same and if wholesale price remains same for some period than retail price may remain same. So from that we get following two hypothesis:

H3. The deviation between arrival - wholesale and wholesale-retail should not vary much compared to values for same time in past years.

Also, one should make a note of that, even in H1 and H2, when wholesale price or retail is increasing, then it should be in considerable amount. It should not be like, there is marginal increase in wholesale price and retail price is boosting up or there is little decrement in arrival and wholesale price goes up by unacceptable level.

Also, we assume that mandis in the same region will behave in similar man-

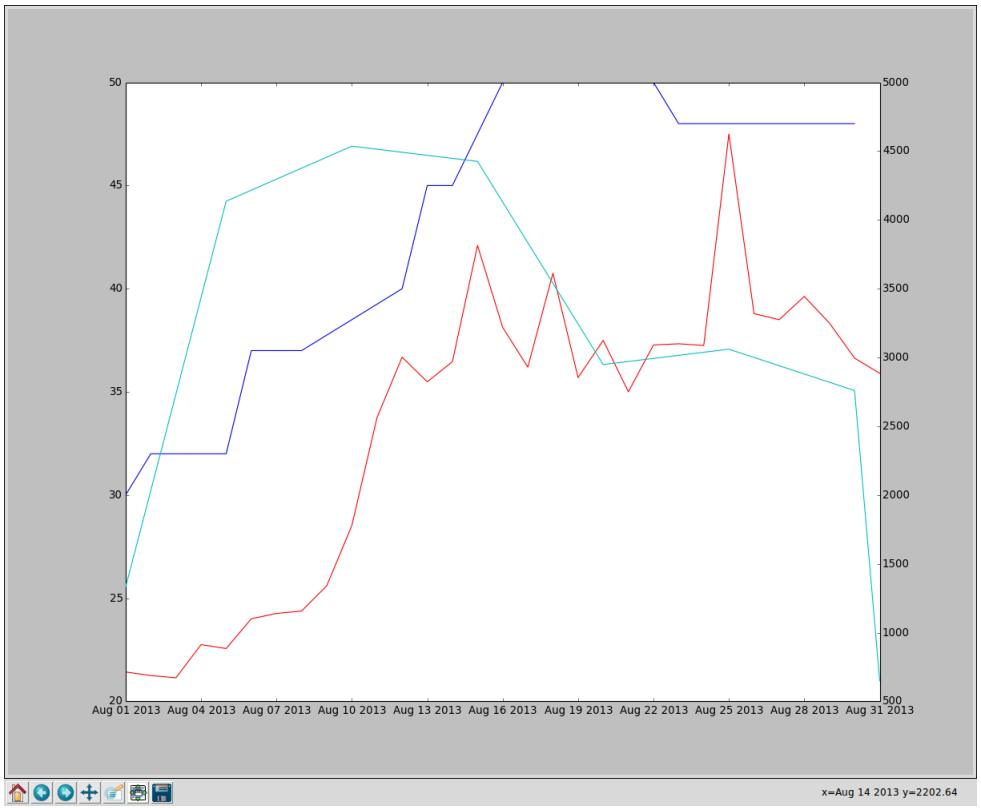

Figure 3.13: Ahmedabad , Aug 2013. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival)

ner, because production and effect of other factors will be same in one region. So, based on that we have following hypothesis:

H4. Mandis in the same region should follow the same relationship between arrival and wholesale price, as that of, taken whole region combinely.

Our attempt would be to create a library which takes all these series as an input and try to detect possibilities of above stated anomalies or highlight all the important unusual behavior of parameters seen in the input series.

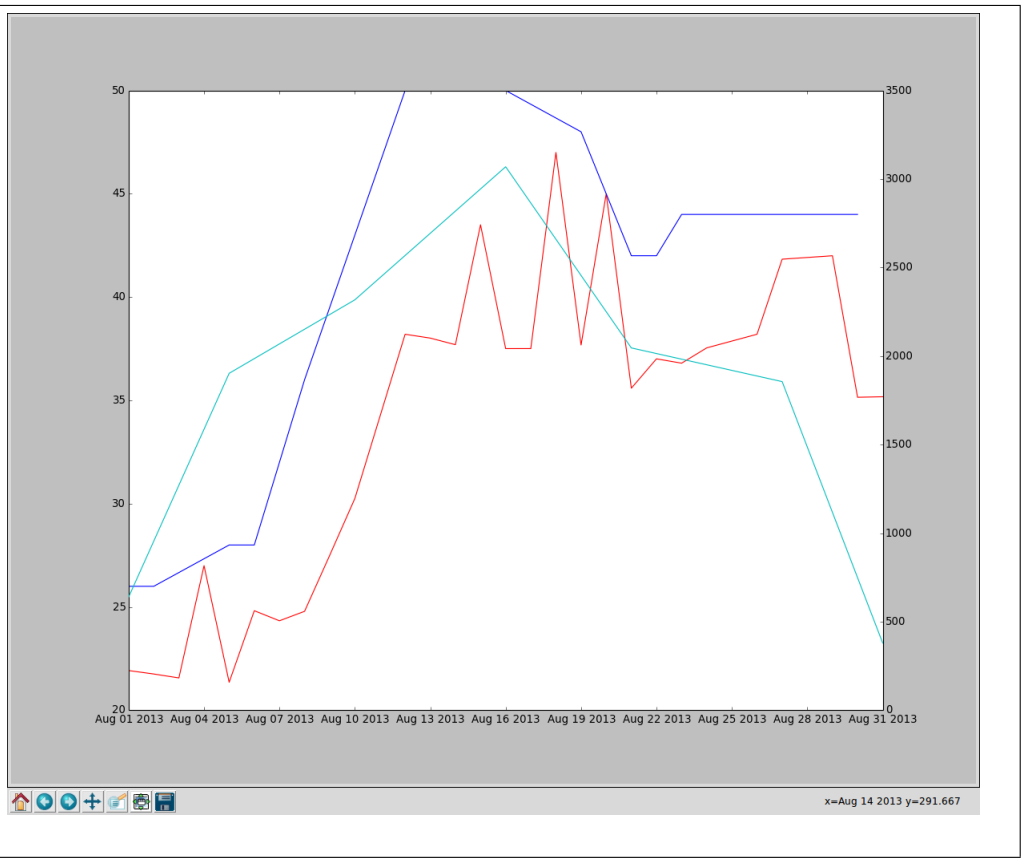

Figure 3.14: Rajkot , Aug 2013. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival)

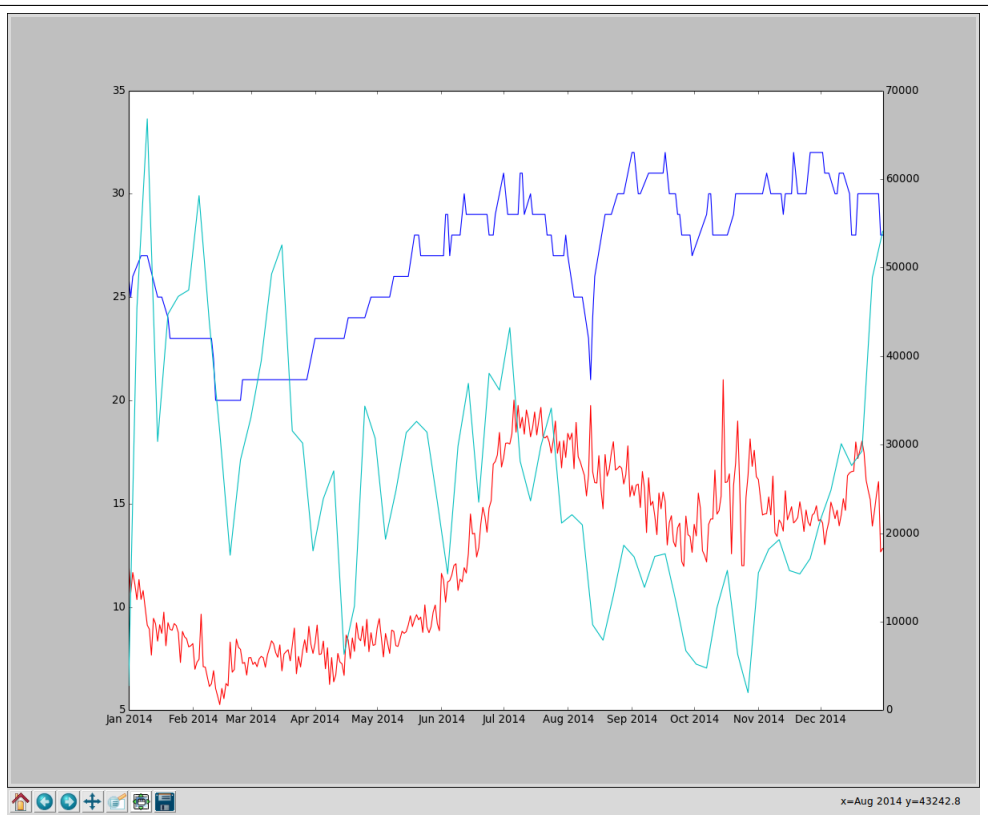

Figure 3.15: Mumbai , 2014. (Blue - Retail price, Red - Wholesale Price, Cyan - Arrival)

Chapter 4

Proposed Design and Framework

4.1 System Design

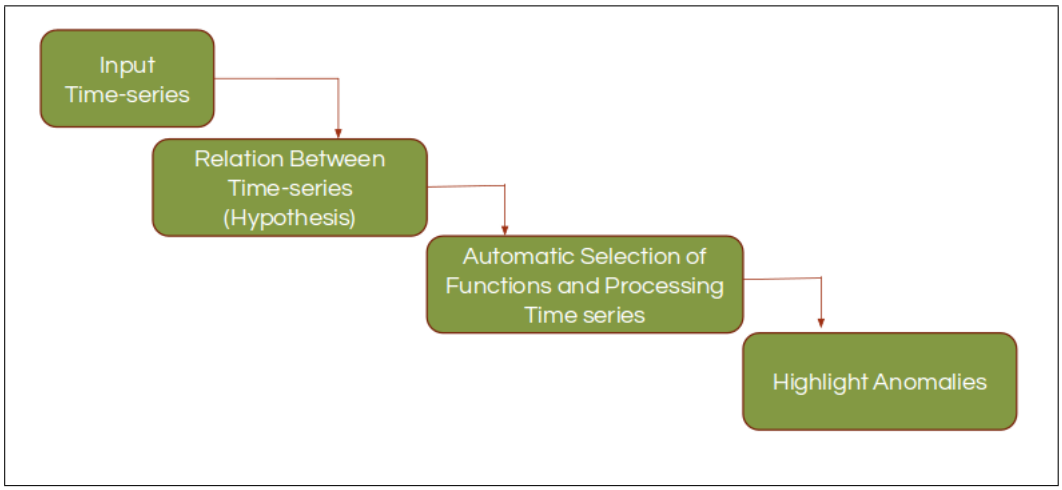

Figure 4.1: System Design

So figure 4.1 explains the overall design of the system and its modules. We have made some assumption regarding system as follows:

1. Given time series should be of the same time period.
2. Each time series should be present in different file.

4.1.1 Input Time-series

Here, give option to user to input number of time series and accordingly, provide option to select time series in CSV format. Currently, we have thought of keeping restriction that CSV file should contain only one time series. Also, ask user whether he wants to smooth time series or not. Smoothing of time

series will be done using Exponential Moving Average technique. α factor for same will be considered as

$$2/(1 + period)$$

. Where period is considered as up to how much of previous values should have effect on today's value. By default, we have kept it as 14. If user has technical knowledge then option to select period can be provided.

4.1.2 Relation Between Time-series (Hypothesis)

Normal behavior of the time series should be stated somehow, so that we can detect anomaly in the data. So, to get this input from user, we will be providing $n*n$ matrix with some default value, where 'n' corresponds to the number of time series user gave as an input. Now by filling each cell in the matrix, user can specify the relation between any 2 time series as follows:

- **1:** Positive Correlation
- **-1:** negative Correlation
- **0:** Random Relation
- **2:** User is not aware but is interested in finding relation, if exist any
- **Default:** User is not aware of any relation and also any doesnt exist according to user

If correlation exists, then we may also take input from user maximum lag factor to consider. If not provided any, then by default it will be taken as -15 to +15. We may also ask user, whether to consider only positive lag, negative lag or both types of lag.

4.1.3 Automatic Selection of Functions and Processing Time series

Now, after taking proper input, it is time to execute appropriate function for each of the provided input types.

4.1.4 Highlight Anomalies

After execution, display what tests were performed and what were their results. Detailed values will be provided if asked specifically. Different Graphs/Charts possible with the output can be generated on user demand. Provide an option to user whether he would like to see the results with some different threshold value other than the taken by library. Provide all the functionalities present in the system and ask user, if he wants to run some specific tests on the input time series.

4.2 Test Criterion for Hypothesis

In the last chapter, we stated some hypothesis. Now, here we state what tests can be performed to detect anomalies for each of the hypothesis.

4.2.1 Hypothesis 1

- Starting with granularity of year-wise, find out cross-correlation, considering various lag factor up to 15, between arrival and wholesale and check that overall it is positive or negative.
- Go on decrease the granularity. Next will be season-wise. For a particular season, how arrival and wholesale price are behaving. Point out where, this correlation becomes positive.
- Similar thing can be done for month-wise as well as fortnight data.
- Note that, correlation mentioned here is just one of the method. There can be various other methods to define the behavior of 2 time-series.

- **Slope Based Detection:** Calculate the slope of both time-series day-wise and compare. *Assumption:* Data is smoothed. Since day-wise is taken, if data is not smoothed then it may generate many spikes. So if data is not smoothed, then before applying this technique, either apply smoothing or take weekly average. If granularity of data is daywise, then smoothing can help. If data is reported once-twice weekly, then taking average weekly and then calculating slope day-wise may help.
- **Linear Regression Based Method:** We can train model using linear regression. Then find out the difference between actual and predicted value for each of the data point and plot a histogram. From these points we can try to find out points which are going very much out of the way using method like MAD test.

4.2.2 Hypothesis 2

- Starting with granularity of year-wise, find out cross-correlation, considering various lag factor up to 15, between arrival and wholesale and check that overall it is positive or negative.
- Go on decrease the granularity. Next will be season-wise. For a particular season, how arrival and wholesale price are behaving. Point out where, this correlation becomes positive.
- Similar thing can be done for month-wise as well as fortnight data.
- Note that, correlation mentioned here is just one of the method. There can be various other methods to define the behavior of 2 time-series.
- **Slope Based Detection:** Calculate the slope of both time-series day-wise and compare. *Assumption:* Data is smoothed. Since day-wise is taken, if data is not smoothed then it may generate many spikes. So if data is not smoothed, then before applying this technique, either apply smoothing or take weekly average. If granularity of data is daywise, then smoothing can help. If data is reported once-twice weekly, then taking average weekly and then calculating slope day-wise may help.

- **Linear Regression Based Method:** We can train model using linear regression. Then find out the difference between actual and predicted value for each of the data point and plot a histogram. From these points we can try to find out points which are going very much out of the way using method like MAD test.
- Spike Detection methods

4.2.3 Hypothesis 3

- One can use prediction based model like ARIMA [19] to test this hypothesis. If method mentioned in [19] is used then there is no need to align time-series into phase, as method mentioned in it take care of it. One can train the model considering some period for 8 years and can be tested on remaining 2 years. Difference in actual and predicted value is seen and one above some threshold value is reported. Also, while generating model, different window size can be considered. If some other technique is used, then one may need to align time series into common phase. That can be done using cross-correlation method, considering lag factor of around 15 days. Lag with the highest correlation value is considered.
- One can also apply graph based anomaly detection technique as mentioned in [17] to find out malicious behavior.

4.2.4 Hypothesis 4

Here, we need to combine data of multiple places into one, to find out the combined behavior and then compare it with each of the mandi. One of the technique, to combine multiple series is,

- For Arrivals: take sum from all Mandis
- For Wholesale-price: Take average

- Apart from center, if one wants to combine statewise, then for retail price also, one can take average for retail price

Other technique, to combine multiple time series is, subspace based transformation, as mentioned in [16]. Then, to compare behavior of each particular mandi time-series with the aggregated one, techniques mentioned to test hypothesis 1 or 2 can be used.

Note (Applicable to Onion Case): Results returned by H1 and H2, consists of all the time periods where arrival-wholesale price and wholesale price-retail price pairs go out of line. There may be case where, for each time of that year it may be going out of line and might not be anomaly. So, to remove such false positives, we can take intersection of results of H1 and results of H3. Similar thing can be done with H2 also, by taking intersection of results of H2 and results of H3.

Chapter 5

Conclusion

Often rise in the price of commodity are considered as a result of unseasonal rainfall, increased demand, unrealistic government policies, supply deficit etc. But supply chain deficits and natural calamities can not only be held responsible for every hike in price of commodity. Sometimes the hike in prices are man-made with the intension of earning more profits with illicit means like excess hoarding, manual creation of supply crunch etc.

These manual interventions to the supply chain of commodity for sake of earning more profit can be found by analyzing time series data of dependent factors since these have distinguishing characteristics. These characteristics can be found by various statistical and machine learning techniques.

So, for given time series first one needs to understand normal behavior before going for detection of anomalies. Time series may have periodicity, seasonality, trends or may have complete randomness. So method to detect anomaly should be designed in a such a way that functionalities covered by it does not miss any type of time series. So, here we have tried to build up a system keeping multiple functionalities for one hypothesis, such that it can cover multiple aspects of any time series.

In future, we are planning to develop this library, where given time series and relation as input, our system will provide cases of anomalies present in the time series.

Bibliography

[1] Agriculture marketing. http://agmarknet.dac.gov.in/. (Visited on 01/04/2016).

[2] fcainfoweb.nic.in/pmsver2/reports/report_menu_web.aspx. http://fcainfoweb.nic.in/PMSver2/Reports/Report_Menu_web.aspx. (Visited on 01/04/2016).

[3] Flying squads to tackle onion hoarding in maharashtra — business standard news. http://www.business-standard.com/article/markets/flying-squads-to-tackle-onion-hoarding-in-maharashtra-113072200152_1.html. (Visited on 01/05/2016).

[4] Govt wakes up from slumber to prevent onion hoarding — latest news & updates at daily news & analysis. http://www.dnaindia.com/ahmedabad/report-govt-wakes-up-from-slumber-to-prevent-onion-hoarding-1876027. (Visited on 01/05/2016).

[5] The great onion robbery: 135% mark-up from mandi to retail - times of india. http://timesofindia.indiatimes.com/india/The-great-onion-robbery-135-mark-up-from-mandi-to-retail/articleshow/7147837.cms. (Visited on 01/05/2016).

[6] Hoarding pushing onion prices up, govt finds - times of india. http://timesofindia.indiatimes.com/india/Hoarding-pushing-onion-prices-up-govt-finds/articleshow/21872501.cms. (Visited on 01/05/2016).

[7] Kerala yet to initiate action against onion hoarders - the hindu. http://www.thehindu.com/news/national/kerala/kerala-yet-to-initiate-action-against-onion-hoarders/article6189123.ece. (Visited on 01/05/2016).

[8] No onion hoarding in delhi, don't politicise price rise: Sheila:ibnlive videos. http://www.ibnlive.com/videos/politics/sheila-alerts-646704.html. (Visited on 01/05/2016).

[9] Onion crisis worsened by hoarding? http://www.ndtv.com/india-news/onion-crisis-worsened-by-hoarding-531668. (Visited on 01/05/2016).

[10] Onion price rise may be due to hoarding: Centre - times of india. http://timesofindia.indiatimes.com/india/Onion-price-rise-may-be-due-to-hoarding-Centre/articleshow/22672225.cms. (Visited on 01/05/2016).

[11] Onion price rise: Nashik farmers put blame on hoarders - timesofindia-economictimes. http://articles.economictimes.indiatimes.com/2014-06-20/news/50739208_1_lasalgaon-apmc-onion-prices-farmers. (Visited on 01/05/2016).

[12] Onion prices rise again at vashi agricultural produce marketing committee — latest news & updates at daily news & analysis. http://www.dnaindia.com/mumbai/report-onion-prices-rise-again-at-vashi-agricultural-produce-marketing- (Visited on 01/05/2016).

[13] Rains, hoarding make vegetables costlier. http://www.deccanchronicle.com/140720/nation-current-affairs/article/rains-hoarding-make-vegetables-costlier. (Visited on 01/05/2016).

[14] Rediff on the net business news: Onion crisis looms as trader-exporter nexus waits for the kill. http://www.rediff.com/business/1999/sep/21onion.htm. (Visited on 01/05/2016).

[15] Why onion prices are so high. http://www.ndtv.com/india-news/why-onion-prices-are-so-high-442629. (Visited on 01/05/2016).

[16] Deepthi Cheboli. *Anomaly detection of time series*. PhD thesis, University of Minnesota, 2010.

[17] Haibin Cheng, Pang-Ning Tan, Christopher Potter, and Steven A Klooster. Detection and characterization of anomalies in multivariate time series. In *SDM*, pages 413–424. SIAM, 2009.

[18] PG Chengappa, AV Manjunatha, Vikas Dimble, and Khalil Shah. Competitive assessment of onion markets in india. *Institute for Social and Economic Change. Competition commission of India*, 1:86, 2012.

[19] SC Hillmer, DF Larcker, and DA Schroeder. Forecasting accounting data: A multiple time series analysis. *Journal of Forecasting*, 2(4):389–404, 1983.

[20] Devesh Kapur and Milan Vaishnav. Quid pro quo: Builders, politicians, and election finance in india. *Center for Global Development Working Paper*, (276), 2011.

[21] Sandip Sukhtankar. Sweetening the deal? political connections and sugar mills in india. Technical report, Mimeo, 2008.