TEST OF ELLIOT WAVE THEORY
BY
TIME SERIES SEGMENTATION ALGORITHMS

Thesis submitted to the
Institute for Graduate Studies in the Social Sciences
in partial fulfillment of the requirements for the degree of

Master of Arts
in
Management Information Systems

by

Tamer Çağatay

Boğaziçi University

2009

Thesis Abstract

Tamer Çağatay, "Test of Elliot Wave Theory by Time Series Segmentation Algorithms"

Discovering the patterns of stock prices in a volatile and rapidly-changing market environment is a challenging problem. Elliot Wave theory, a form of technical analysis, attempts to investigate market price movements. The purpose of this study is to test the principles of Elliot Wave theory on Istanbul Stock Exchange(ISE)-National 100 index. Two time series segmentation algorithms, top-down algorithm and a modified version of bottom-up algorithm, are applied. Standard (compulsory) rules and rules expressed as the golden ratio of Fibonacci are formulated by statistical hypotheses. The existence of Elliot Wave pattern in ISE -National 100 index is tested. Hypothesis tests show that Elliot patterns exist in ISE-National 100 index, stating that Istanbul Stock Exchange is not an efficient market.

Tez Özeti

Tamer Çağatay, "Elliot Dalga Teorisi'nin Zaman Serisi Segmentasyon Algoritmaları ile Testi"

Kısa sürede ve süratle değişen piyasa koşullarında hisse senedi fiyat endeksindeki paternlerin keşfi önemli bir problemdir. Bir teknik analiz yöntemi olan Elliot dalga teorisi, hisse senedi fiyat hareketlerini araştırır. Bu çalışmanın amacı İstanbul Menkul Kıymetler Borsası (İMKB)-Ulusal 100 endeksi üzerinde Elliot Dalga teorisinin varlığını test etmektir. İki zaman serisi segmentasyon algoritması – yukarıdan aşağıya algoritması ve aşağıdan yukarıya algoritmasının uyarlanmış bir versiyonu—uygulanmıştır. Standart (zorunlu) kurallar ve Fibonacci'nin altın oranı ile ifade edilen kurallar istatistiksel hipotezlerle formule edilmiştir. İMKB-Ulusal 100 endeksinde Elliot dalga paterninin varlığı test edilmiştir. Hipotez test sonuçları IMKB'nin etkin bir piyasa olmadığını ortaya koyarak Elliot paternlerinin İMKB-Ulusal 100 endeksinde var olduğunu göstermiştir.

Acknowledgements

I would like to thank my advisor, Assist. Prof. Bertan Badur for his guidance and continuous support. I am also thankful to my jury members, Dr. Ali Tükel and Assoc. Prof. Eyüp Çetin. I would like to present my special thanks to my family and my friends for their continuous support and never ending belief.

Finally, I dedicate my thesis to all hidden heroes or heroines I met in my life unconsciously, who has even a small drop of participation to the point where I stand now.

CONTENTS

FIGURES

TABLES

CHAPTER I

INTRODUCTION

Time series data accounts for an increasingly large fraction of the world's supply of data (Ratanamahatana et al., 2005). Considering the rapidly increasing databases, there has been an explosion of concern in time series data mining (Wang et al., 2005). Time series databases are utilized in many domains such as finance, medicine, meteorology, astrophysics, etc.

Finance, in which multi dimensionality, volatility and vulnerability of data are quite high, is one of the fertile application domains for time series data mining. Availability of several sequential variables such as stock prices, exchange rates, etc. makes finance a generous disciple for time series data mining.

Due to instability and environment-sensitivity in financial markets, understanding the movement of stock prices has always been a challenging problem (Yi, 2007).

Technical analysis is a practice that argues the aptitude to project the future direction of prices through the study of past market data, based on price and volume (Edwards, 2007). The Elliott wave theory is a form of technical analysis that attempts to investigate market price activities (Powers, 2001). It is a theory of price behavior which states that the price of any given stock or commodity in a free market will travel in pre-determined wave-like advances and recessions, in a semi-cyclical fashion (Neely & Hall, 1990). Elliot asserts that prices of actively traded securities move in wave-like motions in an alternating pattern that normally involves three steps forward for every two steps back.

There are several past researches that principles of Elliot Wave theory are tested, and highly accurate results are obtained including stock price indices such as Dow Jones (consist of the most trading companies in New York Stock Exchange), NASDAQ (the National Association of Securities Dealers Automated Quotation System), London Stock Exchange and Tokyo Stock Exchange.

As an alternative to Elliot Wave theory, Efficient Market Theory asserts that prices do not follow a trend, act in a random walk and cannot be predicted. When new information arrives, stock prices adjust rapidly. Numerous competing investors, random arrivals of information to market and attempts of investors as a reaction to information arrival make price movements unpredictable.

This study aims to test Elliot Wave theory on Istanbul Stock Exchange (ISE) National 100 index by applying time series pattern search techniques. Elliot Wave pattern is searched in ISE-National 100 index as an alternative to efficient market theory. Knowledge Discovery in Databases methodology is used throughout the study. Time series segmentation algorithms are implemented. Elliot wave rules, categorized as compulsory and Fibonacci rules, are formulated as hypotheses, then parametric hypothesis tests are applied.

The study is organized as 5 chapters: Chapter I is an overview of the study. Chapter II gives background information about time series data mining algorithms and Elliot Wave theory referring the previous research of these disciplines. Chapter III states the definition of problem, description of data and the methodology that is applied. Chapter IV represents the results and findings of the study. Chapter V finalizes the study with a summary of the outcomes and yields to a conclusion.

CHAPTER II

LITERATURE REVIEW

This chapter presents background information about time series data mining and its algorithms. Elliot Wave theory is explained in details referring the previous researches.

Data mining is a process of extracting hidden patterns from large amounts of data where intelligent methods and algorithms are applied in order to mine the knowledge (Çağatay & Badur, 2007). Conventional data mining methods are designed to deal with cross sectional data, where the order of records is not that of interest. However, there are many other cases such as time series data, where sequential information can significantly enhance the mined data (Agrawal et al., 1995).

Time series databases introduce new aspects and challenges to the tasks of data-mining and knowledge discovery. These new challenges include determination of best representation of time series data, measuring representation level and detecting the patterns hidden in the sequence.

Introduction and Definition of Time Series Data

A time series is a collection of observations that are chronologically ordered (Prechter & Frost, 1998). Time-sequenced data accounts for an increasingly large fraction of the world's supply of data (Keogh et al., 2005). The nature of time series data mainly has large data size and high dimensionality (Wang, Yu, & Han, 2005). Given the exponentially growing sizes of databases, keeping millions of time-stamped data, there has been an explosion of interest in time series pattern analysis (Han & Kamber, 2001).

Time series pattern search is a set of techniques that enables related parties to analyze the market behavior (Rafiei & Mendelzon, 2000), helps find similar patterns in time series data. These techniques mainly employ a high-level-look to data points. There are several methods asserted for different application domains: The transformative approach, leaded by (Agrawal, Faloutsos, & Swami, 1993), mainly based on a discrete Fourier transform (DFT) for each time sequence and picks the first few coefficients to index their respective original sequences (Dhar, 1998) Time chains with matching coefficients are considered similar. On the other hand, the work of Agrawal et al. (1995) allows subsequence matching. Moreover, a framework based on wavelet decomposition is presented by Satchwell (2004). Chan and Fu (1999) applied discrete wavelet transform (DWT) for time series pattern matching. Keogh et al. (2005) proposed a probabilistic model based on linear segmentation of time sequence in accordance with prior knowledge for efficient representation. Xia (1997) proposed methods for efficient retrieval of entire similar series in the time series dataset.

Time series pattern search is applied in many domains such as finance, science/medicine, call centers and multimedia etc. (Agrawal, Faloutsos, & Swami, 1993). In finance, stock prices and exchange rates are some examples of time series data.

Stock prices have its own characteristics compared to other time series data such as electrocardiogram (ECG) (Rafael, 2001). Similar patterns in stock prices generally cannot be recognized by bare eye unless it has exactly the same visual pattern. These patterns are mostly similar but different in amplitude and period. Therefore, representation of hundreds of data points, which is addressed by time series pattern search, gains importance.

Time Series Data Segmentation Algorithms

There are many pattern representation methods. All of these algorithms have different application domains and different approximation levels (Last, Kandel, & Bunke, 2004).

Keogh et al. (2005) projected Piecewise Linear Representation (PLR) to embody data points in which the approximation of a time series T, of length n is referred with K straight lines, known as segments.

Piecewise Linear Representation

Piecewise Linear Representation (PLR) refers to the approximation of a time sequence T, of length n, with K straight lines. These segments are referred as segments. PLR is also known as a time series segmentation algorithm.

As PLR reduces the scale of data into a small set, it enables researchers make the storage and computation more efficient when compared to dealing with raw data.

In terms of pattern search and data mining, PLR is mainly used in the following tasks:

- Exact similarity search.
- Weighed similarity search.
- Concurrent mining of data and time series data.
- Clustering and classification algorithms.

As the description implies, PLR transforms low-level data points into a higher representation. During the transformation process, data is smoothed up until an acceptable error level is reached.

The segmentation problem can be investigated for different acceptable error calculation concepts. These concepts are as follows:

- Find the best representation using exactly K segments, given a time series T.
- Find the best representation such that maximum error level is less than the following threshold criteria:
    - Each segment does not exceed a user-defined threshold.
    - Total error of all segments does not exceed a user-defined threshold.

There are three main categories of time series segmentation algorithms (Last, Kandel, & Bunke, 2004) as:

- Sliding windows
- Top-Down Algorithm
- Bottom-up Algorithm

Additionally, reducing data points into straight lines is performed mainly by two methods:

- Linear Interpolation: The line fitted for the subsequence is simply the line connecting two points (Wang, Yu, & Han, 2005).
- Linear Regression: The line approximated for the subsequence is taken to be the best fitting line in the least squares sense (Last, Kandel, & Bunke, 2004).

Linear interpolation tends closely align the endpoint of consecutive segments, and gives a smooth look to the sequences of segments. Whereas, linear regression can produce a very disjoint look. Additionally, the accuracy of the approximating line produced by linear interpolation is inferior compared to the line produced by linear regression.

The Top-Down Algorithm

The top-down algorithm considers every possible partitioning of the time series and splitting is performed at the optimal breakpoint. Initially, the dataset is fitted to a single segment, and this segment is split into two sub-sections at the optimum point having the minimum error value. Both sub-sections are then tested if their approximation error exceeds user-defined threshold. Otherwise, the algorithm recursively split the sub-sections until each segment piece has an approximation error that deceeds specified error. The algorithm has several different names in different application domains, such as Douglas-Peucker algorithm in cartography, Ramer's algorithm in image processing, and "Iterative end-point fits" in data mining.

Pseudocode of top-down algorithm is given in Figure 1.

```
START
        GET T(s,f) and max_error;
        SET line_error = inf , i=2;
        WHILE 2 < i < f-2
                Split(T,i);
                SET error_subsegment= Calculate_error (T,i) ;
                IF Error_subsegment < line_error THEN
                        Set breakpoint = I;
                        line_error = errror_subsegment;
                ELSE
                        i=i+1;
                ENDIF
        END
        Calculate_Ts_b_error;
        SET Ts_b_error for T(1,breakpoint);
        IF Tib_error >max_error THEN
                SEND T(1,breakpoint);
                GET T(s,f) and max_error;
ELSE
                SET Tb_f_error for T(breakpoint+1,f)= Calculate_Tb_f_error;
                IF Tb_f_error > max_error THEN
                        GOTO START WITH
                        SEND T(breakpoint+1,f);
                ELSE TERMINATE
                ENDIF
ENDIF
```

Figure 1. Pseudocode of top-down algorithm

The flowchart of algorithm is given in Figure 2.

Figure 2. Flowchart of top-down algorithm

.

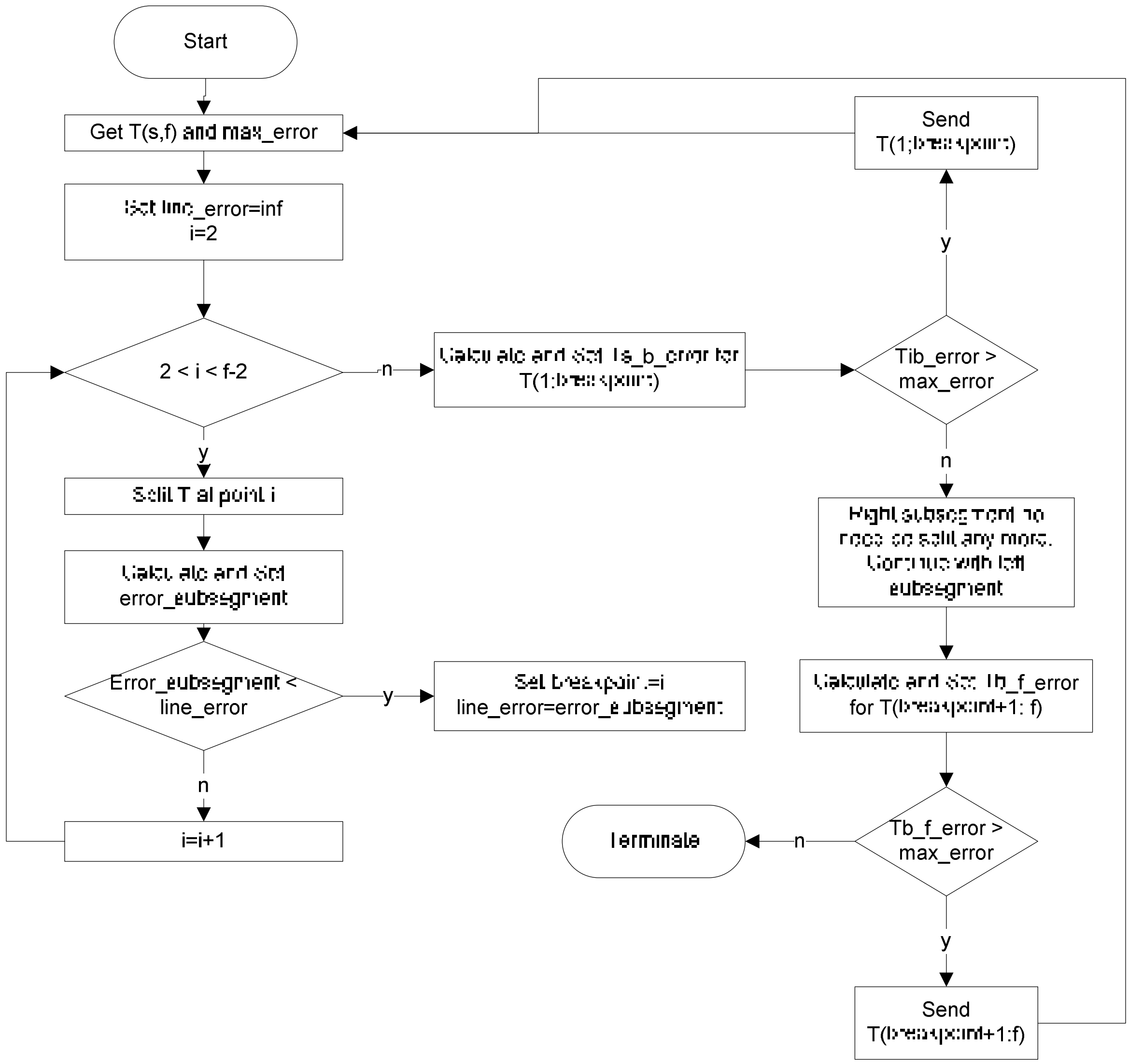
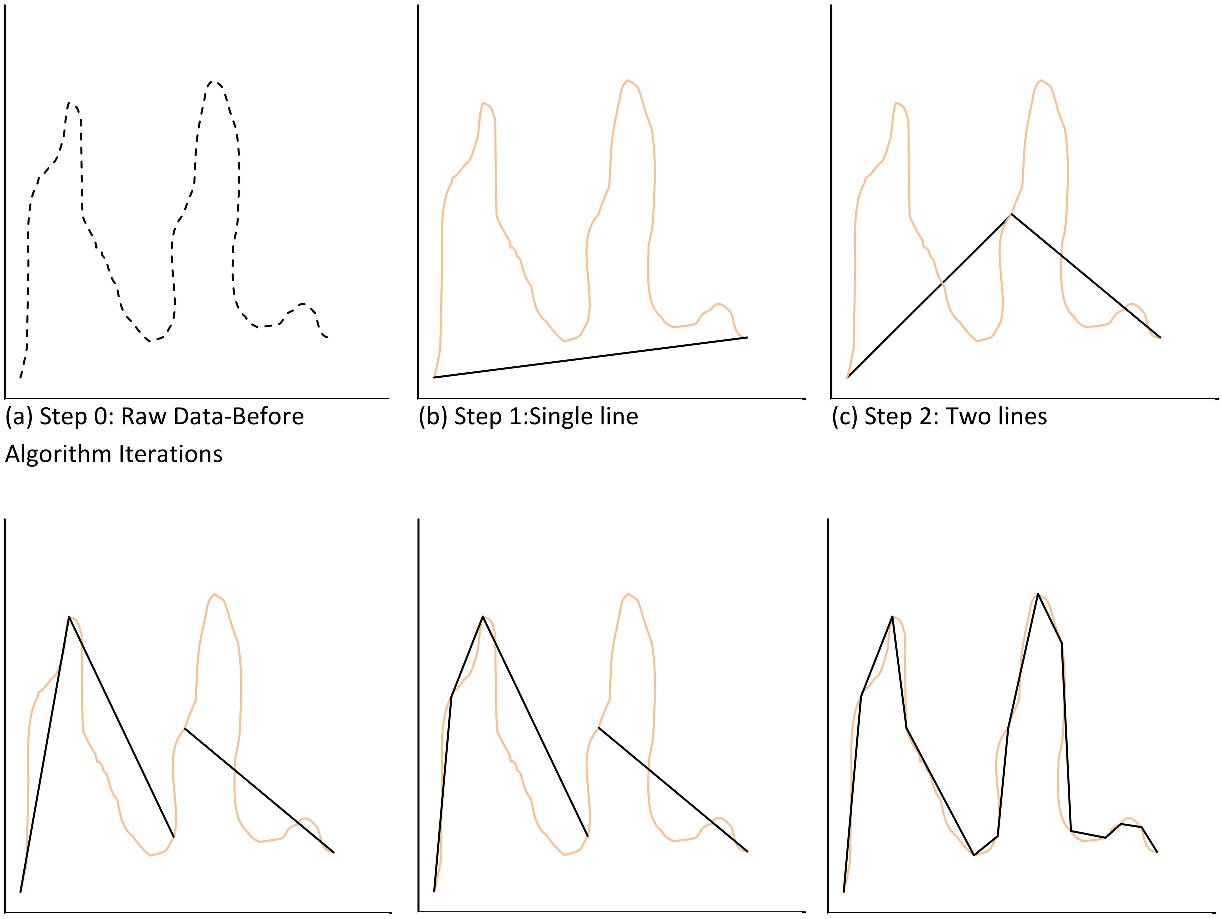


Figure 2. Flowchart of top-down algorithm

The algorithm is applied to support a framework for mining sequence databases at different abstraction levels (Li et al., 1998). Shatkay and Zdonik (1996) proposed it to test fairly accurate queries in time series databases. Park et al. (1999) introduced an adjustment where they first carry out a scan over the whole dataset marking each peak and valley. Kargupta et al., (1999) uses the top-down algorithm to prove the synchronized mining of text and time series. Moreover Smyth and Ge (2001) use the algorithm to produce a representation that can support a Hidden Markov Model approach to both change point detection and pattern matching.

As the algorithm proceeds, the change in segments can be seen in Figure 3 for a hypothetic dataset.



(a) Step 0: Raw Data-Before Algorithm Iterations

(b) Step 1:Single line

(c) Step 2: Two lines

(d) Step 3: Three lines

(e) Step 4: Right line split into two

(f) Step n: End of algorithm

Figure 3. Fitted segments as top-down algorithm proceeds

<u>The Bottom-Up Approach</u>

The bottom-up algorithm is the complement to the top-down algorithm. The algorithm starts by generating the finest possible approximation of the time series, so that n/2 sequences are used to approximate the n length time series. Then, the cost of combining each pair of consecutive segments is calculated, and the algorithm begins to iteratively unite the smallest cost pair until a stopping criteria is met.

This algorithm has been used in different domains such as computer graphics, medicine/science and finance. Two and three-dimensional analogues of this algorithm are common in the field of computer graphics where they are called decimation methods (Hecbert & Garland, 1997). In data mining, the algorithm is widely used in order to support most of the data mining tasks in time series data (Keogh et al., 2004). In addition, this algorithm has usage in medicine to provide high level representation of medical pattern matching systems (Hunter & McIntosh, 1999).

The pseudocode of bottom-up algorithm is given in Figure 4.

```
START
GET T(s,f)
GET max_error
SET Seg_Array = Null
WHILE i<f
        Create Segment T(i,i+1);
        SET Seg_Array = Add (Seg_Array,Segment T(i,i+1));
        i=i+2;
END;
WHILE k<length(Seg_Array)-1
        Merge_cost(k) = Calculate_Error(merge(Seg_Array(k),Seg_Array(k+1)));
        k=k+1;
END;
WHILE Min(merge_cost)<max_error
        SET p = min(merge_cost);
        SET Seg_Array(p) = merge(Seg_Array(p),Seg_Array(p+1));
        DELETE Seg_Array(p+1);
        UPDATE Merge_cost(p) =
Calculate_Error(merge(Seg_Array(p),Seg_Array(p+1)));
        UPDATE Merge_cost(p-1) = Calculate_Error(merge(Seg_Array(p-
1),Seg_Array(p)));
END;
TERMINATE;
```

Figure 4. Pseudocode of bottom-up algorithm

The flowchart of the algorithm is given in Figure 5.

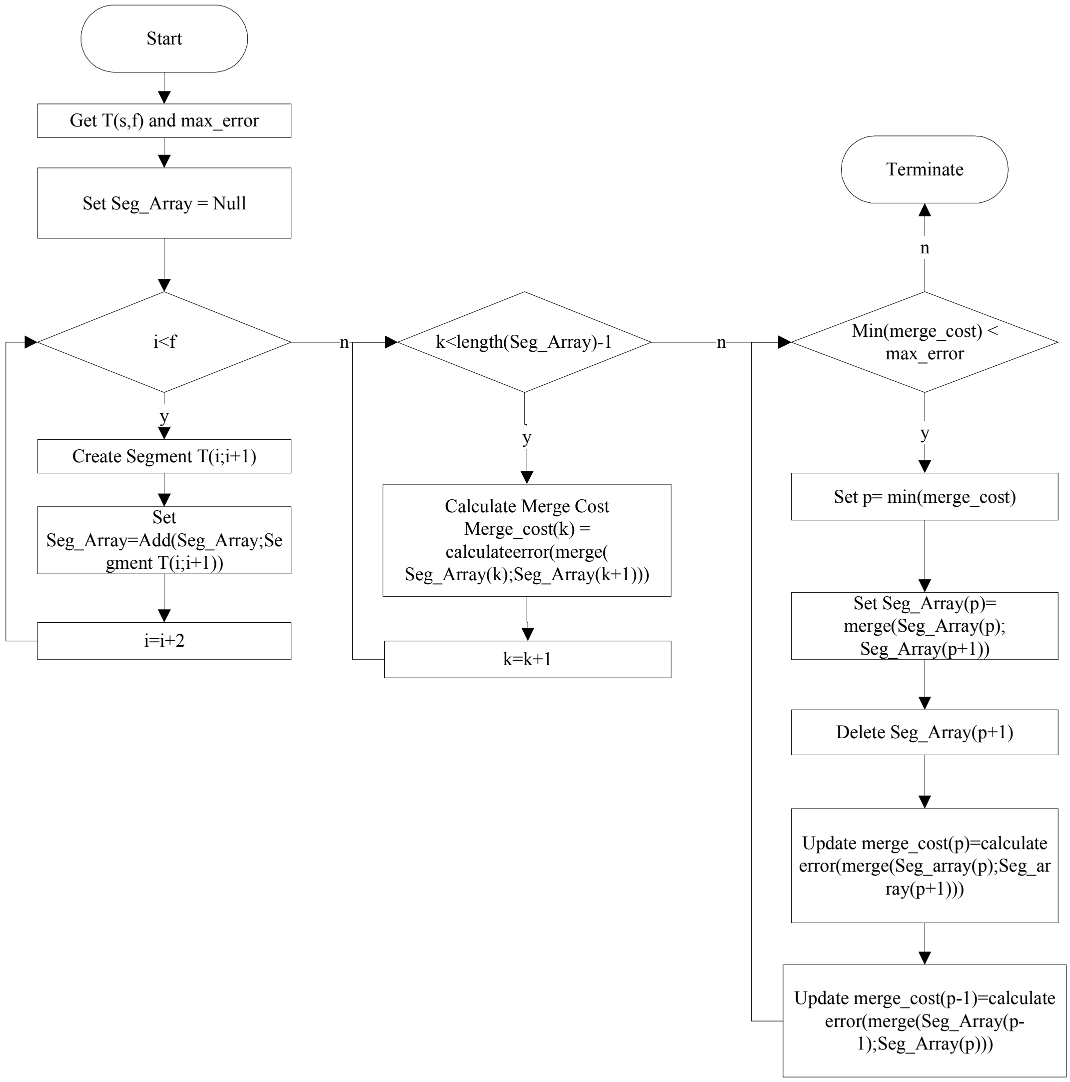


Figure 5. Flowchart of bottom-up algorithm

The comparison of top-down and bottom-up algorithm is as follows:

- The performance of both algorithms depends on the specified error threshold. If error threshold is zero than, there is no significant difference between the algorithms.
- Top-down algorithm works best when split point is in the middle of the lines. When the breakdown points are at the end of lines, then the algorithm should be stopped after a user-defined number of iterations.

- Bottom-up algorithm works best when merged segments have same length. On the other hand, when merged segments consist of too long and too short sub-lines, then approximation error is higher.

Elliot's Wave Theory

Definition

Technical analysis is a practice that claims the ability to forecast the future direction of prices through the study of past market data, primarily price and volume (Person, 2004). The Elliott wave theory is a form of technical analysis that attempts to analyze market price movements (Powers, 2001). It is a theory of price behavior which states that the price of any given stock or commodity in a free market will travel in pre-determined wave-like advances and recessions, in a semi-cyclical fashion (Neely & Hall, 1990). Elliot asserts that prices of actively traded securities move in wave-like motions in an alternating pattern that normally involves three steps forward for every two steps back. Frost and Prechter (1978) define the theory as follows:

"The 'Elliot Wave Principle' is Ralph Nelson Elliot's discovery that social, or crowd, behavior trends and reverses in recognizable patterns."

Additionally, Elliot Wave Theory posits that collective investor psychology (or crowd psychology) moves from optimism to pessimism and back again, and these swings create patterns, as evidenced in the price movements of a market at every degree of trend (Droke, 2000).

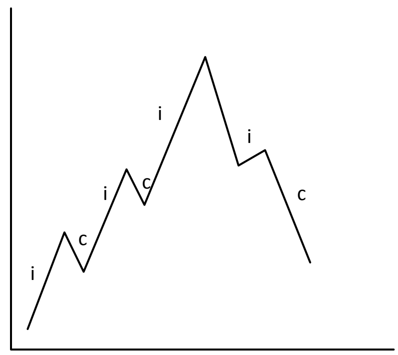


Figure 6. An Elliot wave

Elliott wave theory contends that stock market price movements take place in repetitive cycles of similar structure. Every cycle is made up of two overarching movements. There is a wave in the direction of the primary trend, which is called an impulsive wave. This is followed by a wave in the opposite direction, which is called a corrective wave. The main wave in the direction of the primary trend has five smaller waves, and the main corrective wave has three smaller waves. Thus there are five waves in the direction of the primary trend, and these are followed by a three-wave correction in the opposite direction. This eight-wave or five-three cycle occurs repetitively. The relative lengths of the waves vary, but the basic eight-wave structure remains invariant. Each of the eight waves of a cycle in turn subdivides into smaller waves, which can be broken into even smaller waves.

A basic Elliot wave is given in Figure 6. Upward movements are labeled as impulse, and downward movements are labeled as correction. Additionally impulses are named with numbers from 1 to 5, but the correction movements are named as a, b and c.

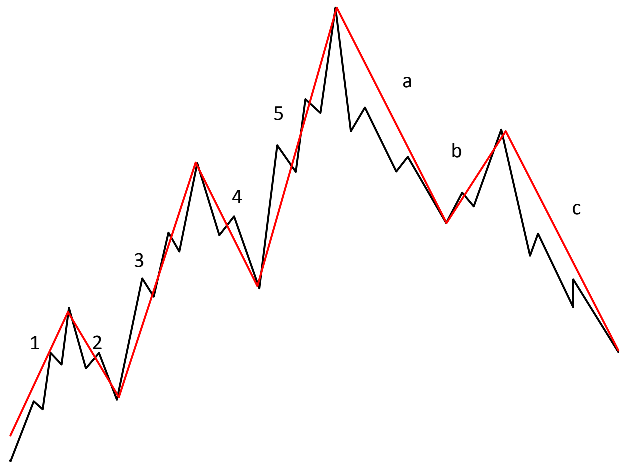


Figure 7. Elliot wave in different degrees

Figure 7 displays an Elliot Wave that is composed of sub-waves in a different time granularity. Each sub-wave follows an Elliot pattern, moreover when they are concatenated; they form a larger Elliot Wave with impulse and corrective segments. These kinds of Elliot Waves with sub-waves are called “Multi Fractal Elliott Waves” (Neely & Hall, 1990).

These sub-waves show the degree of an Elliot wave and have a classification convention. The standard Elliot Wave classification is as follows:

- Grand Supercycle: Time range from multi-decade to multi-century.
- Supercycle: Time range from a few years to a few decades.
- Cycle: Time range from one year to a few years.
- Primary: Time range from a few months to a couple of years.
- Intermediate: Time range from weeks to months.
- Minor: Only weeks.
- Minute: Days.
- Minuette: Minutes.

The naming notation for each degree is different than others. Table 1 displays naming conventions of Elliot Waves in different degrees:

Table 1. Naming Convention of Elliot Wave in Different Degrees

Wave Degree	Impulse	Correction
Supercycle	(I) (II) (III) (IV) (V)	(A) (B) (C)
Cycle	I II III IV V	A B C
Primary	(1) (2) (3) (4) (5)	(A) (B) (C)
Intermediate	(1) (2) (3) (4) (5)	(a) (b) (c)
Minot	1 2 3 4 5	A B C
Minute	i ii iii iv v	a b c
Minuette	1 2 3 4 5	a b c

Elliot Wave theory is a contradiction to Efficient Market Theory, where an “efficient” market is defined as a market where there are numerous profit-maximizing and actively competing participants, with each trying to analyze future market values of individual securities, and where important current information is almost freely available to all participants (Fama, 1970). The Efficient Market hypothesis states that the market rapidly adjusts the new information and prices of stocks reflect all of the available information.

EHM has been a widely accepted theory which claims that prices are defined in a random walk procedure, making price behavior completely unpredictable. In contrast to Elliot Wave theory, EMH asserts that the changes in prices of stock are expected to follow a random walk (Fama, 1970). Additionally, an efficient market contradicts all forms of analysis as it is impossible to beat the market with technical, fundamental or time series analysis because there are considered to have no better performance than random guessing (Fama, 1970).

Fibonacci Numbers and Golden Ratio

Elliott's market model relies heavily on analyzing price charts. Previous studies mainly focused on developing price moves to distinguish the waves and wave structures, and discern the value of price in the following time period, thus the application of the wave principle is a form of pattern recognition (Mazza, 2007).

The structures Elliott described also meet the common definition of a fractal (self-similar patterns appearing at every degree of trend) (Prechter & Frost, 1998). Elliott wave practitioners claims that just as naturally-occurring fractals often expand and grow more complex over time, the model shows that collective human psychology develops in natural patterns, via buying and selling decisions reflected in market prices: "It's as though we are somehow programmed by mathematics. Seashell, galaxy, snowflake or human: we're all bound by the same order." (Fishcher, 1941).

The cycle on which Elliot positioned his theory is basically discovered by Leonardo Fibonacci, and named as Fibonacci numbers or Fibonacci sequences (Fishcher, 1941).

Initiation point of Fibonacci numbers is explained by a rabbit story:

“The problem was presented in terms of the reproductive capabilities of rabbits – namely, how many pairs of rabbits could one pair produce in a year? The first pair are allowed to reproduce in the first month, but subsequent pairs can only reproduce from their second month onwards. Each birth consists of two rabbits. Assuming that none of the rabbits dies, then a pair is born during the first month, so there are two pairs. During the second month, the first pair reproduces, creating another pair. During the third month, both the original pair and the first-born pair have produced new pairs. Consequently, there are three adult pairs and two young pairs.” (Plummer, 2008)

If the analysis is continued, the results conveys to a sequence of numbers, known as Fibonacci numbers: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144 and so on. Table 2 displays the sequence.

Table 2. Fibonacci Sequence

Month	Adult Pairs	Young Pairs	Total
1	1	1	2
2	2	1	3
3	3	2	5
4	5	3	8
5	8	5	13
6	13	8	21
7	21	13	34
8	34	21	55
9	55	34	89
10	89	55	144
11	144	89	233
12	233	144	377

The Fibonacci numbers are crucial because they are of interest almost every domain where a recursive cycle is in focus such as mathematics, chemistry, electronics, medicine, psychology and finance (Plummer, 2008).

There are three important properties of the sequence. The first is that each term in the sequence is the sum of the two terms that immediately preceds it. Such sequences, in which every term can be represented as a linear combination of preceding terms, are called recursive sequences (Plummer, 2008).

$$F_n = F_{n-1} + F_{n-2}$$

where F_n is the nth term in Fibonacci series.

The second important feature is that each term in the sequence, when divided by the following term, approximates the ratio 0.618, which is called as the golden ratio of Fibonacci. The divergence from 0.618 is much greater for earlier values when compared to later elements of sequence. Additionally, the ratio of each term in the sequence, divided by the previous term, approximates 1.618.

The third feature of the sequence is that alternative combinations of terms in the sequence results in constant ratios. An example of a Fibonacci ratio is given as:

$$\frac{t_n}{t_{n+2}} = \frac{t_{n+2}}{t_{n+4}} = \frac{t_{n+4}}{t_{n+6}} = \frac{t_{n+6}}{t_{n+8}} = \cdots$$

where t_n is the nth element of Fibonacci sequence.

There are several ratios that can be derived from the Fibonacci sequence. These ratios are related in various ways. However, two primary ratios are 0.618 and 1.618.

Elliot states about Fibonacci series that "All human activities have three distinctive features; pattern, time and ratio, all of which observer the Fibonaci summation series. … and later I found that the basis of my discoveries was a law of Nature known to the designers of the Great Pyramid 'Gizeh' which may have been constructed 5000 years ago…" in his writings (Hill, 1979).

Elliot Wave Rules

An Elliot wave is defined as five impulsive movement and three corrective movements in the opposite direction. However, not every wave carrying these features is qualified as an Elliot wave. Initiated from the Fibonacci numbers, there are some rules that a wave should obey to be identified as an Elliot wave. Examples of an Elliot wave and non-Elliot waves are given in Figure 8.

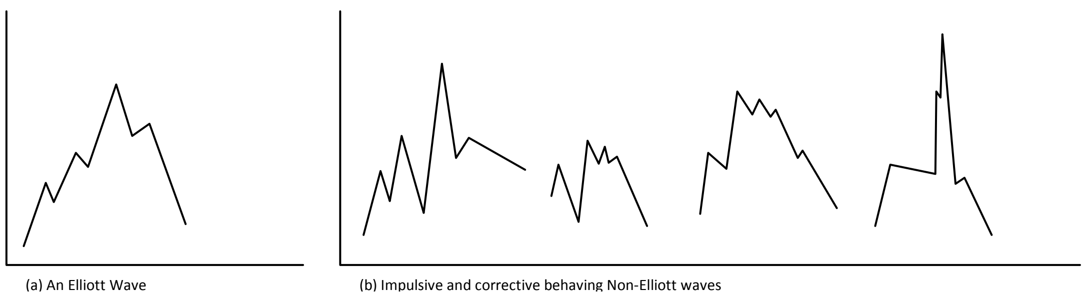


Figure 8. Examples of Elliot waves and non-Elliot waves

Elliot wave rules are mainly categorized in two distinctive groups:

- Intuitive rules: These rules do not generally composed of a numerical measure and are always in the form of a comparison of wave segments. Most of the time they overlap and support each other (Prechter & Frost, 1998). There are basically a few intuitive rules, but many derived rules asserted in different researches such as Neely and Hall (1990) and Prechter and Frost (1998). These rules are less strict then Fibonacci rules.

- Fibonacci rules: These rules are constructed over the Fibonacci numbers. Using the golden ratio and other Fibonacci ratios, each wave segment is compared with exact approximations. As the boundaries of comparison are strictly drawn, waves are harder to support these rules.

Intuitive rules are grouped in two categories as:

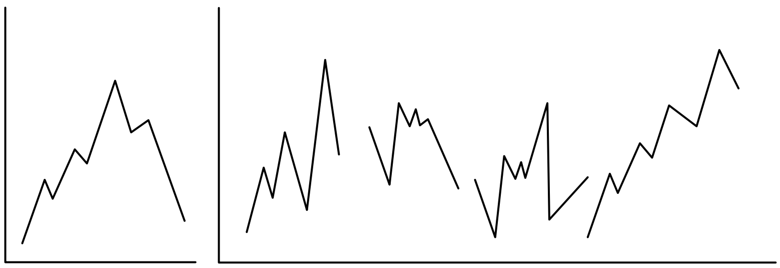
- Compulsory rules
- Derived or guiding rules

<u>Compulsory Elliot Rules</u>

These rules are asserted by Elliot, and should be satisfied by the waves.

RULE 1: A cycle is never complete unless it is composed of a fully 8 movements (Droke, 2000).

The definition of an Elliot wave implies that there should be exactly 8 segments in a wave, and this rules guarantees 8-stepped wave structure. Unless this rule is satisfied, other rules are not required to verify. Additionally, the wave should be in the form of a 5+3 wave structure. Wave examples that violate complete cycle rule are given in Figure 9.



(a) An Elliott Wave with 8 segments (b) Incomplete or non 5-3 pattern waves (non-

Figure 9. Examples of incomplete waves

RULE 2: An upward movement should be followed by a downward movement (Droke, 2000).

This rule is the main supporter of up-down movement principle of the theory. It is formed around the idea that every action is responded by a reaction. Thus, the price should be in the form of a zigzag independent of its degree. Wave examples that violate complete cycle rule is given in Figure 10.

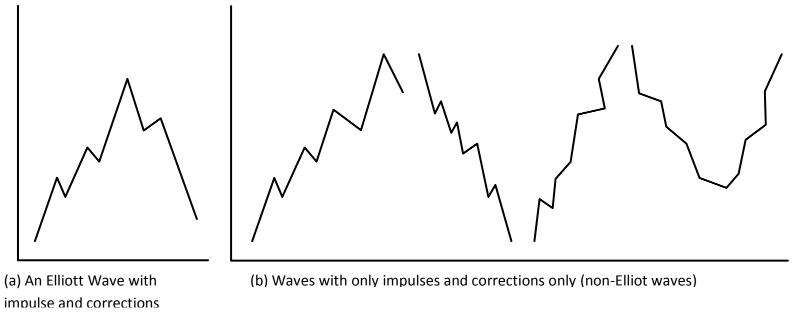


Figure 10. Waves with only impulses or corrections

RULE 3: 3rd Segment should be the longest segment at duration.

The second impulse of the wave is observed to affect more than the other price movements. So the duration of increase in a bull market should be the longest of all others in the wave. There several variations of this rule are given as follows:

- Third segment should be longer than second segment (Neely & Hall, 1990).
- Fourth segment should be shorter than third segment (Droke, 2000).
- Third segment should never be the shortest of all (Droke, 2000).

RULE 4: 2nd segment cannot retrace 1st segment at price.

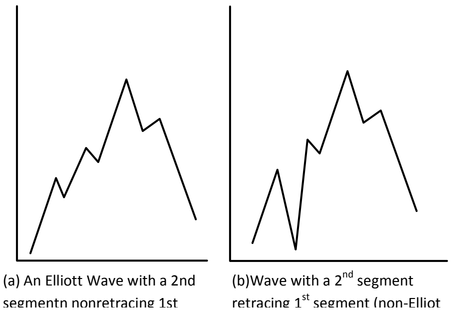


Figure 11. An example wave violating rule 4

The first response to an action is never as strong as the action itself, so corrective segment cannot put back first segment at price. An example can be seen in Figure 11.

RULE 5: 4th segment cannot trace 3rd segment at price

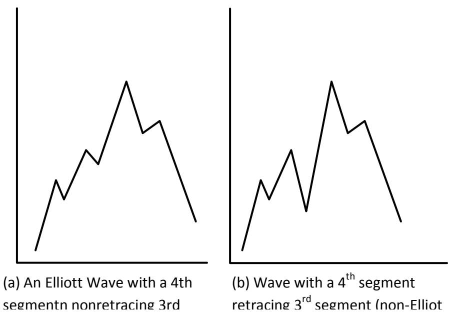


Figure 12. An example wave violating rule 5

Similar to Rule 4, 4th segment cannot put back 3rd segment in price. This is explained as a minor reaction to the previous action immediately. Then the market gets ready to response in A, B, C segments (Neely & Hall, 1990). An example can be observed in Figure 12.

RULE 6: 1st and 5th segments should be same in duration

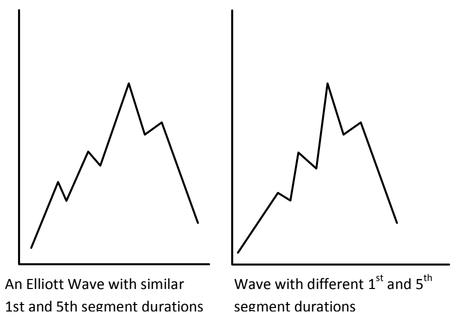
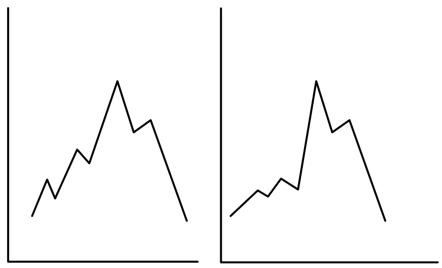


Figure 13. An example wave violating rule 6

As the main impulse of the wave is the 3rd segment, 5th segment is expected to be as short as 1st segment. In this rule, constancy is evaluated as similar or slightly different (Teseo, 2001). A wave example that violate complete cycle rule is given in Figure 13.

RULE 7: 1st and 5th segments should be same in price.



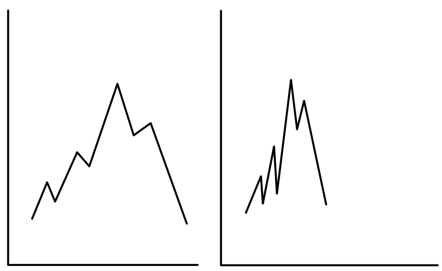
(a) An Elliott Wave with similar 1st and 5th segment

(b) Wave with different 1st and 5th segment prices

Figure 14. An example wave violating rule 7

The height of the 1st and 5th segments should be similar or slightly different from each other. An example is given in Figure 14.

RULE 8: If 2nd segment is a sharp decrease, 4th segment should be a slow decrease, or vice versa.



(a) An Elliott Wave with similar 2nd and 4th different

(b) Wave with sharp corrections in 2nd and 4th segments

Figure 15. An example wave violating rule 8

This rule is explained in the following way: If the first reaction to first action is a strong one, the market behaves more softly in the second time, or vice versa. Wave examples that violate complete cycle rule is given in Figure 15.

RULE 9: 3rd segment should climb up 1st segment.

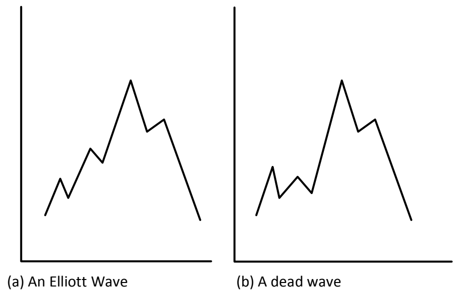


Figure 16. An example wave violating rule 9

After the second corrective movement, 3^{rd} segment should trace over the 1^{st} segment. Otherwise, wave is assumed to be a dead wave (Person, 2004). Figure 16 displays an example wave that disobeys rule 9.

RULE 10: 4th segment cannot climb down the 1st segment

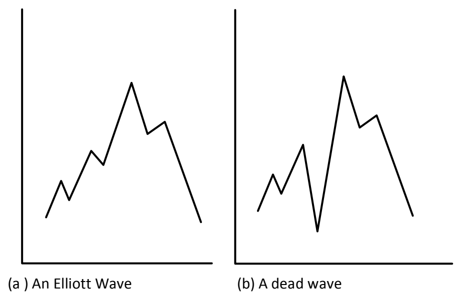


Figure 17. An example wave violating rule 10

As the impulsive behavior requires, 4^{th} corrective segment cannot put back 1st segment. Otherwise, the wave is assumed to be a dead wave. A dead wave is given in Figure 17.

<u>Derived or Guiding rules</u>

Many other rules for almost every exceptional case in price movements are derived in several studies (Frost & Prechter, 1978). As these rules are mostly case specific, it is easy to find traces of these derived rules in the compulsory rules.

Furthermore, derived rules are only the supporters of compulsory rules when they are incapable of handling the scenarios that have been rarely met. In order to apply these rules, a wave does need to be identified as an Elliot wave. Rules can be applied to any wave that has the pattern of 5-3 structure (Neely & Hall, 1990).

Some of the derived and guiding rules asserted in the literature are as follows:

- 2nd segment is a sharp correction (a clarified version of a compulsory rule).
- 4th segment is a slow decrease.
- 1st segment is the longest, 5th segment is the shortest of all.
- A segment should be similar to C segment.
- 5th segment affects shorter that 1st segment.
- 5th segment affects shorter that 3rd segment.
- C segment affects sorter than A segment.
- 2nd segment should retrace the 4th segment of the previous segment.
- 5th segment affects shorter than 4th segment.
- 3rd segment is sharper than 5th segment.

Fibonacci Rules

Fibonacci rules are the formulated versions of the compulsory or derived rules. The formulation is completely based on the Fibonacci numbers and golden ratios. Mostly, several ratios are used together to define the boundaries of the rules. When compared to compulsory rules, Fibonacci rules are stricter, and hard to satisfy. Some examples of guiding Fibonacci rules are listed as follows:

- 3rd segment should be %161, %261 of the 1st segment at price.
- 5th segment should climb up %61, %100 or %161 of 1st segments end point
- 2nd segment should climb down %61 of the 1st segment at price

- 4th segment should climb down %61 of the 3rd segment at price.
- 5th segment moves at least % 61 of the 4th segment at duration.
- 5th segment should be at least % 61 of the 4th segment at price.

In addition to the compulsory Fibonacci rules mention previously, there are guiding Fibonacci rules as well. Some of the guiding rules are as follows:

- 4^{th} segment is mostly retraces % 38.2 of 3^{rd} segment.
- 5^{th} segment affects more than a total of the 1^{st}, 2^{nd} and 3^{rd} segment.
- 5^{th} segment should be at least %38.2 of 1^{st} segment.
- 4^{th} segment climbs down to the beginning point of 1^{st} segment.

CHAPTER III

PROBLEM DEFINITION
AND
METHODOLOGY

This chapter states the definition of the problem initially. Then description of data is given with summary statistics. Steps of KDD methodology are explained in detail. Finally Elliot wave rules formulated as hypothesis are presented.

Problem Definition

Time-sequenced data accounts for an increasingly large fraction of the world's supply of data (Keogh et al., 2005). Given the exponentially growing sites of databases, there has been an explosion of interest in time series data mining (Wang et al., 2005)Pattern search is one the data mining techniques that enables to extract mostly repeating pattern within a series of data points.

Pattern search has several application domains, including finance, medicine/science, entertainment, etc. In finance, multidimensionality due to interaction between stock prices, exchange rates and inflation rates, volatility and environment-sensitivity are quite high. Additionally, due to their instable characteristics, understanding the behaviors of stock prices has always been a challenging subject (Yi, 2007). Similar movements and patterns within an index in different time periods cannot be easily recognized.

Fama (1970) and Baumol (1965) state that the prices in the market are defined in a random walk process, making the price behavior completely unpredictable. This hypothesis is called as Efficient Market Hypothesis. Efficient Market Hypothesis claims that stock prices do not follow a trend and act randomly (Fama, 1970).

In contrast to efficient market theory, Elliot asserts that stock prices are not unpredictable, and behave in a predictable and traceable manner (Rafael, 2001). Elliot asserts that having time series price data in different time granularity, such as minute, day, week or month, enables to discover the repeating cycle (Fishcher, 1941).

Elliot wave theory is applied to different stock markets including Dow Jones (consist of the most trading companies in New York Stock Exchange), NASDAQ (the National Association of Securities Dealers Automated Quotation System), London Stock Exchange, Tokyo Stock Exchange. These implementations resulted in highly accurate estimations using Elliot Wave theory (Powers, 2001).

The objective of this study is to test Elliot Wave theory pattern rules on Istanbul Stock Exchange (ISE) – National 100 index by applying time series pattern search techniques. The study is critical, because Elliot Wave patterns hidden in National 100 index is a violation of the efficient market hypothesis.

Description of Data

For this study, Istanbul Stock Exchange National-100 Index is used. Price index dataset is retrieved from the official website of the Central Bank of the Republic of Turkey (CBRT) general statistics database (2008).

Sample period of the dataset starts on the 01.01.1986 and ends on 31.07.2008. Index is readjusted by CBRT, so that index value in 01.01.1986 is 1.

Descriptive statistics is given in Table 3.

Table 3. Description of ISE National -100 Dataset

Time Series	Sample Period		Data Range		Amount of Data Points
	Start Date	Finish Date	Min	Max	
ISE National -100	01.01.1986	31.07.2008	1	54708.42	5892 (All points with 0's included) 5266 (Holidays and weekends are excluded)

As seen in Table 3, the dataset comprises the historical data of 22 years of Istanbul Stock exchange. Time series plot of data is given in Figure 18.

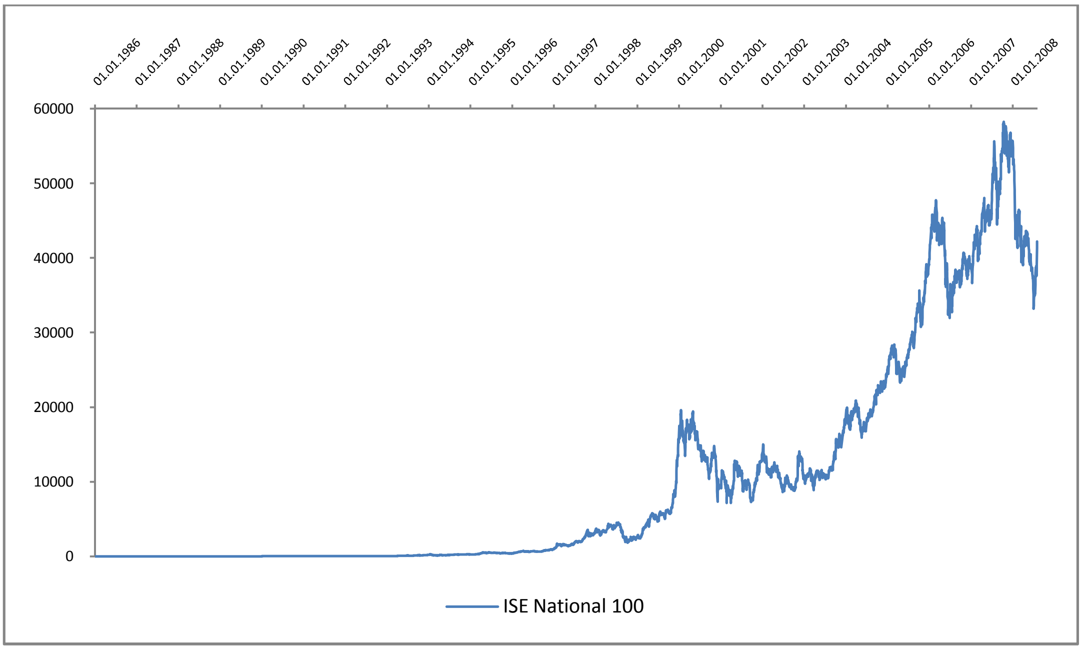


Figure 18. The plot of ISE National-100 index

In order to consider the differences between the years, summary statistics are given in a yearly base in Table 4.

Table 4. Data Statistics of ISE National-100 on a Yearly Basis

Time Series	Year	Date Range		Data Range		Amount of Data Points
		Start Date	Finish Date	Min	Max	
ISE National -100	1986	01.01.1986	31.12.1986	1	1.73	261
	1987	01.01.1987	31.12.1987	1.72	13.32	261
	1988	01.01.1988	31.12.1988	3.62	8.58	261
	1989	01.01.1989	31.12.1989	3.72	22.18	260
	1990	01.01.1990	31.12.1990	22.18	57.5	261
	1991	01.01.1991	31.12.1991	25.18	54.34	261
	1992	01.01.1992	31.12.1992	31.42	51.29	262
	1993	01.01.1993	31.12.1993	39.96	206.83	261
	1994	01.01.1994	31.12.1994	129.81	291.45	260
	1995	01.01.1995	31.12.1995	246.44	546.54	260
	1996	01.01.1996	31.12.1996	387.79	975.89	262
	1997	01.01.1997	31.12.1997	995	3543	261
	1998	01.01.1998	31.12.1998	1852.28	4530.99	261
	1999	01.01.1999	31.12.1999	2408.87	15747.94	261
	2000	01.01.2000	31.12.2000	7329.61	19577	260
	2001	01.01.2001	31.12.2001	7159.66	13782.76	261
	2002	01.01.2002	31.12.2002	8627.42	14999.51	261
	2003	01.01.2003	31.12.2003	8892.65	18625.02	261
	2004	01.01.2004	31.12.2004	15922.44	24971.68	262
	2005	01.01.2005	31.12.2005	23285.94	39837.27	260
	2006	01.01.2006	31.12.2006	31950.56	47728.5	260
	2007	01.01.2007	31.12.2007	36629.89	58231.9	261
	2008	01.01.2008	31.07.2008	33208.24	54708.42	153

Methodology

Raw data is mostly confusing and hard to analyze unless it is processed, restructured and reshaped. In order to gather valuables information from the masses of data, a disciplinary set of procedures are required. In this study, Knowledge Discovery in Databases (KDD) methodology is applied. KDD is a methodology that consists of a series of iterative steps, in which data are transformed from raw data into valuable information (Han & Kamber, 2001). The process is represented in Figure 19.

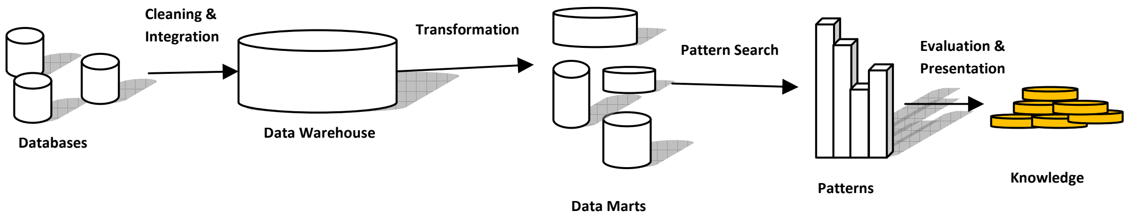


Figure 19. Knowledge discovery in databases

These steps of KDD are as follows:

1. Data Cleaning: In this phase, noisy and inconsistent data are handled.
2. Data Integration: Multiple data sources are combined, if data are collected from different data stores.
3. Data Selection: In accord with the focus of analysis, data relevant to the analyzed task are retrieved from the database.
4. Data Transformation: Data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.
5. Data Mining: This step is the most essential part of KDD methodology. Intelligent methods, algorithms are applied in this step to extract data patterns. Data mining step may be accompanied with operational know-how or experience (Han & Kamber, 2001). Additional or out-of-scope patterns are represented to user for further studies.
6. Pattern Evaluation: Achievement of data mining step leads to several patterns including both interesting and meaningful ones and trivial ones. In order to identify truly interesting patterns an evaluation process should be executed.
7. Knowledge Presentation: Interesting patterns are represented in a highly visual environment to end user.

Data mining is the task of discovering interesting patterns from large amounts of data by applying functionalities such as association, classification, clustering, trend analysis, time series data mining etc (Han & Kamber, 2001).

In this study, time series data mining functionality of data mining is used. Two time series segmentation algorithms, top-down and bottom-up algorithms are applied. As steps of KDD methodology differs for each algorithm, KDD methodology is implemented for both of the algorithms.

KDD Methodology with Top-Down Algorithm

Top-down algorithm is an algorithm that transforms data to a represented formation starting from a higher granularity moving to a lower granularity. The steps of KDD methodology are applied when implementing top-down algorithm.

Data Cleaning

The National-100 index data is a dataset prepared on a daily basis, including the weekends, holidays and the days that stock exchange is closed.

Starting from the 01.01.1986, all days that stock exchange is closed are flagged with a variable, and excluded from the scope of the analysis.

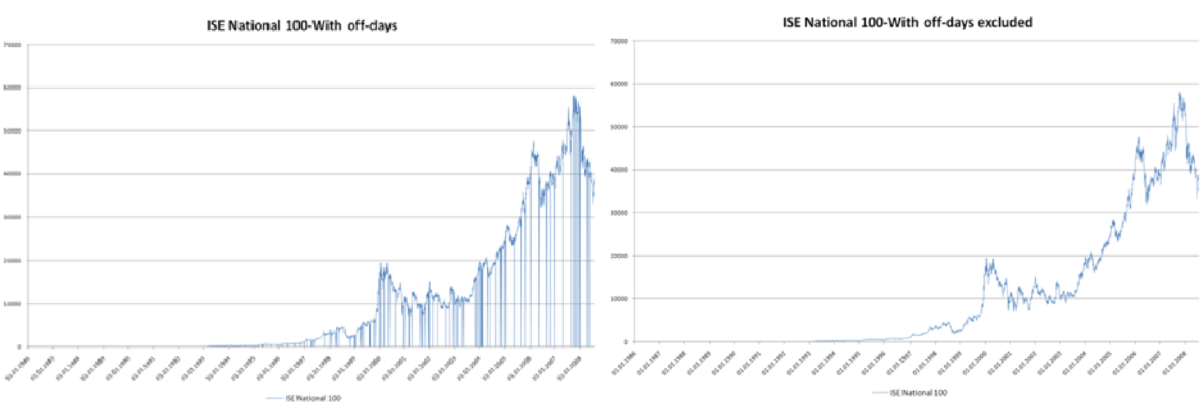


Figure 20. Data plot with and without close days of stock market

In Figure 20, data series with and without holidays and closed days are plotted.

Data Integration, Selection and Transformation

In this study, ISE National-100 index is obtained from a single data source, thus data integration and selection steps are skipped. Additionally, the usage of linear regression requires an index variable added as a new attribute to the dataset.

Data Mining

In Top-down algorithm which data points are fitted to lines that are called as segments. Initially, the dataset is fitted to a single segment, and this single segment is split into two sub-segments at the optimum point having the minimum error value. Both sub-segments are then tested if their approximation error exceeds user-defined threshold. Otherwise, the algorithm recursively split the sub-segments until each sub-segment has approximation error that exceeds the specified error. In Figure 21, actual data points, breakpoint having the minimum approximation error and fitted sub-segments are presented.

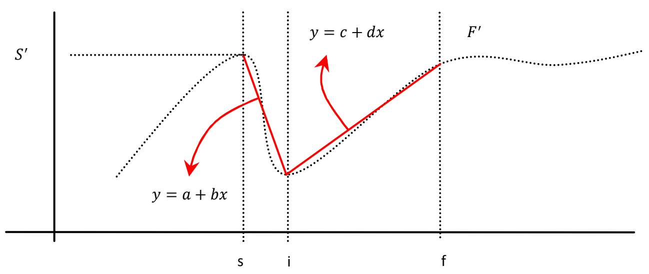


Figure 21 The breakpoint and fitted lines

In this study, approximation error is computed by the sum of squares error (SSE) given as:

$$SSE(T_k) = \sum_{1}^{n} (y_t - y_{tr})^2$$

where (T_k) is segment k, n is the number of points in segment k, y_t is the actual point t, and y_{tr} is the point on the fitted segment. The differences between the observed data points and fitted data points are squared and summed.

In Top-down algorithm, each point is tested as a breakpoint and the total error is computed. This error can be explained as the cost of splitting the segment at a particular point. Then comparing the cost of all points, algorithm determines the point with least cost and splits the segment into two, and continues with the sub-segments.

When a line is to be split, then total error is calculated as follows:

$$\sum_{t=s}^{i}\left(y_t - (a + bt)\right)^2 + \sum_{t=i+1}^{f}\left(y_t - (c + dt)\right)^2$$

where s is the starting point of sequence, f is the end point of sequence, i is the point that the segment is to be split at, a is the intercept of sub-segment fitted to the left-hand side of i , and b is the slope of is the sub-segment fitted to the left-hand side of i , c is the intercept of sub-segment fitted to the right-hand side of i , and d is the slope of is the sub-segment fitted to the right-hand side of i. Figure 21 displays an example of actual points, breakpoint, and fitted lines.

Testing each data point, the algorithm determines the point with smallest total error. Thus it can be formulated with the following objective function:

$$Min\ Z = \left(\sum_{t=s}^{i}\left(y_t - (a + bt)\right)^2 + \sum_{t=i+1}^{f}\left(y_t - (c + dt)\right)^2\right)$$

Subject to

$$(a - c) + (b - d)i = 0$$

$$a + bs = S'$$

$$c + df = F'$$

where S' is the starting point of index, and F' is the ending point of index.

Thus, when the starting and end point of a sequence is specified, the optimal breakpoint, the slope and intercepts of the sub-segments can be computed.

Top-down algorithm requires a user-defined error threshold in order to stop sub-segment splitting. After a segment is split, calculated error is compared to threshold error in order to determine whether the algorithm should stop or continue splitting.

In this study, approximation error is calculated by the following methods:

- Sum of squares error (SSE): The squares of error between the actual points and estimated points are computed and summed.

$$SSE(T_k) = \sum_{1}^{n} (y_t - y_{tr})^2$$

- Average Error: Absolute difference between the observed points and the fitted points are computed, summed and averaged over the points of interest in the corresponding segment. This average is expected to be less than the threshold for a segment to stop split.

$$E_{ave} = \frac{\sum_{1}^{n}(|y_t - y_{tr}|)}{n}$$

- Percentage Error: The difference between actual data and estimated data is calculated as a percent of actual data point.

$$E_{\%} = \sum_{1}^{n} \frac{(|y_t - y_{tr}|)}{y_t}$$

Since data range from 1 to 50,000, percentage error produces more accurate results when compared to SSE and average error.

Clustering of Segments

Using top-down algorithm, 5266 data points are divided into 417 segments. Each segment has a slope and intercept. In order to understand the natural groups in these segments, clustering is applied.

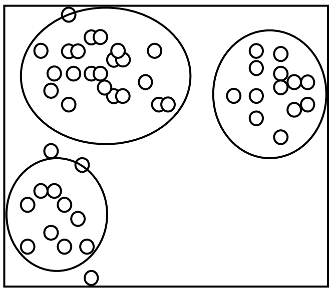


Figure 22. Clusters determined by k-means

Clustering is a data mining functionality, which partitions objects into groups or clusters, so that objects within a cluster are "similar" to one another and "dissimilar" to objects in other clusters (Han & Kamber, 2001). Similarity is commonly defined in terms of closeness to objects are in space, based on a distance function. A representation of clustered data points is given in Figure 22.

K-means algorithm is one of the clustering algorithms. The algorithm takes the input parameter, k, which is the number of clusters that will be formed. Then it partitions a set of n objects into k clusters so that the resulting intracluster similarity is high but the intercluster similarity is low.

K-means algorithm is applied on segments produced by top-down algorithm. Firstly, since the units of variables, slope and duration, are different, z-transformation is applied and variables are standardized.

After z-transformation is complete, k-means algorithm is applied varying k's. Thus, a set of k's are tested and clustering error is computed. Clustering error is defined as:

$$E = \sum_{i=1}^{k} \sum |p - m_i|^2$$

where E is the sum square-error for all points in the dataset, p is the actual point, and m_i is the center of the cluster k (Han & Kamber, 2001).

k value, at which k versus clustering error plot flattens is determined as the number of clusters.

KDD Methodology with Bottom-Up Algorithm

In Bottom-up algorithm, beginning from the finest possible approximation, segments are merged until a user specified stopping criteria is met. The steps of KDD methodology are applied when implementing bottom-up algorithm.

Data Cleaning, Integration, Selection & Transformation

Data cleaning, integration, selection and transformation steps of KDD methodology are skipped due to the fact that data is obtained from single data source and no transformation is required.

Data Mining

As mentioned in the literature summary, bottom-up algorithm is a way of time series segmentation method, in which granular data is combined and straight lines are constructed. Just like the top-down algorithm, linear regression is applied as the merging technique.

As the nature of Elliot's Wave is in the form of successive impulsive-corrective movements, bottom- up algorithm is modified to capture the Elliot's wave pattern. Pseudocode of algorithm is given in Figure 23.

```
START
        IF actual_point>next_point THEN
                Set Flag= 'down'
        ELSE IF actual_point<next_point THEN
                Set Flag= 'up'
        ELSE
                Set Flag= 'same'
        END;
        Reset Flag WHERE Flag= 'same'
        Calculate_regression_line;
        Calculate_mid_point;
TERMINATE;
```

Figure 23. Pseudocode of modified bottom-up algorithm

Flowchart of the modified version of bottom-up algorithm is given in Figure 24.

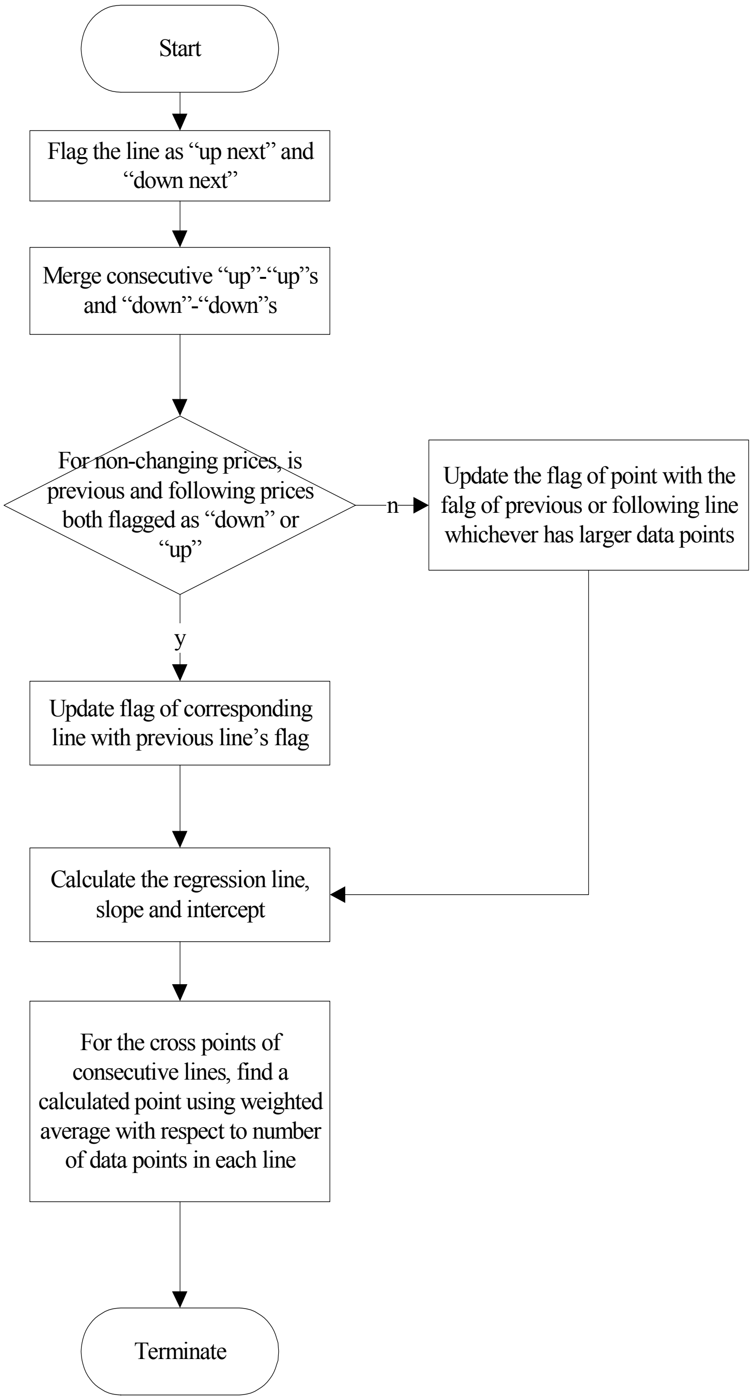


Figure 24. Flowchart of the modified version of bottom-up algorithm

Using modified version of the algorithm following improvements in segmentation is achieved:

- Lines with approximately zero slope which formed by constant prices, disappeared.

- If two successive segments are in the same direction (either up-up or down-down), they are merged and another segment is generated.

Testing Elliot Wave Rules

Creating Segment Groups

An Elliot wave is defined as five impulsive movement and three corrective movements in the opposite direction (Davidson, 2006).

However, definition of Elliot Wave does not state where the wave initiates or ends. The theory mainly focuses on the pattern of wave. Thus, an Elliot wave can emerge in any section of the time series independent of time. Additionally, waves that start from different time points can overlap with each other. In order to find successive Elliot waves, it is crucial to identify a starting point in time. Different starting points may result in capturing different Elliot Waves.

Segments, generated by modified version of bottom-up algorithm, are systematically decomposed to segment groups using sliding-windows method. A window is defined as a frame that can consist of exactly 8 segments. Starting from the first segment, the window is slided by 2 segments to the right. Window is slided by two segments each time, because an Elliot Wave should start with an impulsive movement. The first window will span the set of segments labeled as 1 to 8. Then first sliding is performed, meanly, windows is moved by 2 segments to the right. At this time, the window will span the set of segments labeled as 2 to 10. Sliding continues in this manner.

Sliding windows results in 8-length segment sets that overlap with other segment sets. In order to remove overlapping, 4 segment groups are formed in which successive non-overlapping segment sets are grouped. A representation of segment sets and groups generated by sliding windows is given in Figure 25.

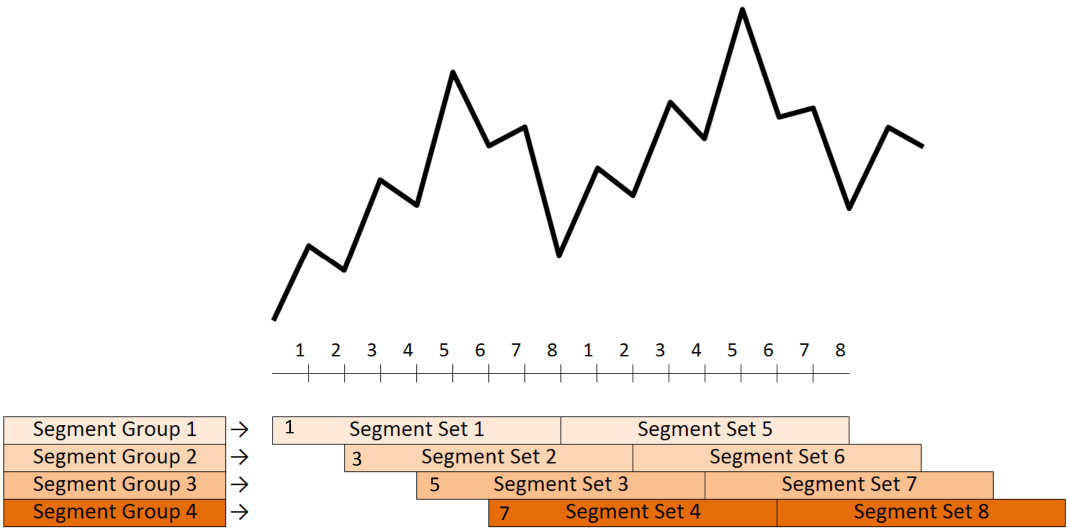


Figure 25. Segment groups

Characteristics of four segment groups in ISE-National 100 are given in Table 5:

Table 5. Summary of Segment Groups

Segment Group	Number of Elliot Wave Candidates	Number of Segments
1	291	2328
2	291	2328
3	290	2320
4	290	2320
All Segments		2333

Elliot Wave rules are implemented on each of the segment group separately.

Testing Segment Groups against Rules

Rules and Hypotheses to be Tested

When exact matching is tested, a real Elliot wave may violate Elliot Wave rules (Neely & Hall, 1990). As a result, waves are tested against the rules by formulated hypothesis.

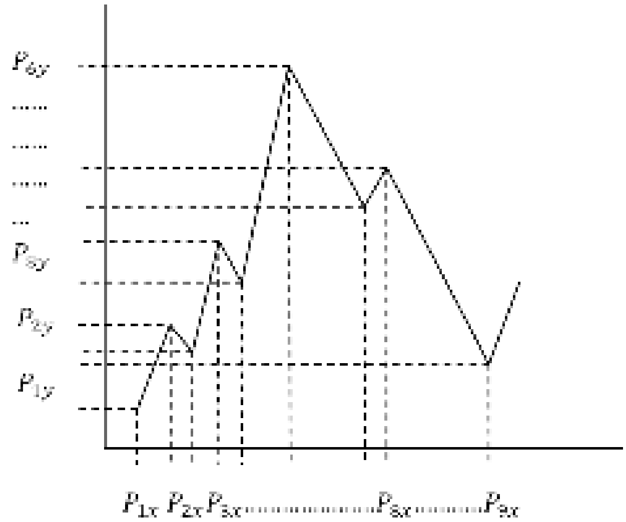


Figure 26. Segment coordinates

Additionally, rules defined as “A cycle is never complete unless it is composed of a fully 8 movement” and “An upward movement should be followed by a downward movement” is not tested, because applying modified version of bottom-up algorithm ensures the existence of these patterns. The coordinates of an Elliot Wave is given in Figure 26.

The rules and hypothesis are given as follows:

R1: *3rd Segment should be the longest segment at duration.*

$$H_0: P_{3x} - P_{2x} = \max(P_{2x} - P_{1x}, P_{3x} - P_{2x}, P_{4x} - P_{3x}, P_{5x} - P_{4x}, P_{6x} - P_{5x}, P_{7x} - P_{6x}, P_{8x} - P_{9x})$$

$$H_1: P_{3x} - P_{2x} \neq \max(P_{2x} - P_{1x}, P_{3x} - P_{2x}, P_{4x} - P_{3x}, P_{5x} - P_{4x}, P_{6x} - P_{5x}, P_{7x} - P_{6x}, P_{8x} - P_{9x})$$

The hypothesis in R1 is tested by multivariate hypothesis testing. In multivariate hypothesis test, when sample size is large enough hypothesis and confidence regions for mean can be constructed without the assumption of a normal population. Multiple confidence intervals construct a multi-dimensional result set.

Let $X_1, X_{2,\ldots,}X_n$ be a random sample from a population with mean μ and positive definite covariance matrixΣ. When n-p is large, the hypothesis $H_0: \mu = \mu_0$ is rejected in favor of $H_1: \mu \neq \mu_0$, at a level of significance α, if the observed

$$n(\overline{x} - \mu_0)'S^{-1}(\overline{x} - \mu_0) > \chi_p^2(\alpha)$$

where $\overline{x}$ and S are the sample mean vector and covariance matrix of the X_j's respectively, n is the sample size, p is degree of freedom and $\chi_p^2(\alpha)$ is the upper 100 αth percentile of a chi-square distribution with p d.f.

R2: 2nd segment cannot retrace 1st segment at price.

$$H_0: P_{3y} - P_{1y} > 0$$

$$H_1: P_{3y} - P_{1y} \leq 0$$

This requires the end point of 2^{nd} segment to be at a higher level when compared to the beginning point of 1^{st} segment.

R3: 4th segment cannot retrace 3rd segment at price

$$H_0: P_{5y} - P_{3y} > 0$$

$$H_1: P_{5y} - P_{3y} \leq 0$$

The second drop of prices cannot be larger than total increment of segment 3.

R4: 1st and 5th segments should be same in duration

$$H_0: (P_{2x} - P_{1x}) - (P_{6x} - P_{5x}) = 0$$

$$H_1: (P_{2x} - P_{1x}) - (P_{6x} - P_{5x}) \neq 0$$

R5: 1st and 5th segments should be same in price

$$H_0: (P_{2y} - P_{1y}) - (P_{6y} - P_{5y}) = 0$$

$$H_1: (P_{2y} - P_{1y}) - (P_{6y} - P_{5y}) \neq 0$$

R6: If 2nd segment is a sharp decrease, 4th segment should be a slow decrease, or vice versa.

$$H_0 = \frac{(P_{2y} - P_{3y})}{(P_{3x} - P_{2x})} \geq -1 \ and \ \frac{(P_{4y} - P_{5y})}{(P_{5x} - P_{4x})} \geq -1$$

R7: 3rd segment should climb up 1st segment.

$$H_0: P_{4y} - P_{2y} > 0$$

$$H_1: P_{4y} - P_{2y} \leq 0$$

R8: 4th segment cannot climb down the 1st segment

$$H_0: P_{5y} - P_{2y} > 0$$

$$H_1: P_{5y} - P_{2y} \leq 0$$

Fibonacci Rules are as follows;

F1: 3rd segment should be %161 or %261 of the 1st segment at price.

$$H_0: P_{4y} - (1.61 * P_{2y}) = 0 \ Or \ \ P_{4y} - (2.61 * P_{2y}) = 0$$

$$H_1: P_{4y} - (1.61 * P_{2y}) \neq 0 \ And \ \ P_{4y} - (2.61 * P_{2y}) \neq 0$$

Rule testing will be applied separately to each of “or” statements.

F2: 5th segment should climb up %61, %100 or %161 of 1st segments end point.

$$H_0: P_{6y} - (.61 * P_{2y}) = 0 \ Or \ \ P_{6y} - P_{2y} = 0 \ or \ P_{6y} - (1.61 * P_{2y}) = 0$$

$$H_1: P_{6y} - (.61 * P_{2y}) \neq 0 \ and \ \ P_{6y} - P_{2y} \neq 0 \ and \ P_{6y} - (1.61 * P_{2y}) \neq 0$$

Rule testing are applied separately to each of “or” statements.

F3: 2nd segment should climb down %61 of the 1st segment at price.

$$H_0: P_{3y} - (1 - .61)P_{2y} = 0$$

$$H_1: P_{3y} - (1 - .61)P_{2y} \neq 0$$

F4: 4th segment should climb down %61 of the 3rd segment at price

$$H_0: P_{5y} - (1 - .61)P_{4y} = 0$$

$$H_1: P_{5y} - (1 - .61)P_{4y} \neq 0$$

F5: 5th segment moves at least % 61 of the 4th segment at duration.

$$H_0: (P_{6x} - P_{5x}) - (P_{5x} - P_{4x}) * 1.61 \geq 0$$

$$H_1: (P_{6x} - P_{5x}) - (P_{5x} - P_{4x}) * 1.61 < 0$$

F6: 5th segment should be at least % 61 of the 4th segment at price.

$$H_0: P_{6y} - (1.61)P_{4y} \geq 0$$

$$H_1: P_{6y} - (1.61)P_{4y} < 0$$

Normality Tests

Hypotheses constructed for rules mainly require the normality assumption, thus Kolmogorov-Smirnov goodness-of-fit test is applied to test the normality. However, the hypothesis of rule 1 consists of many variables that form multiple confidence intervals. Thus a multivariate test is performed. When the sample size is large, multivariate hypothesis test and confidence regions for μ can be constructed without the assumption of a normal population, because deviation from a normal population is overcome by large sample size (Mardia, Kent, & Bibby, 1979). Therefore, goodness-of-fit test is not applied for rule 1.

In Fibonacci rules, F1 and F2, hypotheses consist of two statements combined with a logical "or" operator. Each statement is tested separately. Normality condition is searched for both of the statements simultaneously. Statements are renamed as follows:

F1;

F11 $P_{4y} - (1.61 * P_{2y}) = 0$

F12 $P_{4y} - (2.61 * P_{2y}) = 0$

F2;

F21 $P_{6y} - (.61 * P_{2y}) = 0$

F22 $P_{6y} - P_{2y} = 0$

F23 $P_{6y} - (1.61 * P_{2y}) = 0$

CHAPTER IV

RESULTS AND FINDINGS

In this chapter, segmentation results by top-down and bottom-up algorithms are given. Statistical tests of compulsory and Fibonacci rules and hypothesis results are presented.

Results of KDD Methodology with Top-Down Algorithm

Top-down algorithm is implemented on data set using three different error computation criteria. In Figure 27, Figure 28, Figure 29 segment plots can be seen for each of the three error criteria, SSE, Average Error and Percentage error respectively.

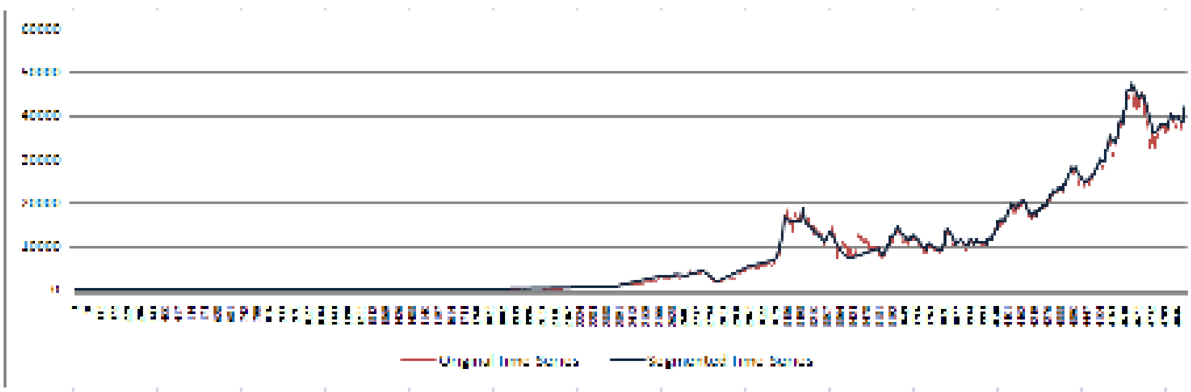


Figure 27. Segments determined by SSE

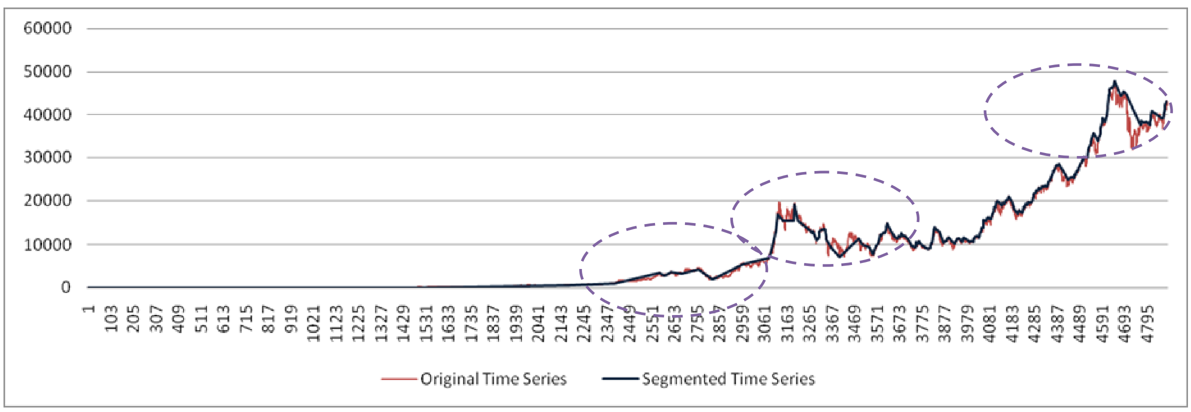


Figure 28. Segments determined by average error

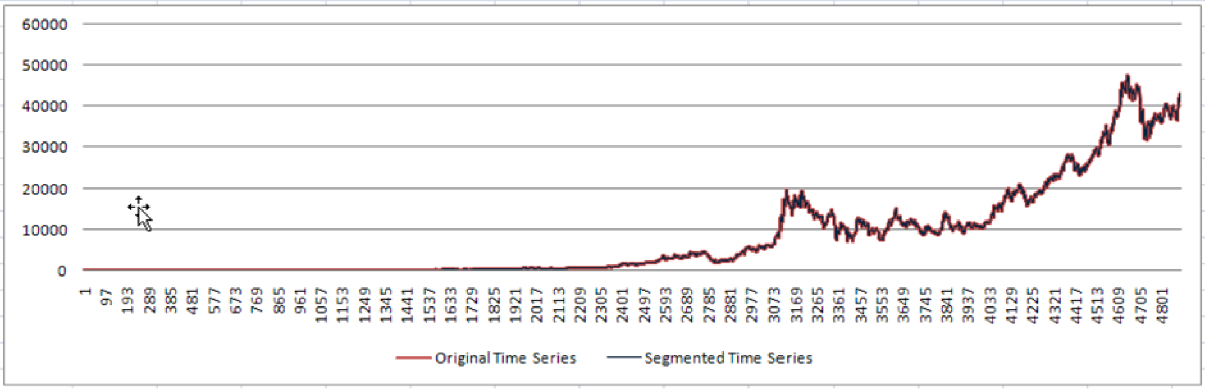


Figure 29. Segments determined by percentage error

In Table 6, comparison of three methods, in terms of error values are given. Three error criteria are used when computing the approximation error. However, SSE is used when comparing the three error computation methods.

Table 6. Comparison of Threshold Error Criteria

	Total number of data points	Number of segments	SSE	Total Error
Threshold Error by SSE	5266	307	147,123.23	12,134.34
Threshold Error by Average Error	5266	256	158,147.87	15,768.55
Threshold Error by Percentage Error	5266	417	91,128.54	9,918.23

Both visually and numerically, SSE and Average Error cannot capture some of the segments as accurate as percentage error.

As seen is the Table 6, the least error value is achieved by employing percentage error criteria. Considering the exponentially growing nature of stock prices, SSE and average error criteria result in higher values when compared to percentage error criteria.

After determining segments by top-down algorithm with percentage error criteria, natural groups in these segments are investigated.

As explained in chapter 3, optimal number of clusters, k, is the point where k versus SSE plot flattens when k varies from 2 to 20.

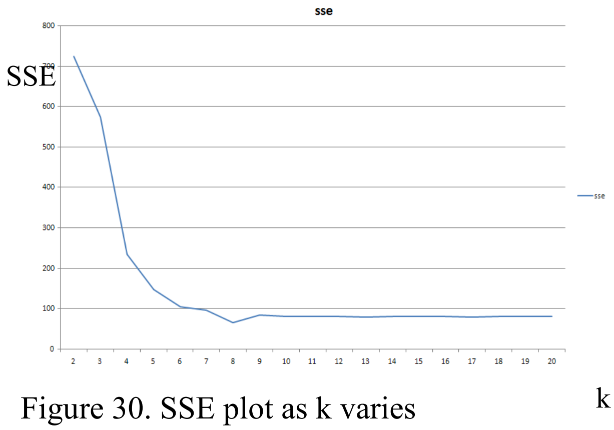


Figure 30. SSE plot as k varies

Figure 30 displays the SSE curve, as k changes. After cluster 7, error starts to flatten. Thus k is determined to be 7.

k-means algorithm is applied on 417 segments and clusters concluded by the algorithm is given in Figure 31.

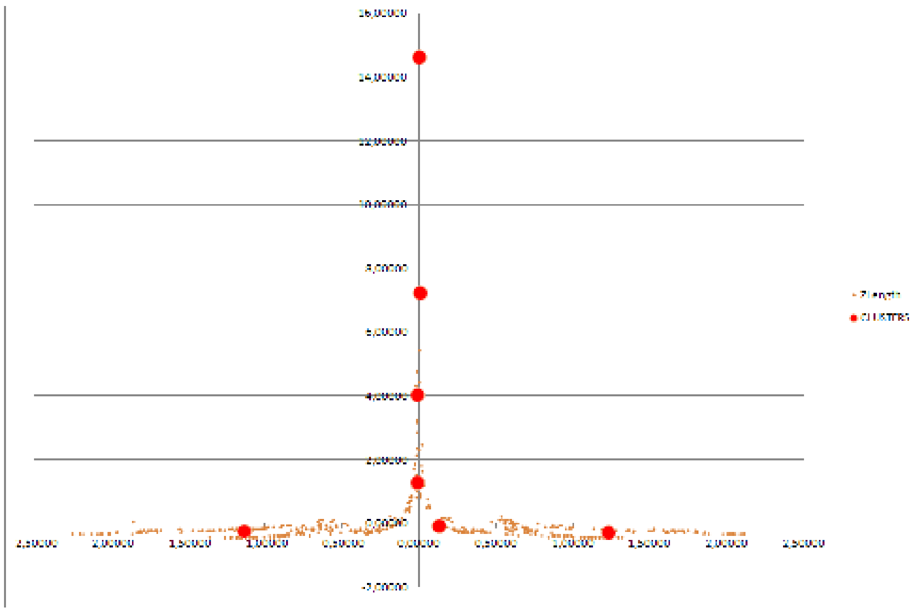


Figure 31. Data points and cluster centers for each cluster

Centers for each cluster, total number of segments are given in Table 7. Clusters are named examining the slope and duration values of clusters.

Table 7. Cluster Centers

Clusters	Slope	Duration	# of cases	% of cases	Cluster Names
1	-0.032	33.117	34	7.1%	Constant segments
2	5.694	4.526	108	22.5%	Sharply Increasing movements
3	-0.023	83.571	7	1.5%	Constant segments
4	-5.221	5.472	141	29.3%	Sharply Decreasing movements
5	0.625	8.561	126	26.2%	Slowly Increasing Movements
6	0.023	277	1	0.2%	Constant long segments-Outliers
7	0.048	141.999	1	0.2%	Constant long segments-Outliers

The results of clustering show that 81% of all segments have either an increasing or decreasing trend, however 9% of segments are horizontal.

Horizontal segments are not accepted in Elliot's wave, because the movement should be either increasing or decreasing.

Table 8. Slope Direction Summary

		Third Segment's Slope	
First S. Slope	Sec. S. Slope	Decrease	Increase
Decrease	Decrease	59	56
	Increase	35	32
Increase	Decrease	57	69
	Increase	52	58

As shown in Table 8, about %25 percent of successive segments are in the same direction (either up-up or down-down). Thus, due to observed same-directional trends in segments, top-down algorithm is not appropriate to use.

Results of KDD Methodology with Bottom-Up Algorithm

Modified version of the Bottom-Up algorithm is executed on the dataset. 2333 segments are generated. Error statistics are given in Table 9 and Table 10.

The comparison of actual ISE National-100 prices and fitted segments can be seen in Figure 32.

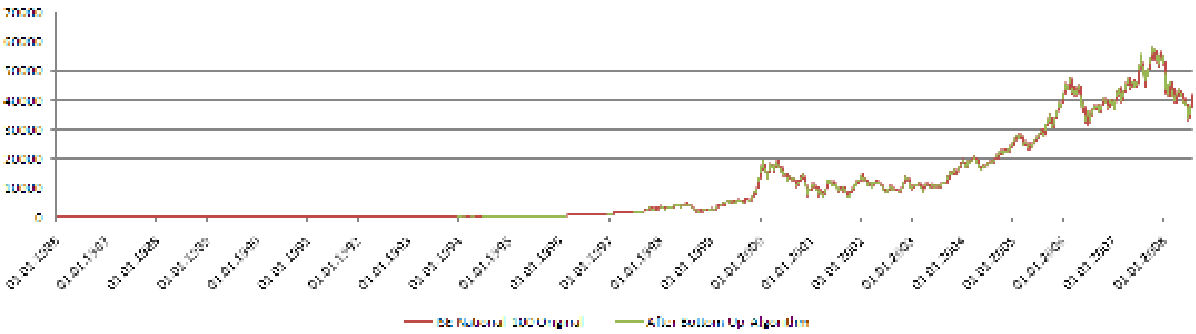


Figure 32. Segments determined by modified version of bottom-up algorithm

Furthermore, this algorithm produces segments having a percentage error of 12.5% at maximum.

Table 9. Error Values

Error Description	Maximum	Minimum	Average
Total error	1403	0	1.22
Percentage Error	12.5 %	0	1 %

On the other hand, in this algorithm total number of segments is almost five times higher when compared to top-down algorithm.

Table 10. SSE of Bottom-Up Algorithm

	Total number of data points	Number of segments	SSE	Total Error
Threshold Error by SSE	5266	2333	50,367.67	17,134.34

The comparison of two algorithms is given in Table 11.

Table 11. Comparison of Algorithms

Indicator	Top-Down	Bottom-Up
SSE	SSE is higher	SSE is lower
Percentage Error	Percentage error is higher	Percentage error is lower
Segment count	Number of segments is smaller	Number of segments is larger
Slope	Constant sloped lines can be encountered	Only increasing and decreasing slopes can be encountered
Price movement direction	A segment can be followed by a segment with same signed slope	All segments followed by a different signed segment
Visual fit	Visual fit is worse	Visual fit is better
Ease of implementation	Ease of implementation is low	Ease of implementation is high

As a result modified version of bottom-up approach is selected as the time series segmentation technique.

Normality Test Results

In order to apply hypothesis test, Kolmogorov-Smirnov goodness-of-fit test is performed for all rules except R1 in which multivariate hypothesis testing is implemented.

The results of Kolmogorov-Smirnov normality test for each segment groups are given in Table 12, Table 13, Table 14, Table 15, and Table 16 for each segment groups.

Table 12. K-S Test Results for Segment Group 1

SG	Rules	N	Normal Parameters		Most Extreme Differences			KS - Z	Critical KS-Z Value
			Mean	Std. Deviation	Absolute	Positive	Negative		
Segment Group 1	R2	291	-10.97	0.952	0.0266	0.0214	-0.0266	0.453	0.797
	R3	291	29.68	0.713	0.0225	0.0201	-0.0225	0.383	0.797
	R4	291	0.08	0.002	0.0171	0.0171	-0.0125	0.291	0.797
	R5	291	-17.54	0.683	0.0212	0.0204	-0.0212	0.361	0.797
	R6	291	-13.54	0.643	0.0412	0.0304	-0.0112	0.361	0.797
	R7	291	-35.02	0.907	0.0249	0.0226	-0.0249	0.424	0.797
	R8	291	-386.23	1.167	0.0226	0.0222	-0.0226	0.385	0.797
	R9	291	-162.16	0.424	0.0351	0.0351	-0.0254	0.598	0.797
	R10	291	139.46	0.286	0.0313	0.0249	-0.0313	0.533	0.797
	F11	291	-7165.9	9.604	0.0228	0.0228	-0.0197	0.388	0.797
	F12	291	-1.891	2.505	0.0226	0.0226	-0.0201	0.385	0.797
	F21	291	4595.40	6.075	0.0225	0.0201	-0.0225	0.383	0.797
	F22	291	36.39	1.205	0.0235	0.0205	-0.0235	0.400	0.797
	F23	291	-7094.3	9.608	0.0236	0.023	-0.0197	0.392	0.797
	F3	291	6714.62	8.943	0.0226	0.0199	-0.0226	0.385	0.797
	F4	291	6758.11	9.029	0.0227	0.0198	-0.0227	0.387	0.797
	F5	291	-1.59	0.002	0.0246	0.0196	-0.0246	0.418	0.797
	F6	291	-7037.8	9.374	0.0226	0.0226	-0.0196	0.385	0.797

Table 13. K-S Test Results for Segment Group 2

SG	Rules	N	Normal Parameters		Most Extreme Differences			KS - Z	Critical KS-Z Value
			Mean	Std. Deviation	Absolute	Positive	Negative		
Segment group 2	R2	291	29.68	0.713	0.0225	0.0201	-0.0023	0.383	0.797
	R3	291	55.91	0.725	0.0217	0.0217	-0.0022	0.370	0.797
	R4	291	.08	0.002	0.0166	0.0137	-0.0017	0.283	0.797
	R5	291	-47.85	0.695	0.0244	0.0198	-0.0024	0.416	0.797
	R6	291	-13.54	0.643	0.0412	0.0304	-0.0112	0.361	0.797
	R7	291	71.47	0.753	0.0235	0.0235	-0.0023	0.400	0.797
	R8	291	-295.29	0.991	0.0243	0.0206	-0.0024	0.414	0.797
	R9	291	-139.46	0.286	0.0313	0.0313	-0.0025	0.533	0.797
	R10	291	147.68	0.280	0.0299	0.0248	-0.003	0.510	0.797
	F11	291	-703.85	9.374	0.0226	0.0226	-0.002	0.385	0.797
	F12	291	-1.874	24.736	0.0225	0.0225	-0.002	0.383	0.797
	F21	291	4678.82	6.199	0.0225	0.0198	-0.0023	0.383	0.797
	F22	291	133.65	1.006	0.0218	0.0218	-0.002	0.371	0.797
	F23	291	-697.90	9.316	0.0227	0.0227	-0.0021	0.387	0.797
	F3	291	6758.11	9.029	0.0227	0.0198	-0.0023	0.387	0.797
	F4	291	6786.13	9.072	0.0227	0.0199	-0.0023	0.387	0.797
	F5	291	-1.33	0.002	0.0223	0.02	-0.0022	0.380	0.797
	F6	291	-709.18	9.358	0.0224	0.0224	-0.0021	0.382	0.797

Table 14. K-S Test Results for Segment Group 3

SG	Rules	N	Normal Parameters		Most Extreme Differences			KS - Z	Critical KS-Z Value
			Mean	Std. Deviation	Absolute	Positive	Negative		
Segment Group 3	R2	290	55.53	726.306	0.022000	0.0220	-0.0217	0.374	0.079
	R3	290	45.86	795.184	0.023100	0.0231	-0.0224	0.393	0.079
	R4	290	-.07	2.167	0.016200	0.0132	-0.0162	0.275	0.079
	R5	290	15.05	760.661	0.023400	0.0234	-0.0223	0.398	0.079
	R6	290	-13.54	0.643143	0.041200	0.03	-0.0112	0.361	0.079
	R7	290	61.63	670.172	0.022600	0.0226	-0.0201	0.384	0.079
	R8	290	-320.21	1.0073	0.025900	0.0207	-0.0259	0.441	0.079
	R9	290	-147.16	280.930	0.030000	0.0300	-0.0250	0.510	0.079
	R10	290	151.29	314.344	0.031500	0.0253	-0.0315	0.536	0.079
	F11	290	-704.21	9.3363	0.022500	0.0225	-0.0206	0.383	0.079
	F12	290	-1.87E4	2.4724	0.022500	0.02	-0.0201	0.383	0.079
	F21	290	4627.23	6.1513	0.022600	0.0194	-0.0226	0.384	0.079
	F22	290	86.30	1.0763	0.024200	0.0198	-0.0242	0.412	0.079
	F23	290	-701.50	9.4043	0.022800	0.0228	-0.0207	0.38	0.079
	F3	290	6736.59	9.0493	0.022800	0.0199	-0.0228	0.38	0.079
	F4	290	6758.31	8.9973	0.022600	0.0201	-0.0226	0.384	0.079
	F5	290	-1.53	2.431	0.018900	0.0164	-0.0189	0.321	0.079
	F6	290	-711.71	9.5153	0.022700	0.0227	-0.0198	0.386	0.079

Table 15. K-S Test Results for Segment Group 4

SG	Rules	N	Normal Parameters		Most Extreme Differences			KS - Z	Critical KS-Z Value
			Mean	Std. Deviation	Absolute	Positive	Negative		
Segment Group 4	R2	290	45.86	795.184	0.0231	0.0231	-0.0224	0.393	0.079
	R3	290	-11.01	954.443	0.0266	0.0213	-0.0266	0.452	0.079
	R4	290	-.08	2.127	0.0153	0.0153	-0.0122	0.260	0.079
	R5	290	45.38	704.905	0.0227	0.0227	-0.0199	0.386	0.079
	R6	290	43.38	73.905	0.023	0.0343	-0.0199	0.386	0.079
	R7	290	24.59	805.786	0.0242	0.0213	-0.0242	0.412	0.079
	R8	290	-392.87	1.3023	0.0245	0.0217	-0.0245	0.417	0.079
	R9	290	-151.29	314.344	0.0315	0.0315	-0.0253	0.536	0.079
	R10	290	162.71	425.149	0.0351	0.0254	-0.0351	0.597	0.079
	F11	290	-711.71	9.5153	0.0227	0.0227	-0.0198	0.386	0.079
	F12	290	-1.88E4	2.4994	0.0226	0.0226	-0.0197	0.384	0.079
	F21	290	4554.46	6.013	0.0225	0.0198	-0.0225	0.383	0.079
	F22	290	-10.55	1.2393	0.0236	0.0221	-0.0236	0.401	0.079
	F23	290	-715.96	9.6833	0.0227	0.0227	-0.0198	0.386	0.079
	F3	290	6758.31	8.9973	0.0226	0.0201	-0.0226	0.384	0.079
	F4	290	6737.77	8.9503	0.0226	0.0198	-0.0226	0.384	0.079
	F5	290	-1.61	2.844	0.0202	0.0175	-0.0202	0.343	0.079
	F6	290	-719.66	9.6123	0.0227	0.0227	-0.0196	0.386	0.079

Table 16. K-S Test Results for All Segments

SG	Rules	N	Normal Parametersa		Most Extreme Differences			KS - Z	Critical KS-Z Value
			Mean	Std. Deviation	Absolute	Positive	Negative		
All Segments	R2	1162	-.039763	112.800	0.0255	0.0214	-0.0255	0.869	0.039
	R3	1162	-.039763	1128007	0.0255	0.0214	-0.0255	0.869	0.039
	R4	1162	.01	2.178	0.0147	0.0147	-0.0143	0.501	0.039
	R5	1162	.021151	7.6586289	0.0212	0.0210	-0.0212	0.722	0.039
	R6	1162	-.04024	1.1232487	0.0324	0.0245	-0.0254	0.864	0.039
	R7	1162	-.040609	1.1217087	0.0254	0.0220	-0.0254	0.86	0.039
	R8	1162	-43108	1.6129455	0.0238	0.0238	-0.0236	0.811	0.039
	R9	1162	-17003	4.4289313	0.0351	0.0351	-0.0254	1.196	0.039
	R10	1162	17.0181	4.4305340	0.0350	0.0254	-0.0350	1.193	0.039
	F11	1162	-7317	9.5798661	0.0222	0.0222	-0.0195	0.756	0.039
	F12	1162	-19406	2.5195287	0.0221	0.0221	-0.0196	0.753	0.039
	F21	1162	4700	6.2155081	0.0223	0.0191	-0.0223	0.760	0.039
	F22	1162	-.02389	1.6072661	0.0250	0.0203	-0.0250	0.852	0.039
	F23	1162	-7.35502	9.7373167	0.0224	0.0224	-0.0200	0.76	0.039
	F3	1162	6921	9.0879939	0.0222	0.0196	-0.0222	0.756	0.039
	F4	1162	6.9212	9.0879939	0.0222	0.0196	-0.0222	0.756	0.039
	F5	1162	-15169	2.7134004	0.0213	0.0184	-0.0213	0.726	0.039
	F6	1162	-7352	9.6271187	0.0222	0.0222	-0.0195	0.756	0.039

In all segment groups, all rules satisfy normality assumption, thus parametric hypothesis tests can be applied.

Hypothesis Test Results

R1: Segment groups are tested by a multivariate hypothesis test for R1. The hypothesis $H_0: \mu = \mu_0$ is rejected in favor of $H_1: \mu \neq \mu_0$, at a level of significance α, if the observed

$$n(\overline{x} - \mu_0)' S^{-1} (\overline{x} - \mu_0) > \chi_p^2(\alpha)$$

where $\overline{x}$ and S are the sample mean vector and covariance matrix of the X_j's respectively, n is the sample size, p is degree of freedom and $\chi_p^2(\alpha)$, which is the critical value, is the upper 100 αth percentile of a chi-square distribution with p d.f. $\overline{x}$ is the average of difference of each segment from 3rd segment, so seven $\overline{x}$'s are constructed and descriptive statistics for each segment group is given in Table 17.

Table 17. Difference Matrix

	$\bar{x}_{s2-s1}$	$\bar{x}_{s3-s2}$	$\bar{x}_{s4-s3}$	$\bar{x}_{s5-s4}$	$\bar{x}_{s6-s5}$	$\bar{x}_{s7-s6}$	$\bar{x}_{s8-s7}$
Segment Group 1	1.21	-0.27	-0.13	-0.48	-0.08	-1.19	-0.18
Segment Group 2	1.03	-0.31	-0.22	-0.74	-0.32	-1.31	-0.40
Segment Group 3	1.37	-0.11	-0.22	-0.63	-0.31	-1.24	-0.18
Segment Group 4	1.30	-0.23	-0.30	-0.73	-0.22	-1.16	-0.14

Applying $n(\overline{x}-\mu_0)'S^{-1}(\overline{x}-\mu_0) > \chi_p^2(\alpha)$, in which $\chi^2_{>100}(0.05) = 14.7$, for each segment group, test results given in Table 18 is obtained:

Table 18. Test Results for R1

Segment Groups	Rule Statistics	H_0 Test Result
Segment Group 1	16,456> 14,7	Accept
Segment Group 2	12,54> 14,7	Reject
Segment Group 3	19,122> 14,7	Accept
Segment Group 4	14,77> 14,7	Accept
All Segments	7.3443>14.7	Reject

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 18.

R2: 2nd segment cannot retrace 1st segment at price.

$$H_0: P_{3y} - P_{1y} > 0$$

$$H_1: P_{3y} - P_{1y} \leq 0$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 19:

Table 19. Test Results for R2

Segment Groups	$z = \frac{\bar{x}-\mu}{s}$	z_α	H_0 Test Result
Segment Group 1	-11.51348242	-1.65	Reject
Segment Group 2	41.62079918	-1.65	Accept

Segment Group 3	0.076455378	-1.65	Accept
Segment Group 4	-2.057672187	-1.65	Reject
All Segments	-0.36	-1.65	Accept

- *R3: 4th segment cannot retrace 3rd segment at price*

$$H_0: P_{5y} - P_{3y} > 0$$

$$H_1: P_{5y} - P_{3y} \leq 0$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 20:

Table 20. Test Results for R3

Segment Groups	$z = \frac{\bar{x} - \mu}{s}$	z_α	H_0 Test Result
Segment Group 1	41.62079918	-1.65	Accept
Segment Group 2	77.10862648	-1.65	Accept
Segment Group 3	0.057672187	-1.65	Accept
Segment Group 4	-2.01153554	-1.65	Reject
All Segments	-0.35250708	-1.65	Reject

- *R4: 1st and 5th segments should be same in duration*

$$H_0: (P_{2x} - P_{1x}) - (P_{6x} - P_{5x}) = 0$$

$$H_1: (P_{2x} - P_{1x}) - (P_{6x} - P_{5x}) \neq 0$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 21.

Table 21. Test Results for R4

Segment Groups	$z = \frac{\bar{x} - \mu}{s}$	$z_{-\alpha/2}, z_{\alpha/2}$	H_0 Test Result
Segment Group 1	38.31417625	+-1.96	Reject
Segment Group 2	35.3200883	+-1.96	Reject
Segment Group 3	-0.0323027	+-1.96	Accept

Segment Group 4	-0.03761166	+-1.96	Accept
All Segments	3.004591368	+-1.96	Reject

- *R5: 1st and 5th segments should be same in price*

$$H_0: (P_{2y} - P_{1y}) - (P_{6y} - P_{5y}) = 0$$

$$H_1: (P_{2y} - P_{1y}) - (P_{6y} - P_{5y}) \neq 0$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 22:

Table 22. Test Results for R5

Segment Groups	$z = \frac{\bar{x} - \mu}{s}$	$z_{-\alpha/2}, z_{\alpha/2}$	H_0 Test Result
Segment Group 1	-25.6769479	+-1.96	Reject
Segment Group 2	-68.8062510	+-1.96	Reject
Segment Group 3	0.019785423	+-1.96	Accept
Segment Group 4	0.064377469	+-1.96	Accept
All Segments	0.27617210	+-1.96	Accept

- *R6: If 2nd segment is a sharp decrease, 4th segment should be a slow decrease, or vice versa.*

$$H_0 = \frac{(P_{2y} - P_{3y})}{(P_{3x} - P_{2x})} \geq -1 \; and \; \frac{(P_{4y} - P_{5y})}{(P_{5x} - P_{4x})} \geq -1$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 23:

Table 23. Test Results for R6

Segment Groups	$z_{-1} = \frac{\bar{x} - \mu}{s}$	$z_1 = \frac{\bar{x} - \mu}{s}$	$z_{-\alpha/2}, z_{\alpha/2}$	H_0 Test Result
Segment Group 1	0.040261	0.895711	+-1.96	Accept
Segment Group 2	-0.367928	0.974367	+-1.96	Accept
Segment Group 3	0.206596	0.02703	+-1.96	Accept
Segment Group 4	-0.255553	0.122805	+-1.96	Accept
All Segments	-0.928041	0.048554	+-1.96	Accept

- *R7: 3rd segment should climb up 1st segment.*

$$H_0: P_{4y} - P_{2y} > 0$$

$$H_1: P_{4y} - P_{2y} \leq 0$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 24:

Table 24. Test Results for R7

Segment Groups	$z = \frac{\bar{x} - \mu}{s}$	z_α	H_0 Test Result
Segment Group 1	-38.57788392	-1.65	Reject
Segment Group 2	94.82074673	-1.65	Accept
Segment Group 3	-1.98237389	-1.65	Reject
Segment Group 4	0.030516787	-1.65	Accept
All Segments	-3.6202809	-1.65	Accept

- *R8: 4th segment cannot climb down the 1st segment*

$$H_0: P_{5y} - P_{2y} > 0$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 25.

Table 25. Test Results for R8

Segment Groups	$z = \frac{\bar{x} - \mu}{s}$	z_α	H_0 Test Result
Segment Group 1	-330.81345	-1.65	Reject
Segment Group 2	-297.79054	-1.65	Reject
Segment Group 3	-2.3179841	-1.65	Reject
Segment Group 4	-2.30174347	-1.65	Reject
All Segments	-0.00267262	-1.65	Accept

Fibonacci Rules;

- *F1: 3rd segment should be %161 or %261 of the 1st segment at price.*

$H_0: P_{4y} - (1.61 * P_{2y}) = 0$ *Or* $P_{4y} - (2.61 * P_{2y}) = 0$

$H_1: P_{4y} - (1.61 * P_{2y}) \neq 0$ *And* $P_{4y} - (2.61 * P_{2y}) \neq 0$

Each statement in null hypothesis is tested separately. Since the operator is logical"or", H_0 is accepted when either of the statements is in confidence interval. Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 26:

Table 26. Test Results for F1

Segment Groups	$z_{1.61} = \frac{\bar{x} - \mu}{s}$	$z_{2.61} = \frac{\bar{x} - \mu}{s}$	$z_{-\alpha/2}, z_{\alpha/2}$	H_0 Test Result
Segment Group 1	-746.09507	-7.54345	+-1.96	Reject
Segment Group 2	-750.740278	-755.967	+-1.96	Reject
Segment Group 3	-0.75419971	-0.75647	+-1.96	Accept
Segment Group 4	-0.74784103	-0.7523	+-1.96	Accept
All Segments	-7.6378905	-0.00077	+-1.96	Accept

- *F2: 5th segment should climb up %61, %100 or %161 of 1st segments end point.*

$H_0: P_{6y} - (.61 * P_{2y}) = 0$ *Or* $P_{6y} - P_{2y} = 0$ *or* $P_{6y} - (1.61 * P_{2y}) = 0$

$H_1: P_{6y} - (.61 * P_{2y}) \neq 0$ *And* $P_{6y} - P_{2y} \neq 0$ *and* $P_{6y} - (1.61 * P_{2y}) \neq 0$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 27:

Table 27. Test Results for F2

Segment Groups	$z_{0.61} = \frac{\bar{x} - \mu}{s}$	$z_0 = \frac{\bar{x} - \mu}{s}$	$z_{1.61} = \frac{\bar{x} - \mu}{s}$	$z_{-\alpha/2}, z_{\alpha/2}$	H_0 Test Result
Segment Group 1	757.0475902	30.17648	-738.822	+-1.96	Reject
Segment Group 2	754.7024335	132.7564	-748.791	+-1.96	Reject
Segment Group 3	0.752272801	0.080204	-0.74612	+-1.96	Accept
Segment Group 4	0.756806248	-0.00851	-0.73851	+-1.96	Accept
All Segments	7.56173E-05	-1.5E-09	-7.6E-05	+-1.96	Accept

Rule testing will be applied separately to each of "or" statements.

- *F3: 2nd segment should climb down %61 of the 1st segment at price.*

$$H_0: P_{3y} - (1 - .61)P_{2y} = 0$$

$$H_1: P_{3y} - (1 - .61)P_{2y} \neq 0$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 28:

Table 28. Test Results for F3

Segment Groups	$z = \frac{\bar{x} - \mu}{s}$	$z_{-\alpha/2}, z_{\alpha/2}$	H_0 Test Result
Segment Group 1	750.7963196	+-1.96	Reject
Segment Group 2	748.4227508	+-1.96	Reject
Segment Group 3	0.744456846	+-1.96	Accept
Segment Group 4	0.751173725	+-1.96	Accept
All Segments	7.6155405	+-1.96	Reject

- *F4: 4th segment should climb down %61 of the 3rd segment at price*

$$H_0: P_{5y} - (1 - .61)P_{4y} = 0$$

$$H_1: P_{5y} - (1 - .61)P_{4y} \neq 0$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 29:

Table 29. Test Results for F4

Segment Groups	$z = \frac{\bar{x} - \mu}{s}$	$z_{-\alpha/2}, z_{\alpha/2}$	H_0 Test Result
Segment Group 1	748.4226	+-1.96	Reject
Segment Group 2	747.9886479	+-1.96	Reject
Segment Group 3	0.751173725	+-1.96	Accept
Segment Group 4	0.752823464	+-1.96	Accept
All Segments	0.076157776	+-1.96	Accept

- *F5: 5th segment moves at least % 61 of the 4th segment at duration.*

$$H_0: (P_{6x} - P_{5x}) - (P_{5x} - P_{4x}) * 1.61 \geq 0$$

$$H_1: (P_{6x} - P_{5x}) - (P_{5x} - P_{4x}) * 1.61 < 0$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 30:

Table 30. Test Results for F5

Segment Groups	$z = \frac{\bar{x} - \mu}{s}$	$z_{-\alpha/2}, z_{\alpha/2}$	H_0 Test Result
Segment Group 1	-585.85114	+-1.96	Reject
Segment Group 2	-466.33941	+-1.96	Reject
Segment Group 3	-0.6293706	+-1.96	Accept
Segment Group 4	-0.5661040	+-1.96	Accept
All Segments	-0.00055904	+-1.96	Accept

- *F6: 5th segment should be at least % 61 of the 4th segment at price.*

$$H_0: P_{6y} - (1.61)P_{4y} \geq 0$$

$$H_1: P_{6y} - (1.61)P_{4y} < 0$$

Critical Values are determined using the significance level of %5. Whether null hypothesis is rejected or not rejected is given in Table 31:

Table 31. Test Results for F6

Segment Groups	$z = \frac{\bar{x} - \mu}{s}$	z_{α}	H_0 Test Result
Segment Group 1	-750.74028	-1.65	Reject
Segment Group 2	-757.7032	-1.65	Reject
Segment Group 3	-0.7478413	-1.65	Accept
Segment Group 4	-0.7480918	-1.65	Accept
All Segments	-7.6367605	-1.65	Reject

The summary of all rules towards all segments are given in Table 32:

Table 32. Summary of All Rules

	Segment Group 1	Segment Group 2	Segment Group 3	Segment Group 4	All Segments
R1	Accept	Reject	Accept	Accept	Reject
R2	Reject	Accept	Accept	Reject	Accept
R3	Accept	Accept	Accept	Reject	Reject
R4	Reject	Reject	Accept	Accept	Reject
R5	Reject	Reject	Accept	Accept	Accept
R6	Accept	Accept	Accept	Accept	Accept
R7	Reject	Accept	Reject	Accept	Accept
R8	Reject	Reject	Accept	Reject	Reject
F1	Reject	Reject	Accept	Accept	Accept
F2	Reject	Reject	Accept	Accept	Accept
F3	Reject	Reject	Accept	Accept	Reject
F4	Reject	Reject	Accept	Accept	Accept
F5	Reject	Reject	Accept	Accept	Accept
F6	Reject	Reject	Accept	Accept	Reject
# of Accepted Rules	3	4	13	11	8
# of Rejected Rules	11	10	1	3	6
Success Ratio	21%	29%	93%	79%	57%

When glared at the comparison matrix, with %5 significance level, segment group 3 has the maximum fit to Elliot rules.

The results of hypothesis tests show that Elliot wave pattern exists in ISE-National 100 index. Since the existence of Elliot Wave patterns in National-100 index controverts the random walk movements in stock prices, Istanbul Stock Exchange is not an efficient market.

CHAPTER V

CONCLUSION

In this study, Elliot Wave theory is tested on Istanbul Stock Exchange (ISE)- National-100 index by implementing time series pattern search techniques. The steps of Knowledge Discovery in Databases methodology are applied.

In order to test Elliot Wave theory, it is required to represent the data in a higher granularity. Thus, time series segmentation functionality of data mining is used. Two algorithms of piecewise linear representation technique, top-down and bottom-up algorithms, are implemented. In order to capture the implusive-corrective patterns of Elliot Waves, bottom-up algorithm is modified. Applying modified version of bottom-up algorithm, segments representing the National-100 index values are formed.

Since Elliot Wave theory does not specify the conditions where a wave should start or end, a window of 8-segment length is formed. This segment window is slided throughout the segments resulting in 4 segment groups that capture all possible Elliot Waves.

Elliot Wave rules, categorized as compulsory and Fibonacci rules are formulated as hypotheses. The hypotheses are then tested for each segment group using parametric hypothesis testing procedure.

The test results show that Elliot Wave pattern exists in Istanbul Stock Exchange- National 100 index. Since the presence of Elliot Wave patterns in National-100 index controverts the random walk movements in stock prices, the efficient market hypothesis is not applicable in Istanbul Stock Exchange and it can be concluded that Istanbul Stock Exchange is not an efficient market.

As future work, the study can be extended to other financial indicators such as sub-indices in stock market (financial, services, industrial, and technology indices), exchange rates, interest rates, company stock indices, etc.

REFERENCES

Agrawal, R., Faloutsos, C., & Swami, A. (1993). Efficient Similarity Search in Sequence Databases. *Proceedings of the 4th International Conference of Foundations of Data Organization and Algorithms (FODO)*, (pp. 69 – 84). London, UK.

Agrawal, R., Lin, K., Sawhney, H. S., & Shim, K. (1995). Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. *Proceedings of the 21st International Conference on Very Large Databases*, (pp. 490 – 501). Zurich, Switzerland.

Baumol, W. (1965). *The Stock Market and Economic Efficiency.* New York: Fordham University Press.

Çağatay, T., & Badur, B. (2007). The Similarity of Stock Prices and Indices in Progressing Information Technology Environment. *Proceedings of Picmet 2007*, (pp. 115-116). Oregon, USA.

Central Bank of the Republic of Turkey. (2008, August 15). Retrieved August 15, 2008, from Central Bank of the Republic of Turkey: http://tcmbf40.tcmb.gov.tr/cbt-uk.html

Chan, K., & Fu, W. (1999). Efficient Time Series Matching by Wavelets. *Proceedings of the 15th IEEE International Conference on Data Engineering*, (pp. 126-133).

Davidson, A. (2006). *Complete Guide to Online Stock Market Investing (2nd Edition).* London, UK: Kogan Page, Limited.

Dhar, V. (1998). Data mining in finance:Using counterfactuals to generate knowledge from organizational information systems. *Information Systems* , 423-437.

Droke, C. (2000). *Elliot Wave Simplified: Making Stock Market Profits with R.N. Elliott's Simple Theory.* Columbia, USA: Marketplace Books.

Edwards, R. (2007). *Technical Analysis of Stock Trends.* New York, USA: CRC Press.

Fama, E. F. (1970). Efficient Market Capitals: A Review of Theory and Emprical Work. *Journal of Finance* , 383-417.

Fishcher, R. (1941). *Fibonacci Applications and Strategies for Traders.* New York, USA.: John Wiley & Sons Inc.

Frost, A. J., & Prechter, R. R. (1978). *Elliot Wave Principle.* Georgia, USA: New Classics Library.

Ge, X., & Smyth, P. (2001). Segmental Semi-Markov Models for Endpoint Detection in Plasma Etching. *IEEE Transactions on Semiconductor Engineering.*

Han, J., & Kamber, M. (2001). *Data Mining: Comcepts and Techniques.* London: Academic Press.

Hecbert, P., & Garland, M. (1997). Survey of Polygonal Surface Simplification Algorithms. *Proceddings of the 24th International Conference on Computer Graphics and Interactive Techniques.*

Hill, J. (1979). *The Complete Wrintings of R.N. Elliott with Practical Application.* N. Carolina, USA.: Commodity Research Institute.

Hunter, J., & McIntosh, N. (1999). *Event Detection in Complex Time Series Data.* Edinburg, UK: Springer Berlin / Heidelberg.

Kargupta, H., Park, B., Hershbereger, D., & Johnson, E. (1999). Collective data mining: A new perspective toward distributed data mining. *Advances in Distributed Data Mining.* AAAI/MIT Press.

Keogh, E., Chu, S., Hart, D., & Pazzani, M. (2004). Segmenting Time Series: A Survey and Novel Approach. In M. Last, A. Kandel, & H. Bunke, *Data Mining in Time Series Databases* (pp. 1-21). Singapore: World Scientific Publishing Co. Pte. Ltd.

Keogh, E., Ratanamahatana, C. A., Lin, J., & Gunopulos, D. (2005). Mining time series data. In *The Data Mining and Knowledge Discovery Handbook* (pp. 1069-1103). USA.

Last, M., Kandel, A., & Bunke, H. (2004). *Data Mining in Time Series Databases.* Singapore: World Scientific Publishing Co. Pte. Ltd.

Li, C., Yu, P., & Castelli, V. (1998). MALM: A Framework for Mining Sequence Database at a Different Abstarction Levels. *Proceedings of the 9th International Conference on Information and Knowledge Management*, (pp. 267-272).

Mardia, K. W., Kent, J., & Bibby, J. (1979). *Multivariate Analysis.* London, UK: Academic Press.

Mazza, D. (2007, February). Beyond Elliott wave. *Futures* , pp. 32-34.

Neely, G., & Hall, E. (1990). *Mastering Elliott Wave.* New York, USA.: Windsor Books.

Person, J. L. (2004). *Complete Guide to Technical Trading Tactics : How to Profit Using Pivot Points, Candlesticks and Other Indicators.* Hoboken, NJ, USA: John Wiley & Sons, Incorporated.

Plummer, T. (2008). *Forecasting Financial Markets : The Psychology of Successful Investing (5th Edition).* London, UK: Kogan Page, Limited.

Powers, M. J. (2001). *Starting Out in Futures Trading (6th Edition).* Blacklick, OH, USA: McGraw-Hill Professional.

Prechter, R. R., & Frost, A. (1998). *Elliott Wave Principle: Key to Market Behavior.* West Sussex, UK.: John Wiley & Sons Inc.

Rafael, R. (2001). *Technical Analysis for Direct Access Trading : A Guide to Charts, Indicators and Other Indispensable Market Analysis Tools.* Blacklick, OH, USA: McGraw-Hill Professional Book Group.

Rafiei, D., & Mendelzon, A. O. (2000). Querying Time Series Data Based on Similarity. *IEEE Transactions on Knowledge and Data Engineering* , 675 – 693.

Satchwell, C. (2004). *Pattern Recognition and Trading Decisions : The Next Generation of Technical Analysis.* Blacklick, OH, USA: McGraw-Hill Professional Publishing.

Shatkay, H., & Zdonik, S. (1996). Approximate queries and representations for large data sequences. *Twelfth International Conference on Data Engineering*, (pp. 536-545). LA, USA.

Teseo, R. (2001, Oct). The Elliott-Fibonacci connection. *Futures* , pp. 52-54.

Wang, H., Yu, P., & Han, J. (2005). Mining Data Streams. In *The Data Mining and Knowledge Discovery Handbook* (pp. 777-792). USA.

Yi, B. (2007, January 28). *Similarity Search and Data Mining in Large Databases of Time Sequences*. Retrieved January 28, 2007, from http://www.postech.ac.kr/~bkyi/research.html, 2002.