A Temporal Pattern Identification and Summarization Method for Complex Time Serial Data

Saif Ahmad

Submitted for the Degree of
Doctor of Philosophy from the
University of Surrey

Information Extraction and Multimedia Group
Department of Computing
School of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, UK

May 2007

Contents

List of Figures

List of Tables

Abstract

Most real-world time series data is produced by complex systems. For example, the economy is a social system which produces time series of stocks, bonds, and foreign exchange rates whereas the human body is a biological system which produces time series of heart rate variations, brain activity, and rate of blood circulation. Complex systems exhibit great variety and complexity and so does the time series emanating from these systems. However, universal principles and tools seem to govern our understanding of highly complex phenomena, processes, and dynamics. It has been argued that one of the universal properties of complex systems and time series produced by complex systems is 'scaling'. The multiscale wavelet analysis shows promise to systematically elucidate complex dynamics in time series data at various timescales. In this research we investigate whether the wavelet analysis can be used as a universal tool to study the universal property of scaling in complex systems. We have developed and evaluated a wavelet time series analysis framework for automatically assessing the state and behaviour of complex systems such as the economy and the human body. Our results are good and support the hypothesis that 'scaling' is indeed a universal property of complex systems and that the wavelet analysis can be used as a universal tool to study it. We conclude that a system based on universal principles (e.g. 'scaling') and tools (e.g. wavelet analysis) is not only robust but also renders itself useful in diverse environments.

Key words: Complex systems, scaling, time series analysis, wavelet analysis.

Email: s.ahmad@surrey.ac.uk
WWW: http://www.computing.surrey.ac.uk/

1 Introduction

Signal processing has become an integral part of current scientific and technological activity. The signals that need to be processed exist in almost all realms of modern life. Signal processing is used in myriad fields like telecommunications, transmission and analysis of satellite images and medical imaging [92], [96], [159] – to name just a few. Many of these require analysis, transmission and synthesis of complex time series. The record of a stock price is a signal as is the record of temperature readings that could facilitate the study of global warming.

Most real-world signals are produced by complex systems. For example, the economy is a social system that produces time series of stocks, bonds, currency rates and so on. Similarly, the human body is a biological system that produces signals like *electrocardiograms* (ECGs), *electroencephalograms* (EEGs), and rate of respiration. Real data tend to be extremely irregular. The somewhat common characteristics of real-world time series produced by social, biological, and physical systems are: (1) the data produced by these systems is immense, that is the data we encounter is *high-frequency*; (2) the data does not follow well-established theoretical phenomena, that is the data may exhibit intricate phenomena like time varying volatility, nonstationarity, time changing probability distributions, long memory, and discontinuities; (3) *noise*, as a result of the artefact of the recording mechanism or other reasons could contaminate the data; and finally, (4) the data may not even be strictly in the form of a time series which is defined as a series of measurements at *equally-spaced* intervals of time. According to Yves Meyer, such signals often look like "complicated arabesques" [107] – tantalizing curves that contain all the information about the *state of a system* but conceal it from our comprehension.

In such a scenario, the search for useful analytical tools is paramount. Key issues that the putative analyst or system designer must consider are robustness of procedure to erroneous assumptions, the ability to handle complex relationships, and systemacity and simplicity of implementation. With this background, the potential promised by the wavelet multiscale analysis is readily apparent. While wavelets do not meet all the criteria required to assess the state and behaviour of complex systems, they meet some of them sufficiently well to indicate that they may well be able to provide new and better insights into the analysis of complex time series data produced by these systems.

One of the key benefits of a wavelet approach for analyzing complex time series is their flexibility in handling very irregular data series [34]. Given the complexity of real-world data, the ability of the wavelet transform to represent highly complex structures without knowing the underlying functional form is of great benefit in time series analysis research. Another important property that wavelets possess is that of being able to precisely locate *discontinuities* and *isolated shocks* to

the system. Furthermore, the wavelet representation is able to deal well with the nonstationarity and stochasticity of complex signals by providing a multiscale characterization of the local dynamics of these signals.

A standard assumption in traditional time series analysis is that the signal *f(t)* is *smooth* while the innovations $\varepsilon(t)$ are *irregular*. A natural first approach of extracting the signal *f(t)* from the observed signal $y(t) = f(t) + \varepsilon(t)$, is by locally smoothing *y(t)*. However, if a signal is more irregular than the *noise*, such a procedure may no longer provide a useful approximation of the signal. The process of *smoothing* to remove the contamination of *noise* may distort the appearance of the original signal. If the noise is below a threshold and the signal variation is well above the threshold then isolation of the signal from the noise component is possible by selectively shrinking the wavelet coefficient estimates [34], [35].

Perhaps, the most important property of wavelets for analyzing complex time series is their ability to decompose a signal on a *scale-by-scale* basis. Most social, biological, and physical systems contain variables that operate on a variety of timescales simultaneously so that the relationships between variables may well differ across timescales. The property of *scaling* (also known as 'self-organized complexity') in complex systems is a *universal property* [140], [153]. A detailed discussion on scaling in complex systems is provided in Chapter 2. The wavelet analysis shows promise to systematically elucidate complex dynamics in time series data at various timescales. In other words, the wavelet analysis can be regarded a *universal tool* to study the *universal property* of *scaling* in complex systems. Universal tools and principles (properties) are important in developing robust autonomous systems that can work well in uncertain and diverse environments. In this research we investigate whether it is possible to develop a robust autonomous system for time series analysis, based on universal principles and tools.

1.1 Aims

The major aim of this research is to develop an automatic procedure for analyzing complex time serial data. We have attempted to develop a procedure by taking into account three important aspects. First, the procedure should be based on a superior analytical method that can provide time series analysis without human assistance. Second, it should be based on insights derived from research in complex systems, time series analysis, statistics, and artificial intelligence (AI). Finally, it should have the robustness and capability to analyze signals across different domains. The research question we would like to address is this: *is it possible to build a system that can automatically analyze time series produced by various complex systems?* We will examine whether *scaling* actually exists as a universal property in complex systems. We will experiment with a variety of wavelet filtering methods to establish whether universal automatic procedures

may be used to analyze complex time series in various domains quickly, cheaply, and without the need of human intervention. We will also study how an effective analysis of time series emanating from complex systems can throw light on the state and behaviour of these systems.

To sum up, the specific aims of the present investigation are:

a) Experimentation with various complex time series to examine whether scaling is actually a universal property of these time series and systems.

b) Experimentation with a variety of wavelet filtering methods for analyzing univariate and multivariate time series data from different domains to determine the usefulness of wavelets as a universal tool for analyzing complex systems.

c) A computer implementation or the development of a wavelet framework that can be used for analyzing complex time series data across various social, biological, and physical systems.

1.2 Contributions

The major contribution of this research is the development of a robust interdisciplinary time series analysis system based on universal properties (scaling) and tools (wavelet analysis). Figure 1.1 summarizes the approach and philosophy for developing an interdisciplinary time series analysis system.

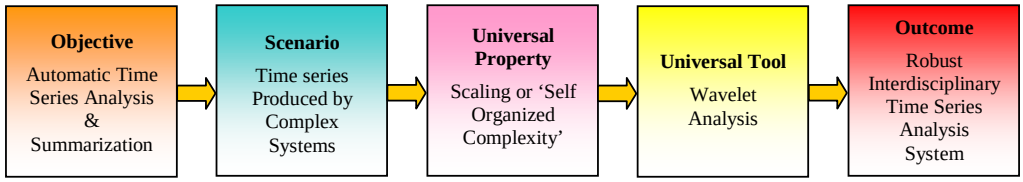


Figure 1.1 Approach and philosophy for developing an interdisciplinary time series analysis system.

This research claims to make the following specific contributions to the field of automatic time series analysis and artificial intelligence (AI):

1. It supports the hypothesis that scaling is indeed a universal property of time series produced by complex systems, and that wavelet analysis can be used as a universal tool to analyze these time series.

2. In addition, it supports the hypothesis that the decomposition of complex time series need not be based only on arbitrary criteria such as first differencing to remove the trend or a twelfth differencing to remove the seasonal component. The wavelet analysis can provide

a systematic (formal) and unique decomposition on a timescale basis to elucidate complex dynamics. Furthermore, the wavelet decomposition can be carried out automatically.

3. The development of a robust interdisciplinary time series analysis system which complements the role of subjective judgement, cost, and time, in univariate and multivariate time series analysis.

In *Appendix C*, I provide details of papers that I have authored and co-authored in peer-reviewed conference proceedings and professional magazines during the course of this research. I also provide details of the research projects that I have worked on and other related professional activities (*Appendix C*).

1.3 Thesis Outline

This thesis is organized into five chapters. This chapter (Chapter 1) gives a general overview of our research into analyzing high-frequency complex time serial data, especially in the presence of noise. In this chapter we also discuss the main aims and contributions of our research.

Chapter 2 establishes the concept of complex systems and how various time series analyses can help us study the behaviour of these systems. The concept of 'universal scaling' or 'self-organized complexity' in complex systems is explained in detail. A variety of univariate and multivariate time series analysis methods are reviewed and discussed.

In Chapter 3, we provide a detailed mathematical explanation of wavelet filtering and wavelet-based analysis of variance (ANOVA) and covariance. Specific examples are considered to throw light on each wavelet-based analysis and statistic. Finally, in Chapter 3, the problem of analyzing complex time series data is formulated.

Chapter 4 introduces the system architecture, explaining the detailed functionality of each module with relevant examples. Algorithms are presented to summarize the functionality of each module. Three case studies are performed to examine financial and medical data. This is followed by details of how our system was evaluated.

In Chapter 5 we present a discussion on the outcome of our research. We conclude Chapter 5 with a general assessment of the achievement of the aims set forth in Chapter 1, and consider directions for future work.

2 Motivation

In this chapter, we introduce the concept of complex systems and how various time series analyses can help us study the behaviour of such systems. Many of the systems that we see around us are complex – for example, social, biological or physical systems. Most, if not all, scientific inquiry aims to understand the properties of these systems. Although complex systems exhibit enormous complexity and variety, universal laws and phenomena seem to govern our understanding of their behaviour [7]. Complex systems consist of elements or parts that interact in a certain manner to result in formations. For example, inside proteins, amino acids interact through bonds to form protein foldings; in the nervous system, neurons interact through synapses to form learning; and in an economy, human beings interact through communication, confrontation, and cooperation to form economic evolution.

The Oxford English dictionary defines complex as "consisting of interconnected parts or involved particulars." The behaviour of a complex system is inherently dependent on its parts. However, this is not the case in simple systems – though they too are formed out of parts. The terms "interconnected" or "interwoven" are vital for differentiating between complex systems and simple systems. To understand the behaviour of a complex system, it is important to understand not only the behaviour of the parts but also how the parts act together to form the behaviour of the whole [52]. What makes a complex system difficult to understand is the fact that we cannot describe the system without describing each of its parts and the relationship of each part to the other parts. Some examples of complex systems are governments, families, the human body – physiological perspective, a person – psychosocial perspective, the brain, the ecosystem of the world, weather, a corporation, and a computer. On the other hand, a few examples of simple systems are an oscillator, a pendulum, a spinning wheel, and an orbiting planet.

The generic concept of "scaling" is important in understanding the behaviour of complex systems. From the discussion so far, we can intuitively say that parts or elements of a complex system are complex systems themselves. That is when the parts of a system are complex then the collection of the parts (the system) would also be complex. However, this may not always be the case.

Consider a complex system formed out of atoms, for example, the solid-liquid-gas transitions in condensed matter physics. The system itself is complex but the elements (atoms) that constitute it are simple systems. This is an important notion known as *emergent complexity* – many simple parts interact in a particular manner to give rise to a collective complex behaviour.

Conversely, we can also describe systems whose parts or elements are complex but the system itself is simple. The earth's revolution around the sun comprises a simple system although several

complex systems exist on the earth. This is the notion of *emergent simplicity* – a system composed of complex parts where the collective behaviour is simple. This example illustrates the concept of "scaling" in complex systems. The system may behave in a complex manner on a smaller scale but on a larger scale, all the complex details may not be relevant.

Articulating the behaviour of complex systems mathematically or quantitatively is crucial for understanding the universal properties of these systems. For the past several years, concentrated research in specialized fields has somewhat isolated individual disciplines of science. However, a computer scientist or a system designer must look for universal properties valid over many domains. Is it possible that disparate fields like molecular biology and economics be studied in a unified framework as one discipline? This is possible if we can adapt concepts from the field of complex systems for studying these disciplines. The field of complex systems breaks the barriers across traditional disciplines of science, engineering, management, and medicine by studying universal principles. Universal principles and tools guide and simplify the study of various complex systems that surround us – complex systems concepts provide a systematic guidance to approach many different problems. These concepts and principles are important for a computer scientist to develop robust autonomous systems for analyzing a range of complex systems.

2.1 Universal 'Scaling' in Complex Systems

Complex systems comprise elements or subunits that interact dynamically to result in complex behaviour. However, this *complexity* gives rise to a kind of *universality* that helps and guides the study of these systems. An important universal property of complex systems is "scaling", which has implications quite contrary to traditional Newtonian physics and calculus.

According to the physics of Newton and the related concepts of calculus, physical systems become simple, smooth and without detail at smaller and smaller (spatial and temporal) scales. This is true in the sense that though fine scale structure of planets, materials and atoms is not without detail, for many problems, such detail becomes irrelevant at the larger scale. Since the fine scale details are irrelevant, it is justified to formulate theories that assume that these details do not exist.

However, in complex systems theory, the concept of progressive smoothness on finer scales is not always a useful mathematical assumption – the fine scale details do matter. This introduces an important change in perspective from which we approach the study of various physical, social and biological systems. The concept of progressively increasing structure on finer and finer length scales leads us to the notion of fractals. Fractals are defined as geometric objects whose spatial structure is self-similar. This means that the structure on the coarsest scale is repeated on finer

length scales – by magnifying one part of the object, we find the same structure as that of the original object.

A pictorial representation of the characteristics of complex systems is shown in Figure 2.1 [110]. The dynamically interacting components of the system give rise to a number of scales and the components have similar shape (behaviour) at each scale. Even though the system is complex there is universality with regards to the scales or levels. This universal scaling in complex systems has also been termed as "self-organized complexity" [140], [153].

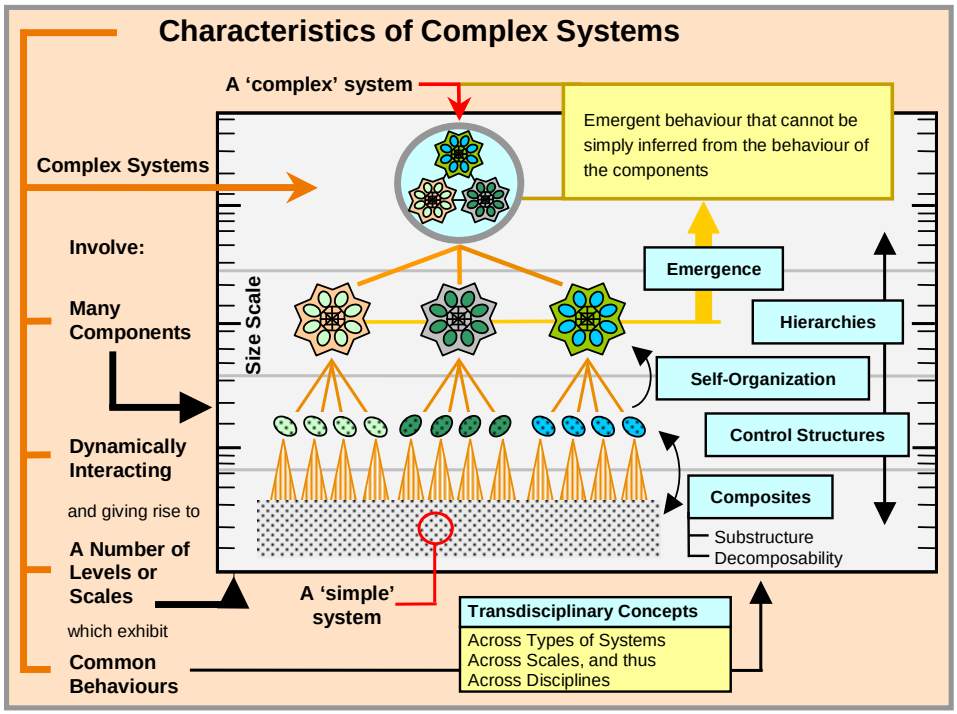


Figure 2.1 A schematic diagram showing the characteristics of complex systems. *Adapted from NECSI: www.necsi.org.*

Let a phenomenon be described by $f(x)$ which is some function of the scale x. According to Newtonian Physics, if there is a well defined length scale at which a particular effect occurs, then for longer length scales, the function $f(x)$ would typically decay exponentially,

$$f(x) \sim e^{-x/\lambda}. \tag{2.1}$$

This implies that the characteristic scale at which this property disappears is λ. However, according to complex systems theory, a system property is relevant over a large range of length scales, and hence it must follow a power law behaviour rather than an exponential behaviour,

$$f(x) \sim x^{\alpha}. \tag{2.2}$$

A function that follows a power-law behaviour defined by Eq (2.2), can also be characterized by the scaling rule,

$$f(Zx) \sim Z^{\alpha} f(x). \qquad 2.3$$

This implies that if we characterize the system on one scale, then on a scale that is larger by the factor Z it has a similar appearance, but scaled by the factor Z^{α}, where α is called the scaling exponent. Notably, as opposed to the behaviour of an exponential, for a power law, there is no particular length at which a system property disappears.

A real-world example of universal scaling or self organized complexity in complex systems is the frequency-size distribution of earthquakes. The mechanism of earthquakes is certainly a complex phenomenon yet it universally satisfies the relation,

$$N \sim S^{-b}, \qquad 2.4$$

where N is the number of earthquakes in a specified time interval and region with their rupture area greater than S. We note that Eq (2.4) is similar to Eq (2.2) and shows a power law behaviour between the number of earthquakes and their rupture area; the scaling exponent -b of Eq (2.4) is analogous to the scaling exponent α of Eq (2.2). This is known as the famous Guttenberg-Richter relation [65]. Other examples of universal scaling can be found in a number of time series originating in physical, biological and social systems. A time series is said to be persistent if its autocorrelation function does not decay at a sufficiently fast rate, that is, a time series in which adjacent values are positively correlated. One approach for quantifying persistence in a time series is to carry out a wavelet based analysis of variance. If the wavelet variance, V, has a power law dependence on the filter width (timescale) λ,

$$V \sim \lambda^{\beta}, \qquad 2.5$$

a time series is said to be a self *affine* fractal [149]. The term *affine* means a function with a constant slope but it distinguishes itself from a *linear* function by the fact that it may have a nonzero value when the independent variables are zero. For example, y = 2x is *linear* in x, whereas y = 2x + 7 is an *affine* function of x. A more detailed discussion on scaling and power law behaviour is provided in Section 2.2.5.

2.2 Univariate Time Series Analysis

The starting point for the study of complex systems is an understanding of the dynamical processes that constitute such systems. Most dynamical processes within complex systems are time dependent: an example from biology would be the variation in heart rate over time whereas an example from economics would be the change in the value of a stock price over time. A complex system is characterized by patterns of variation over time which contain valuable information about the overall state of the system [134]. Therefore, an effective analysis of time series emanating from complex systems can give us vital insights into the present and future

behaviour of these systems. The analysis of patterns of variation over time or time series analysis is performed on a series of data collected continuously or semi-continuously over time. For example, financial tick data, reported at irregular intervals of time may be rendered as a time series of equally spaced intervals by employing appropriate *compression* techniques [33], [54]. Similarly, a heart rate tracing, popularly known as an electrocardiogram (ECG), may be converted to a time series of intervals between consecutive heart beats [152].

Time series analysis can be broadly classified as univariate time series analysis and multivariate time series analysis. As the names suggest, univariate time series analysis deals with the study of a single series of observations recorded sequentially over time. It is assumed that other co-existing time series do not affect the behaviour of the time series or the system under study. On the other hand, multivariate time series analysis deals with the simultaneous study of two or more concurrently recorded series. It is assumed that patterns of variation in one time series affect the patterns of variations in the other series and vice-versa or the system as a whole. In multivariate analysis, the knowledge of the covariance of two or more time series is incorporated in the models built for studying various systems. It is interesting to note here that carrying out a univariate or multivariate analysis is the choice of the analyst and his assumptions about the system and series in question. For example, an analyst who is studying heart rate variations alone is carrying out a univariate analysis although there could be other recordings that might affect the heart-rate series and vice versa, for example, body temperature and blood pressure. This again brings up the notion of complex systems where dynamical processes are often interdependent. Therefore, multivariate analysis might generally prove more useful in the analysis of complex systems.

In this section, we will discuss various univariate time series analyses ranging from very simple and *ad hoc* real world techniques to the more complicated analytically and empirically well grounded techniques. In the next section we will discuss some multivariate time series analysis methods.

2.2.1 Real World Techniques

Real-world techniques for time series analysis comprise stand-alone systems and web-based services that are available to a wide group of users, enabling them to manage time serial data and to perform various kinds of analyses on the data. Most of these systems and services operate in the financial domain and provide a range of financial information and analysis methods. Some of the analyses afforded by these systems are automatic while others are carried out manually by experts and then presented to users.

A typical example of a stand-alone system is the Reuters Kobra cross market display application [129]. This application integrates and combines in one consistent environment both numeric and

textual financial data. Key automatic features in Reuters Kobra include Real-Time News, Real-Time Graphs, Reuters Adfin Analytics like Moving Averages, Bollinger Bands, and Volatility Calculations amongst many others. Then there are manual tools which can be used to better visualize and study a given time series. The user can draw trendlines, price channels, and support and resistance corridors to study stock price movements.

A screenshot of Reuters Kobra is shown in Figure 2.2. As we can see there are four objects inside the main screen (the user can insert as many objects as he likes). Two of the objects are graphics which show fluctuations in the British Pound to US Dollar exchange rate and the volume traded for the same. The user can visualize an instrument time series as a bar-chart, a candlestick chart or as a line plot. The third object is textual data, namely financial news headlines. The user can double click on the news headline of his interest to read the full story. The fourth object is fundamental data that gives information about entities like a company's turnover, dividends, and P/E ratios.

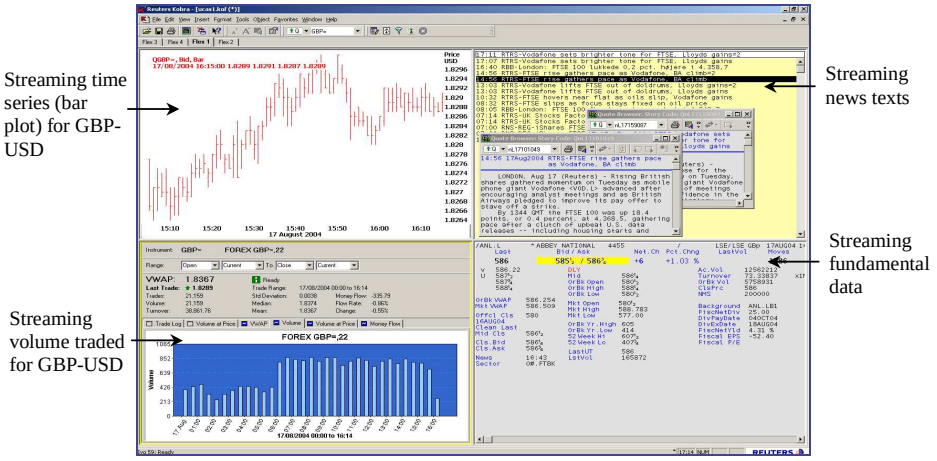
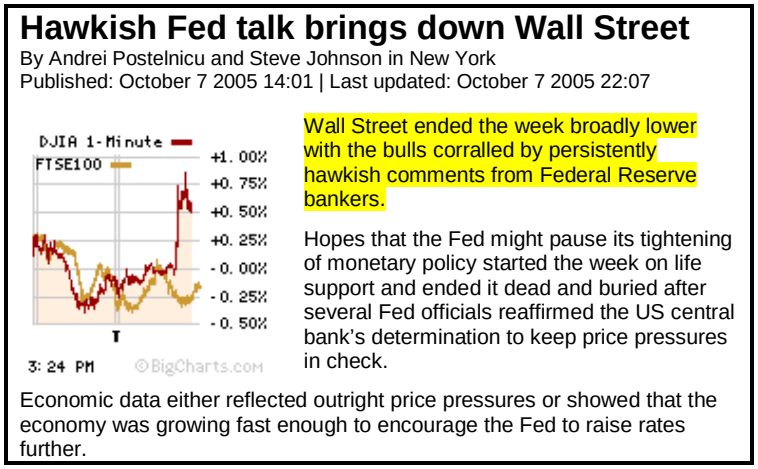


Figure 2.2 Screenshot of the Reuters Kobra interface for managing and visualizing financial Information (*GBP: British Pound, USD: US Dollar*).

A system like the Reuters Kobra is very useful for practicing traders as it provides them with real-time information (both financial news and time series data) along with the necessary analytics that enables them to react immediately to market events and hence make astute trading decisions.

Web-based services for time series analysis are generally provided by various research consultancies, newspapers, and other agencies like famous search engines. In financial sections of newspapers we often come across a news story that talks about the performance of the market and shows a time series as a "picture illustration" accompanying the text. An excerpt from Financial Times online is shown in Figure 2.3 [47]. The picture illustration shows the Dow Jones Industrial Average (DJIA) and the FTSE 100 time series while the accompanying headline and text talk about the performance of Wall Street. The chart has been marked up with a vertical line 'T'

indicating a structural break (variance change) or point on the Dow Jones time series after which the market showed a downtrend. The accompanying text describes the downtrend and the possible reason for it (highlighted yellow in Figure 2.3).



Hawkish Fed talk brings down Wall Street

By Andrei Postelnicu and Steve Johnson in New York

Published: October 7 2005 14:01 | Last updated: October 7 2005 22:07

Wall Street ended the week broadly lower with the bulls corralled by persistently hawkish comments from Federal Reserve bankers.

Hopes that the Fed might pause its tightening of monetary policy started the week on life support and ended it dead and buried after several Fed officials reaffirmed the US central bank's determination to keep price pressures in check.

Economic data either reflected outright price pressures or showed that the economy was growing fast enough to encourage the Fed to raise rates further.

Figure 2.3 Excerpt from Financial Times online.

Such an analysis seems to have been carried out manually by financial experts. The FSTE 100 time series has been plotted below the DJIA to facilitate a visual comparison between the performances of the two indices. Moreover, the ordinates of the chart show percentage changes in the value of the indices rather than their true values.

Yahoo® Finance, quite like the Reuters Kobra application discussed earlier, offers a whole range of financial data and technical analysis tools that can be invoked from a web interface [164].

Figure 2.4 Yahoo® Finance web interface for technical analysis.

A screenshot of the Yahoo® Finance web interface is shown in Figure 2.4. The chart shows the NASDAQ composite index in blue and its volume traded as a bar plot below it. The grey plots are the famous Bollinger Bands whose bandwidth is a measure of market volatility [17]. All the technical analysis methods available are nicely laid out above the chart and the user can click on any of them to carry out the desired analysis.

Yahoo® Finance also provides an analysis known as "Market Overview", which comprises two sections, namely Market Summary and Market Update [164]. An excerpt from Yahoo® Finance Market Overview is shown in Figure 2.5.

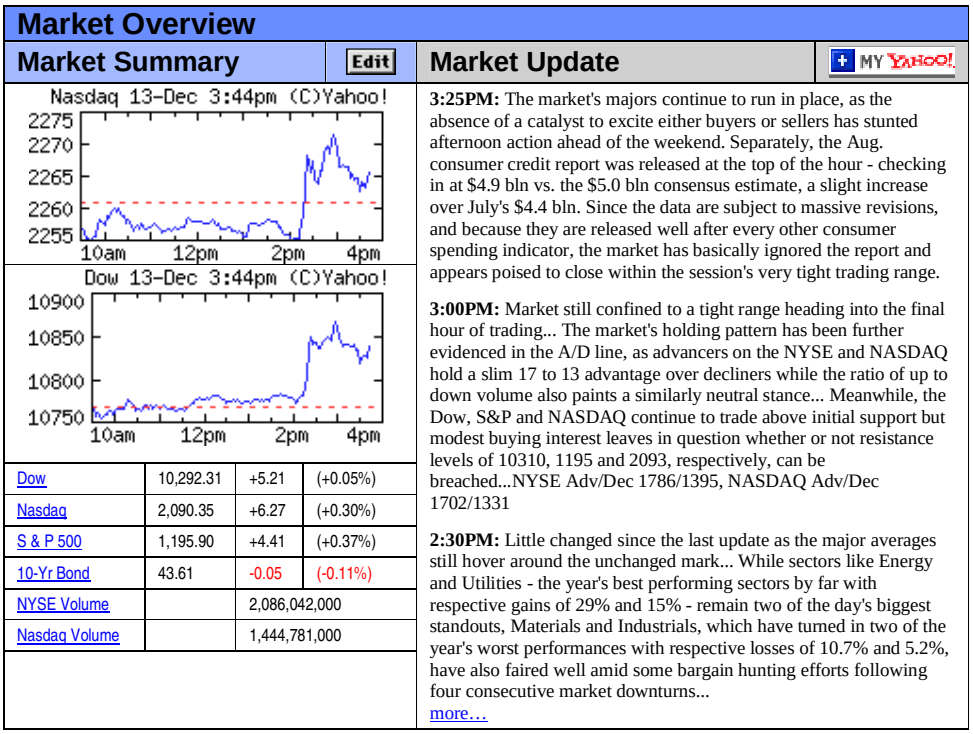


Figure 2.5 An Excerpt from Yahoo® Finance Market Overview.

The Market Summary section displays real-time graphs of major markets marked up with technical analysis indicators like major change-points and support and resistance lines. This section also displays the latest values of major instruments along with the changes from the previous values. The Market Update section provides a live technical analysis commentary about the fluctuations in the market graphs displayed under the Market Summary section. The commentary describes where in time major fluctuations occurred and discusses the possible reasons for these fluctuations. The commentary also predicts future market movements based on technical analysis carried out on the charts and other information like the release of consumer credit reports.

The Market Summary section in Figure 2.5 shows time series plots for NASDAQ and Dow Jones from 10:00 AM to 4:00 PM. The Market Update section shows a time indexed commentary about the movements in the NASDAQ and Dow Jones time series. The user can click on a "More" link to read commentaries going back to 10:00 AM, the opening time of the markets for the day. Again, such an analysis seems to be carried out manually by technical analysts and published on the web as and when events unfold in the marketplace.

Market technical analysis columnist Dale Woodson is editor of Woodson Wave Report, a web-based market-timing newsletter. Woodson Wave Report identifies turning point targets in the Dow, NASDAQ, and S&P 500 index as well as the bond and gold markets using Elliott Wave analysis and Fibonacci ratios [163].

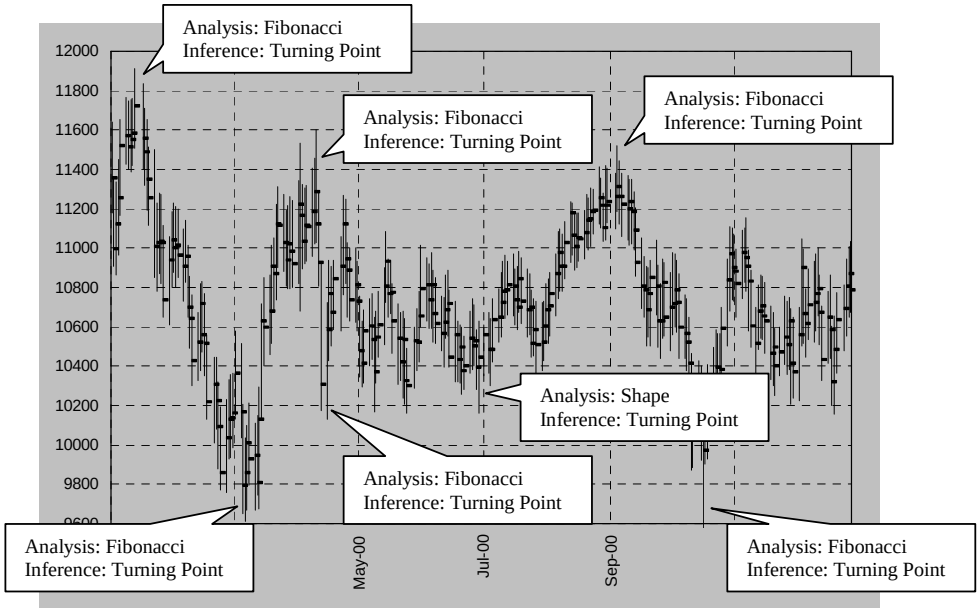


Figure 2.6 Excerpt from Woodson Wave Report on Dow Jones for 2000.

Since publishing the newsletter online in 1998, Woodson Wave Report has been downloaded in twenty-five different countries. Figure 2.6 shows an example of the Woodson Wave Report on Dow Jones Industrial Average for the year 2000. The user, when presented with such a labelling can quickly focus attention on the relevant portions of the graph and draw inferences about the original data. Such an analysis can be termed as knowledge based NLG (natural language generation) labelling to summarize features of interest in a financial time series. The knowledge-based analysis involves the use of Fibonacci ratios and Elliot wave theory to identify turning points while the NLG labelling involves the "marking-up" of key features on the time series graph as shown in Figure 2.6. In Figure 2.6, we just show the turning points identified and the analysis used. However, the actual NLG labelling contains long subjective descriptions about the turning points identified. An example of such descriptions can be found at

http://www.woodsonwave.com/keys.html. Therefore, this analysis seems to be carried out by hand.

A similar labelling of financial time serial data is shown in Figure 2.7. MarketVolume™ provides volume-based technical analysis of major stock market indices [101]. Apart from generating real time "Buy/Sell" recommendations, MarketVolume™ also provides information relating to the trend of the market.

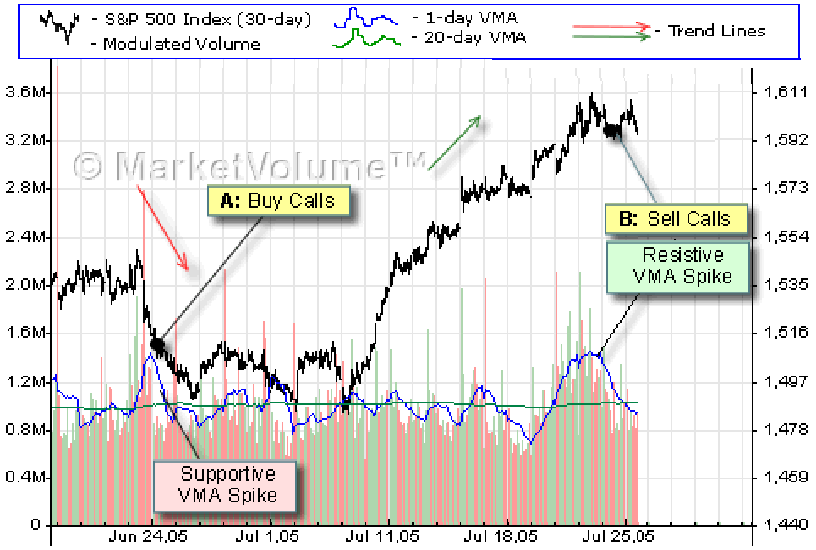


Figure 2.7 Excerpt from MarketVolume™ report for S&P 500 (June - July, 2005).

In Figure 2.7, the uptrend is marked with a green arrow whereas the downtrend is marked with a red arrow. The volume-based analysis shows a one-day moving average (blue) and a twenty-day moving average (green) of the volume and "Buy/Sell" signals that are generated based on the supportive and resistive spikes of these moving averages. It is not certain whether this service is manual or automated.

2.2.2 Conventional Approaches – Statistical Time Series Analysis

In this section we provide a survey of a few conventional approaches to time series analysis. The models discussed in this section can be classified as *probability models*, and *changing variance (non-linear) models* for time series analysis.

We begin by defining the so-called stochastic processes. Many physical processes involve a random element in their structure: a stochastic process can be described as "a statistical phenomenon that evolves in time according to probabilistic laws". For example, the length of a queue, the size of a bacterial colony or air temperatures on successive days at a fixed site, all constitute a stochastic process. The word *stochastic* is Greek in origin and means "pertaining to

chance". An important class of stochastic processes are those, which are *stationary*. From a statistical point of view, a time series is said to be *stationary* if there is no systematic change in mean (trend), if there is no systematic change in variance, and if strictly periodic variations have been removed.

ARMA Method

One of the most consistently successful time-series forecasting methods for the so-called stochastic processes is the Box-Jenkins autoregressive moving average (ARMA) method [18], [130]. Consider the observations $W(t_1)$, $W(t_2)$, ..., $W(t_N)$, obtained from a discrete time series at times t_1, t_2, ..., t_N, where $t_{i+1} - t_i$ is a fixed constant for $i = 1, 2, \ldots N\text{-}1$. The "stationarity" assumption implies that the probability distribution for the series is the same for any instant. For a stationary time series W_t, when the current value of the process is expressed as a finite, linear aggregate of previous values of the process and a shock a_t, where a_t is a purely random Gaussian process with zero mean and variance σ^2. That is,

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + ... + \phi_p W_{t-p} + a_t \qquad 2.6$$

Then W_t is called an AutoRegressive (AR) process of order p, where the coefficients $\phi_i : i=1 \ldots p$ are constants.

The Moving Average (MA) model of order q expresses the process W_t as a finite weighted sum of $a's$, where a_t is a purely random process with zero mean and variance σ^2. That is, for constant coefficients,

$$W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q} \qquad 2.7$$

To achieve greater flexibility in fitting of actual time series, both autoregressive (AR) and moving average (MA) terms can be included in the model. This leads to the mixed autoregressive moving average (ARMA) model. A mixed ARMA process containing p AR terms and q MA terms is said to be an ARMA process of order (p, q). It is given by,

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \phi_p W_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} - ... + \theta_q a_{t-q} \qquad 2.8$$

The importance of the ARMA process lies in the fact that it may often describe a stationary time series using fewer parameters than a pure MA or AR process by itself.

ARIMA Method

In practice most time series are nonstationary. In order to fit a stationary model, such as the ARMA model, it is necessary to remove nonstationary sources of variation. If the observed time series is nonstationary, then we can *difference* the series to make it stationary. This approach is widely used in economics. If we replace the original series W_t by $\nabla^d W_t$, in Eq (2.8) then we have a model capable of describing certain types of nonstationary time series. The operator ∇^d is

known as the *difference operator* where *d* is a positive integer that controls the number of times a series is differenced to make it stationary. In other words, *d* determines the number of differencing iterations. For example, if $d = 0$, the model is equivalent to an ARMA model, similarly, $\nabla^{d=1} W_t = (W_{t2} - W_{t1}), (W_{t3} - W_{t2}), \ldots (W_{tN} - W_{tN-1})$; $\nabla^{d=2} W_t = \{(W_{t3} - W_{t2}) - (W_{t2} - W_{t1})\}, \ldots \{(W_{tN} - W_{tN-1}) - (W_{tN-1} - W_{tN-2})\}$; and so on. The differenced series P_t is defined as,

$$P_t = \nabla^d W_t \qquad 2.9$$

where *d* is the number of differencing operations carried out to make W_t stationary. Such a model is called an *integrated* model because the stationary model, which is fitted to the differenced data, has to be summed or *integrated* to provide a model for the nonstationary data [26]. The autoregressive integrated moving average process (ARIMA) is of the form,

$$P_t = \phi_1 P_{t-1} + \phi_2 P_{t-2} + \phi_p P_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} - \ldots + \theta_q a_{t-q} \qquad 2.10$$

The second class of models that we will discuss in this section are called models of changing variance. These models are primarily concerned with modelling the changes in variance (or volatility). They lead to better estimates of the (local) variance, which allows more reliable prediction intervals to be computed and a better assessment of risk. This is very pertinent to financial time series when there is clear evidence of changing variance in the time plot of the data. We describe below the ARCH and GARCH models which are an important class of models of changing variance.

ARCH and GARCH Models

If a time series is treated as a sequence of random observations, then this random sequence or stochastic process may exhibit some degree of correlation from one observation to the next. This correlation structure can be used to predict future values of the process based on the past history of observations. If a correlation structure exists and it can be exploited, then a time series can be decomposed into two components, namely a deterministic component (i.e. the forecast) and a random component (i.e. the error or uncertainty associated with the forecast). Thus a time series y_t can be expressed as,

$$y_t = f(t-1, X) + \varepsilon_t, \qquad 2.11$$

where $f(t-1, X)$ represents the forecast or the deterministic component of the current value as a function of any information known at time *t-1*, whereas ε_t is the random component which represents the innovation in the mean of y_t. We can also interpret ε_t as the single-period ahead forecast error.

Let us assume that innovations ε_t are generated by the following mechanism,

$$\varepsilon_t = \sigma_t z_t, \qquad 2.12$$

where z_t denotes a sequence of random variables with zero mean and unit variance and σ_t is the local conditional standard deviation of the process. If we further assume that the square of the conditional standard deviation of a process (σ_t) depends on the previous value of the innovation by the following relation,

$$\sigma_t^2 = \kappa + \sum_{j=1}^{Q} A_j \varepsilon_{t-j}^2 , \qquad 2.13$$

where κ and A are chosen in a way to ensure that σ_t^2 is non-negative, then such a model is known as the *autoregressive conditionally heteroscedastic model* of order (Q). The model of Eq (2.13) is abbreviated as ARCH (Q) and was developed by Robert Engle [40], for which he won the 2003 Nobel Prize in Economics.

Generally, ε_t^2 tends to have a relatively slowly decaying autocorrelation function, especially for high frequency return series like daily or weekly series of financial instruments. This would mean that modelling such a series with the ARCH model would require a long lag Q and hence a large number of parameters. However, if the right-hand-side of Eq (2.13) is modified by adding lags of the conditional variance σ_t^2, the resulting model can be formulated with only a small number of parameters and it can still display a slowly decaying autocorrelation function for ε_t^2. Such a model was introduced by one of Robert Engle's graduate students Tim Bollerslev [16], and has the following form,

$$\sigma_t^2 = \kappa + \sum_{i=1}^{P} G_i \sigma_{t-1}^2 + \sum_{j=1}^{Q} A_j \varepsilon_{t-j}^2 . \qquad 2.14$$

The model of Eq (2.14) is known as the *generalized ARCH* model of order (*P, Q*) and abbreviated as GARCH (P, Q). Owing to its simplicity, the first order (P = Q = 1) GARCH model, GARCH (1, 1), has over the years become the most popular ARCH model in practice and is described by the following equations,

$$y_t = C + \varepsilon_t, \qquad 2.14a$$

$$\sigma_t^2 = \kappa + G_1 \sigma_{t-1}^2 + A_1 \varepsilon_{t-1}^2 . \qquad 2.14b$$

We note that the GARCH (1, 1) model described by Eq (2.14a) and Eq (2.14b) requires just four parameters (C, κ, G_1, and A_1) for modelling the conditional variance of a time series. Eq (2.14a) is known as the conditional mean model where the return y_t consists of a sample constant (mean) C, plus an uncorrelated white noise disturbance ε_t. Eq (2.14b) is known as the conditional variance model where the variance forecast σ_t^2 consists of a constant κ plus weighted average of the last period's forecast σ_{t-1}^2 and the last period's squared innovation ε_{t-1}^2.

Figure 2.8 shows a GARCH (1, 1) simulation for the *Deutsche Mark - British Pound* daily return series from January 2, 1984, to December 31, 1991. The returns y_t, are computed using the formula, $y_t = 100ln\ (p_t / p_{t-1})$, where $\{p_t, t = 1, \ldots, N\}$ is the *Deutsche Mark - British Pound* exchange rate time series for the given dates. Plot A is the input return series while Eq (A) and Eq (B) are GARCH (1, 1) equations with the calculated parameters for the input return series y_t. Comparing from Eq (A) and Eq (B) in Figure 2.8 with Eq (2.14a) and Eq(2.14b) we note that C = -6.20e-005, $\kappa = 1.076e\text{-}006$, $G_1 = 0.81$, and $A_1 = 0.15$ for the *Deutsche Mark - British Pound* exchange rate returns time series.

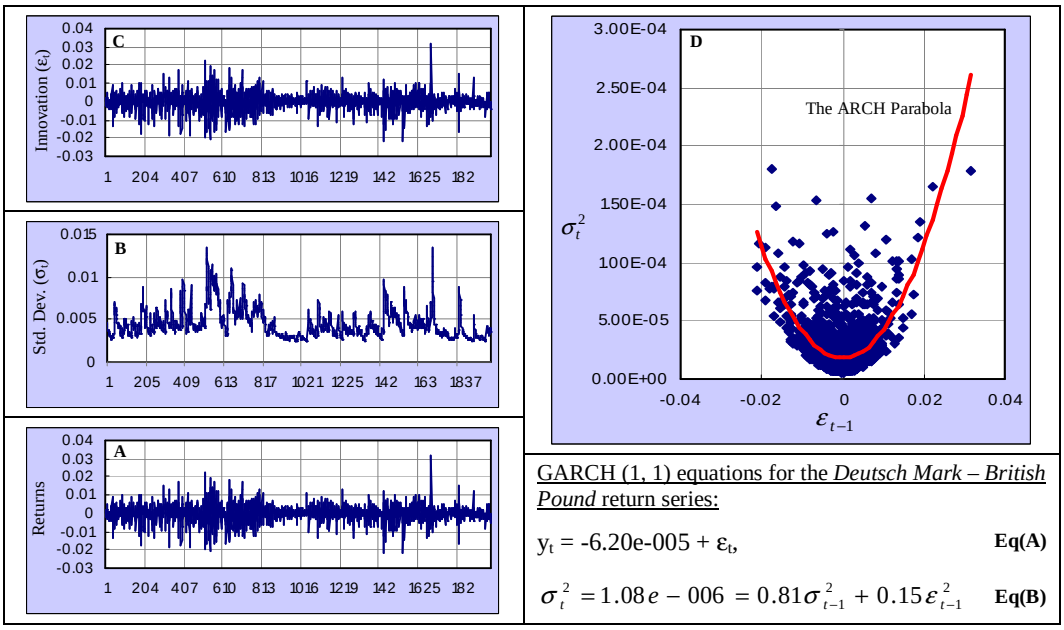


Figure 2.8 GARCH (1, 1) simulation for the *Deutsch Mark – British Pound* return series.

In Figure 2.8, Plot C is the innovation, while Plot B is the standard deviation which are calculated using Eq (A) and Eq (B) respectively. We notice a striking similarity between Plot A (returns) and Plot C (innovations). This is because according to Eq (2.14a), $\varepsilon_t = y_t - C$, where C is the mean of the return series y_t. In this GARCH simulation C = -6.20e-005 which is an extremely small number and hence subtracting it from y_t does not alter the value of y_t significantly.

Plot D of Figure 2.8 is a plot of variance (σ^2_t) at time t versus the innovation (ε_{t-1}) at time t-1: this produces a parabolic curve also called an ARCH and hence the name of the method. Plot D in Figure 2.8 is also known as the *news impact curve* [41]. Robert Engle hypothesized that price returns y_t predominantly evolve through a random mechanism which he called innovations or ε_t (Eq (2.11), and Eq (2.14a)). He attributed the randomness or innovation in price returns to incoming financial news announcements which could randomly constitute "good news" or "bad news" for the market. Negative innovations (or in turn negative price returns) meant downturns in

the actual price time series which resulted from "bad news" whereas positive innovations meant upturns in the price time series which resulted from "good news."

According to Eq (2.14b) and Plot D, Figure 2.8, standard deviation of a stock return at time *t* is related to the innovation at time *t-1* or in other words to the "good news" or "bad news" at time *t-1*. The parabolic curve of Plot D, Figure 2.8, is called the *news impact curve* because it helps us study the impact or effect of financial news on the future market movement or standard deviation. If the parabolic curve (news impact curve) is perfectly *symmetric*, then the "good news" (positive ε_t) and the "bad news" (negative ε_t) have equal impact on the market movement or cause an equal change in standard deviation of future price returns. However, if the news impact curve is asymmetric, that is, it is either skewed to the right or left then either the "good news" or "bad news" has more impact on the market movement. In the example that we have studied in Figure 2.8, the news impact curve is skewed towards the right implying that "good news" has had more impact on the market movement.

2.2.3 NLG Approaches

In this section we discuss Natural Language Generation (NLG) approaches for analyzing time series. Such systems first analyze a given time series with sophisticated time series analysis methods followed by a description in natural language of the results of the analysis. These systems have been termed as *time series summarization systems* since their output comprises a *natural language summary* describing the behaviour of a given time series. NLG is the subfield of Artificial Intelligence (AI) and computational linguistics. It focuses on computer systems that can produce understandable texts in English or other human languages [128], [105]. Figure 2.9 summarizes how typical NLG systems operate.

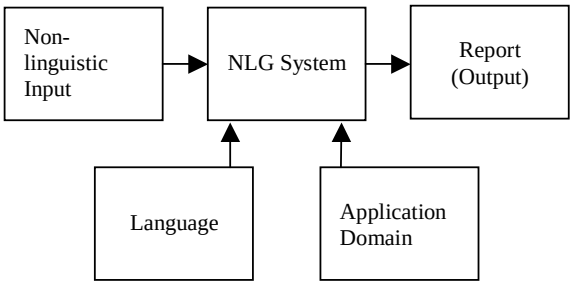


Figure 2.9 Typical NLG system.

The input to such systems consists of some non-linguistic representation of information and they use knowledge about language and the application domain to automatically generate reports and other kinds of texts.

NLG based time series summarization systems perform tasks such as producing weather reports from weather simulations, or summaries of stock market fluctuations. The key goal of research in this area is to develop better technology for producing natural language summaries of time-series data by integrating leading-edge time-series and NLG technology.

SUMTIME-MOUSAM

The SUMTIME project at Aberdeen University aimed at producing "human-like" summaries of time series data using *knowledge acquisition* (KA) techniques [138], [139]. SUMTIME-MOUSAM is a time series summarization system developed as part of the SUMTIME project. The architecture of the SUMTIME-MOUSAM system is shown in Figure 2.10.

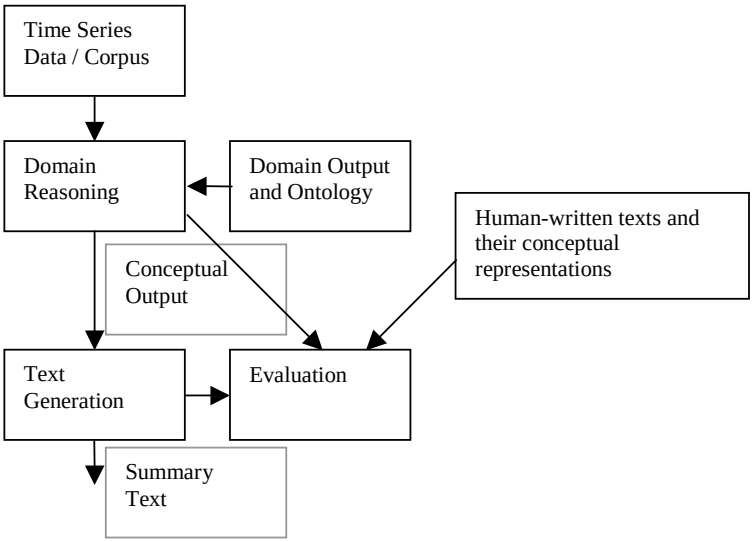


Figure 2.10 Architecture of SUMTIME-MOUSAM.

Input data to the system comprises 40 basic weather parameters like wind direction, wind speed, and gust. Time series analysis is carried out using *linear segmentation*. Segmentation refers to the process of approximating a time series of length *n* with *K* straight lines, where *K* is much smaller than *n*. The approximated signal is known as a *piecewise linear representation* of the input time series [76], [88], [89]. *Forecast* is based on *linear interpolation* of the *segmented* time series.

SUMTIME-MOUSAM also performs microplanning and realization to generate natural language text about various weather parameters and forecasts. The microplanner uses numerical information from the *Segmentation Model* to produce texts about the weather parameters.

SYSTEM TREND

System Trend, which was developed at the Microsoft Research Institute, Australia, uses advanced signal processing techniques like the Continuous Wavelet Transform (CWT) to computationally

detect edges in a given signal. This information is then used to articulate the *trend* in the signal in natural language [19]. The system architecture of TREND is shown in Figure 2.11.

The continuous wavelet transform (CWT) is defined as follows:

$$CWT_x(\tau, a) = \frac{1}{\sqrt{a}} \int x(t) \psi \left(\frac{t - \tau}{a} \right) dt \qquad 2.15$$

The transformed signal is a function of 'τ' (the translation or the position) and 'a' (the scale). '$\Psi(t)$' is the transforming function and is called the *mother* wavelet (also called *prototype* wavelet). High scales correspond to a non-detailed global view (of the signal), and low scales correspond to a detailed view. System TREND uses the derivative cubic spline wavelet as the *prototype* or *mother* wavelet, which is an effective edge detector [102].

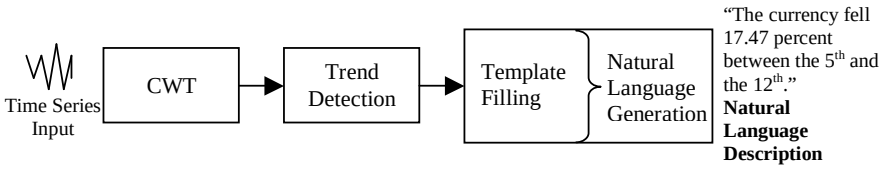


Figure 2.11 System architecture of TREND.

The algorithm for trend detection combines the special properties of the derivative cubic spline wavelet and a theory used in image processing called *scale-space* theory. An algorithm (based on scale-space theory) selects those trends whose wavelet transform extrema are present at several scales and have a higher combined absolute value.

Once the trends have been computationally detected, the numerical values are mapped into verbs and adverbs, which are inserted into a natural language template suitable for natural language generation. Natural Language is generated using the FUF/SURGE, which is a syntactic realization front end for natural language generation systems equipped with a comprehensive grammar of English [39].

```
During 1997, the currency fell 17.47 percent to finish the year at 0.651. It
remained mainly unchanged between the 20th of February and the 30th of May and
decreased considerably between the 30th of May and the 7th of July before
staying mainly unchanged until the 10th of September. It fell dramatically
between the 10th of September and the 31st of December.
```

Figure 2.12 Output from system TREND.

The input to TREND consists of an "annual currency file", which consists of a number of lines each containing the date and the daily currency value. Figure 2.12 shows a sample output from system TREND. The input data in this case comprised the value of the Australian dollar measured against the U.S. dollar during the year 1997.

2.2.4 Artificial Intelligence (AI) / Soft Computing

Time series emanating from complex systems are highly nonlinear since they are a manifestation of many interdependent processes. Therefore, conventional approaches for understanding and predicting the behaviour of such time series based on analytical techniques alone could prove to be difficult. The conventional analytical techniques for the analysis of complex time series are perhaps too specific and inflexible to be able to cope with the intricacies and complexities of real world systems. Models operating in such a scenario should be able to tolerate a high degree of uncertainty and imprecision.

Soft computing is a new branch of computer science, based on fuzzy logic, neural networks and probabilistic reasoning. Soft computing or AI models have the properties of "approximation" and "dispositionality" which makes them tolerant to imprecision and uncertainty [86]. Such a tolerance is not achieved by the so-called hard computing models since they are based on binary logic, crisp systems, numerical analysis and crisp software.

Generally speaking crisp logic is the type of logic that is outside the field of AI. Crisp logic deals with absolute set membership. An item is either a member of the set or not. There is no in-between. For example, membership in the sets of "cat" and "man" is absolute. One single animal (specie) cannot both be a cat and a man.

In soft computing the tolerance for imprecision and uncertainty is exploited for achieving robustness and low cost solutions. Two soft computing methods that have extensively been used for time series analysis and prediction are neural networks and fuzzy logic. More recently, owing to the great complexity of real-world time series data, multiscale wavelet analysis seems to have found a place as a powerful pre-processing tool for soft computing models giving rise to the so-called *hybrid* models [151]. In this section we provide a brief survey of neural network, fuzzy logic, and wavelet-soft computing hybrid models for time series analysis.

Several neural network architectures have been studied for time series analysis in disciplines ranging from economics and hydrology to structural engineering. The most popular architecture seems to be based on multilayer perceptrons (MLPs) which are feed-forward neural networks trained with standard backpropagation algorithms [11], [106], [95], [44], [143], [72], [91], [81]. Other architectures that have been examined for time series analysis include recurrent networks which comprise a network of neurons with feedback connections [49] and radial basis functions (RBF) which are based on the theory of functional approximation [82], [83], [77], [46].

Fuzzy logic time series modelling methods are broadly classified into those using *complex rule generation mechanisms* and *ad hoc data-driven models* [25] for automatic rule generation. The former employ a mixture of methods, like neuro-fuzzy [85], [80] and probabilistic-fuzzy [155]

methods, while the latter utilize data covering criteria in example sets [158], [112]. *Ad hoc* data-driven models have the advantages of simplicity, speed and high performance and often serve as preliminary models that are subsequently refined using other methods [25], [115]. Moreover, since *ad hoc* models employ automatic rule generation, fuzzy rule bases can be built and updated with minimal effort.

Both neural networks which employ connectionist regressions and fuzzy logic models which employ fuzzy-rule bases to model time series data can be termed as global approximators where only one model is used to characterize an entire process. However, real-world time series data is inherently nonstationary and may be a superposition of many sources exhibiting different dynamics. Hence there is a need for developing hybrid soft computing models that can deal effectively with complex real-word data. Recently, there has been an increased interest in multiresolution decomposition techniques like the wavelet transform for elucidating complex relationships in nonstationary time series [118]. The wavelet transform can produce a good local representation of a signal in both time and frequency domain and is not restrained by the assumption of "stationarity" [97]. Motivated by the spatial frequency resolution property of the wavelet transform, several hybrid schemes (local models) have been developed, which combine wavelet analysis with soft computing approaches like neural networks and fuzzy logic for time series prediction [165], [4], [5], [166], [73].

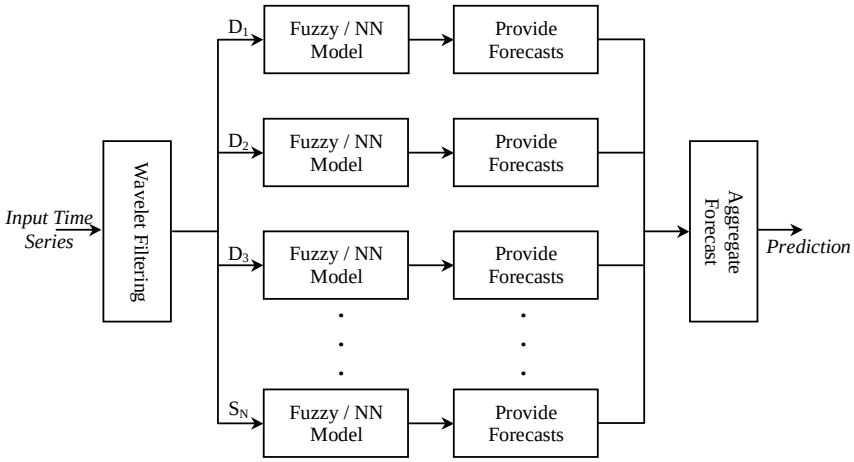


Figure 2.13 Overview of wavelet / soft computing multiresolution forecasting systems. The Ds are wavelet coefficients while S_N is the signal "smooth" or "trend".

A typical wavelet-soft computing hybrid prediction scheme is shown in Figure 2.13. Given a time series *X(n), n = 1, …, N*, the aim is to predict the *lth* sample ahead, *X(N+l)*, of the series. That is *l = 1* for single step prediction. This scheme basically involves three stages. In the first stage, the time series is decomposed into different timescales using the wavelet transform. In the second stage, each wavelet scale is predicted by a separate fuzzy logic model or neural network

architecture, and in the third stage the individual predictions at each scale are combined to generate an aggregate forecast. We have published two papers in which we show improvements in the prediction capability of fuzzy models when they work on wavelet pre-processed data [122], [123].

2.2.5 Empirical Techniques – Power Laws

Power laws are empirical laws that describe the dynamics of a variety of disparate phenomena ranging from earthquakes to solar flares and from stock market fluctuations to avalanches. It is believed that these dynamics arise from the system itself – the theory of "self-organized complexity" has been suggested to represent a universal organizing principle in biological, social and physical systems [140], [153], [6].

Let us revisit the frequency distribution of earthquakes, which we described in Eq (2.4) earlier. According to Eq (2.4), which is also known as the Guttenberg-Richter relation, N is the frequency of earthquakes and S is their rupture area. The magnitude m of earthquakes is defined in terms of their rupture area S by the following relationship:

$$m = log_{10} S. \tag{2.16}$$

The magnitude m is popularly known as the Richter scale and generally an earthquake measuring six or above on the Richter scale is considered as large. Eq (2.16) can also be expressed as,

$$S = 10^{m}. \tag{2.17}$$

Replacing S from Eq (2.17) in Eq (2.4) gives us,

$$N \sim 10^{-bm} \tag{2.18}$$

Introducing an equality sign in Eq (2.18) would give us,

$$N = k10^{-bm} \tag{2.19}$$

where k is a constant. Taking log of Eq (2.19) gives us,

$$log_{10} N = log_{10} k - bm, \tag{2.20}$$

where $log_{10} k$ is a constant and can be replaced by a,

$$log_{10} N = a - bm, \tag{2.21}$$

Eq (2.21) represents a negative sloping straight line of the form $y = a - bx$ where b is the slope of the line and a is the y-intercept.

The only difference in Eq (2.21) is that the y-axis is logarithmic. Hence, if we plot Eq (2.21) on a logarithmic-linear graph, we will get a negative sloping straight line. Empirical results show that a plot of the magnitude of earthquakes (m) against the log of the frequency of their occurrence

($log_{10} N$) reveals a straight line with a negative slope of about -1. This provides a measure of the probability of an earthquake occurring in a given region over a certain period of time. In areas of increased earthquake activity, the straight line is shifted to the right but the slope remains the same. Thus, the difference between the *y intercepts* of two straight lines for two regions provides a measure of the difference in probabilities of an earthquake of all magnitudes occurring in these regions.

Figure 2.14 shows a log-linear plot of the seismic activity in Southern California during 1987-1996 and the average seismic activity in the world for one year [38]. The slopes of the lines describing the two seismic activities are equal to *-1*, that is $b = 1$. However, the line for world seismicity is shifted to the right as compared to the line for Southern California seismicity, indicating a higher possibility of earthquakes all around the world as compared to the possibility of earthquakes in Southern California alone. Power law behaviour has also been studied and described in physics, economics, biology, ecology, and evolution [36], [57].

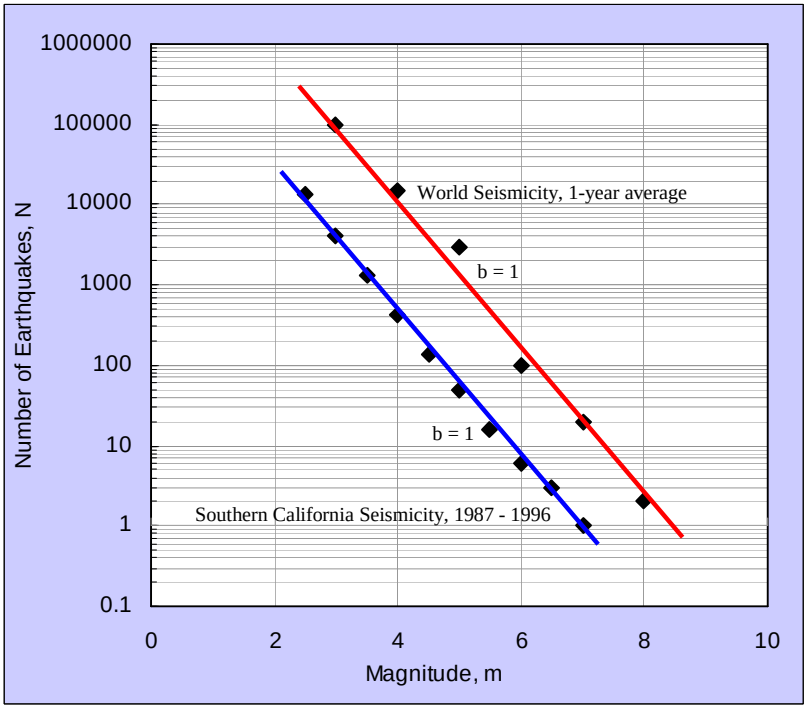


Figure 2.14 World (average for one year) and Southern California (1987-1996) seismicity counts vs. magnitude.

In time series analysis, power laws describe dynamics that have a similar pattern at different timescales; this implies that they are *scale invariant*. A power law describes a time series with a large number of small variations and a smaller and smaller number of large variations. Moreover, the pattern of variation is statically similar regardless of the size of the variation. Magnifying or shrinking the scale of the signal reveals the same statistical relationship that describes the

dynamics of the signal being analyzed. This scale invariant self-similar behaviour is analogous to the property of fractals which are geometric structures first investigated by Benoit Mandelbrot [99]. As discussed in Section 2.1, fractals are defined as geometric objects whose spatial structure is self-similar. This means that the structure on the coarsest scale is repeated on finer length scales – by magnifying one part of the object, we find the same structure as of the original object. Similarly, with regard to time series the patterns of variation appear the same at all timescales – magnifying a portion of a pattern reveals the same pattern [58]. Such a scaling is often referred to as fractal scaling. The power law allows us to measure the long-range correlations that are present in a series of data and alterations in these correlations provide us with a means of assessing the state of a system.

The first step in the evaluation of the power law is the calculation of the Fourier power spectrum [20]. A Fourier power spectrum reveals the various frequency components of a time series. If we plot a log-log representation of a Fourier power spectrum (that is log power vs. log frequency), we obtain a straight line with a slope of approximately -1. This implies that as the frequency increases, the size of the variation (fluctuations) drops by the same factor and this behaviour exists across many scales or frequencies.

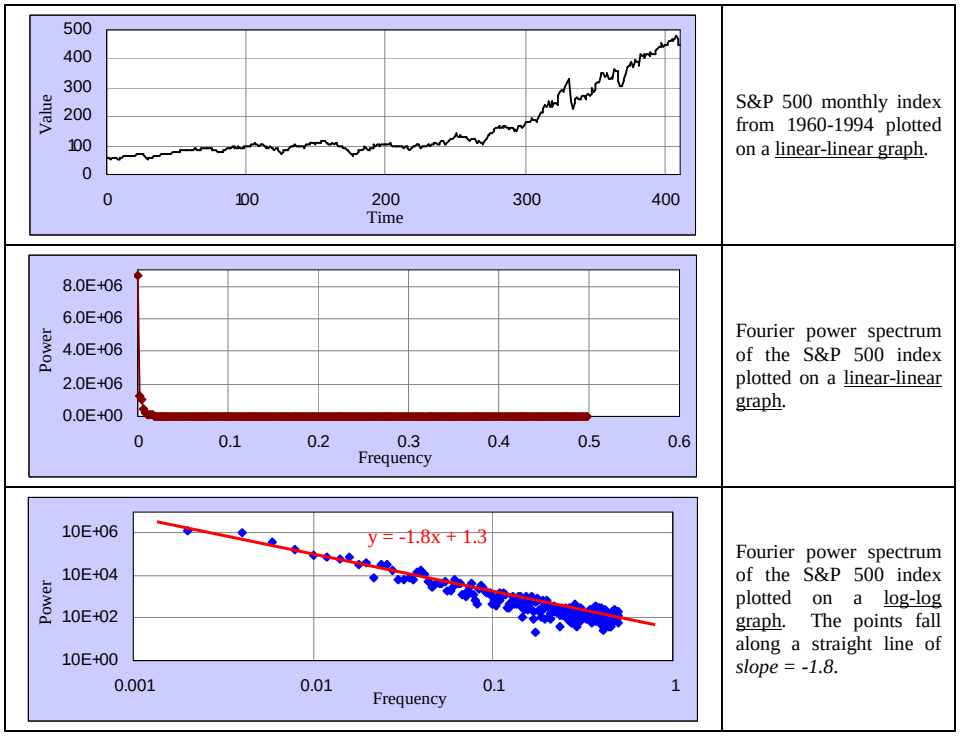


Figure 2.15 Demonstration of power law behaviour for S&P 500 time series using Fourier power spectrum analysis.

In Figure 2.15, we demonstrate the existence of power law behaviour in S&P 500 time series using Fourier power spectrum analysis. The power for very low frequencies is quite high and it

rapidly decreases for higher frequencies (dark red coloured plot in Figure 2.15). The log-log plot of the Fourier power spectrum reveals a straight line of slope -1.8 (blue coloured plot in Figure 2.15) – as the frequency increases, the size of the variation (power) drops by the same factor and this behaviour exists across many scales or frequencies. As per Eq (2.3), power law behaviour is scale invariant – if a variable x is replaced by Ax', where A is a constant, the fundamental power law relationship remains unchanged. A straight line is fitted using linear regression.

Power law behaviour has been observed and studied in biological, economic and physical systems – a change in the intercept and slope helps to assess the state and stability of these systems. In biology, power law behaviour describes fluctuations in heart rate which was first studied by Kobayashi et al. [90], foetal respiratory rate in lambs [148], movement of cells [42] and so on. For heart rate variability, it has been observed that a more negative slope is obtained for aged patients [121] and patients with coronary artery diseases [84]. In patients with recent myocardial infarction, the slope could act as a predictor of all-cause mortality [12]. In marine sciences, power law behaviour can be used to study shoreline changes [149] and in economics it can be used to study "risk, ruin and reward" [100], [87].

There is however one limitation in determining the power law by Fourier spectral analysis in that the technique becomes problematic when applied to nonstationary signals. This limitation might give spurious results regarding the underlying dynamics of the process under study. Wavelet multiscale analysis, which is closely related to Fourier analysis and is capable of dealing with nonstationary signals, seems to be a good alternative to study power law behaviour in real-world time series data. Wavelet-based *analysis of variance* (*ANOVA*) as a function of timescale has proven to be a useful tool for studying power law behaviour in nonstationary time series emanating from complex systems [149], [118], [152]. In Chapter 3, we will discuss in more detail the calculation of the power law using Fourier and wavelet analysis.

2.3 Multivariate Time Series Analysis

We discussed earlier that complex systems comprise parts or elements that interact in a certain manner to give rise to the overall behaviour of the system. We also discussed how it is important to study the interrelationships amongst parts of a complex system to assess its state. In a complex system there may be parameters that change simultaneously to give rise to the overall behaviour of the system. For example, in meteorology, parameters like temperature, air pressure, rainfall etc., may generate a time series each at the same site for the same sequence of time points – these parameters might help us assess the state of the weather in that region. In economics, many different time series measuring various forms of economic activity are recorded at regular intervals and may throw light on the state of an economy. Faced with multivariate data, it is

worthwhile to develop multivariate models that can deal with interrelationships amongst such series to better understand the behaviour of a system.

In this section we will discuss three multivariate time series analysis methods namely, Value at Risk (VaR), Capital Asset Pricing Model (CAPM), and Cointegration. These three models have been extensively used in economics and finance for risk management and prediction. The basic theory behind these multivariate models is the assumption that financial time series do not vary in isolation and that price movements and fluctuations are indeed correlated. More specifically, the VaR model and the CAPM studies stock price correlations to assess financial risk whereas co-integration studies the co-movement of two or more nonstationary time series for prediction purposes.

2.3.1 Value at Risk (VaR) Model

In economics and finance, risk is defined as a measure of uncertainty about the future behaviour of various markets. Value at Risk, popularly known as VaR is one of the modern risk measuring techniques. VaR, which is generally computed for a portfolio, can be defined as a measure of worst expected loss under normal market conditions for a specific time horizon at a given confidence interval. More specifically, "VaR answers the question: how much can I lose with x % probability over a pre-set horizon" [79].

Based on theories and concepts taken from [75] and [23] we provide a mathematical procedure for the calculation of the VaR. The mathematical estimation of the VaR is based on a time horizon *h*, a confidence level *c*, and an initial portfolio value W_0. Let the value of a portfolio at the end of *n* days be W_n, then the continuously compounded return u_n in the last *n* days for this portfolio can be written as,

$$u_n = \ln\left(\frac{W_n}{W_0}\right) \qquad 2.22$$

As per Eq (2.2), if the initial value of the portfolio was W_0, then the value of the portfolio at the end of *n* days would be given by $W_n = W_0 e^{u_n} \approx W_0 (1+u_n)$, since $e^x \approx (1+x)$ for small values of x. Let the expected value (mean) and volatility of return u_n be μ and σ respectively. If the worst possible return on this portfolio is u^*_n, then the lowest portfolio value can be defined as $W^*_n = W_0 (1 + u^*_n)$. Now the VaR can be defined as the loss, relative to the mean μ_n,

$$VaR(\mu_n) = E(W_n) - W_n^* = W_0(1+\mu_n) - W_0(1+\mu_n^*), \qquad 2.23$$

where, μ_n is the mean of the expected return u_n and μ^*_n is the mean of the worst possible return u^*_n. We can write Eq(2.23) as,

$$VaR(\mu_n) = -W_0(\mu_n^* - \mu_n), \quad 2.24$$

If we substitute μ_n = 0 in Eq (2.24), then we can define VaR as the absolute loss, given by the following equation,

$$VaR(0) = -W_0(\mu_n^*), \quad 2.25$$

In Eqs (2.24 & 2.25), μ^*_n corresponds to the *cut-off return*, which is defined as the lowest acceptable rate of return. Using the notion of confidence level, VaR can be derived from the probability distribution of the future portfolio *f (w)*. At a given confidence level *c*, the lowest possible realization of portfolio value *W* is W^*, such that the probability of exceeding this value is *c*. Hence *c* can be defined as,

$$c = \int_{W^*}^{\infty} f(w)dw, \quad 2.26$$

and the probability of a lower realization than W^* is *1-c*, that is,

$$P(w \le W^*) = \int_{-\infty}^{W^*} f(w)dw = 1 - c. \quad 2.27$$

Figure 2.16 shows a normal distribution with areas *c* and *1-c*. For a given confidence level *c*, the lowest possible realization of portfolio value *W*, lies at the border of the two areas and is marked as W^*.

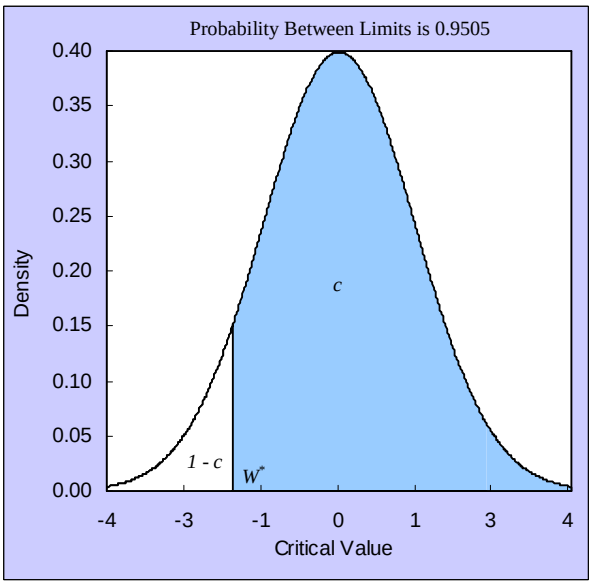


Figure 2.16 The standard normal distribution and the confidence level c.

The above specification is valid for any distribution; however, VaR estimates can be greatly simplified by assuming normal distribution. In the case of a normal distribution the VaR can be estimated directly from the portfolio standard deviation, using a multiplicative factor which

depends on the confidence level and time horizon. The value α, which corresponds to the confidence level *c* in a probability distribution, can be expressed as,

$$-\alpha = \frac{u_n - \mu_n}{\sigma_n}, \qquad 2.28$$

The negative value of α in Eq (2.28) assumes that the critical value *c* is such that we work towards the left of the mean – the interpretation is that over a large number of trading days the value of a portfolio will decline by no more than α%. In Figure 2.16, which demonstrates a 95 % confidence level, we see that α lies at -1.65. We can write u^*_n as,

$$u_n^* = -\alpha\sigma_n + \mu_n . \qquad 2.29$$

Let us assume that σ and μ are expressed on a daily basis. However, the time horizon for VaR is defined as *h* days. To scale the one-day variance into an *h*-day variance, we need to multiply the one-day variance by *h*,

$$\sigma^2_{h-days} = \sigma^2_{1-day} h , \qquad 2.30$$

Taking a square root on both sides of Eq (2.30), we get,

$$\sigma_{h-days} = \sigma_{1-day}\sqrt{h} \qquad 2.31$$

Now we can substitute Eq (2.29) in Eq (2.24), and replace the *1-day* σ with an *h-day* σ to define the general expression for the calculation of VaR, relative to the mean μ,

$$VaR(\mu) = -W_0(u_n^* - \mu_n) = W_0\alpha\sigma\sqrt{h} \qquad 2.32$$

Several strategies have been suggested for lowering risk through methods like *diversification of the portfolio*. A detailed mathematical discussion on portfolio diversification can be found in references [75] and [23].

2.3.2 Capital Asset Pricing Model (CAPM)

The capital-asset pricing model (CAPM), one of the proponents of modern asset pricing theories, introduces the notion of the dependence of expected returns of an asset on the risk involved in holding the asset [126], [131]. For example, if the expected price change of an asset is positive, an investor might want to hold the asset and bear the associated risk in the hope to sell it at a higher price later. On the other hand, an investor not wanting to take risk for an asset that has an "unforecastable" return would willingly make a lesser profit by selling it off earlier.

The CAPM, first proposed by Sharpe and Lintner, predicts that the excess return of a stock (return over the *riskless* rate of return) should be proportional to the market premium (market return over

the *riskless* rate of return) [136], [94]. This implies that the average return of high risk stocks is higher than the average return of low risk stocks and that this relationship is roughly linear.

The CAPM is defined by the following equation,

$$E(r_{it}) = r_{0t} + \left[\frac{Cov(r_{it}, r_{mt})}{\sigma_m^2}\right]\left[E(r_{mt}) - r_{0t}\right], \qquad 2.33$$

where r_{it} is the rate of return for risky assets, r_{0t} is the rate of return for a *riskless* asset, r_{mt} is the market return, and σ^2_m is the variance of the market return. In Eq (2.33), we observe that the market variance (σ^2_m) is not indexed by t since its just one value for the variance of the market return time series (r_{mt}). A complete derivation of Eq (2.33) can be found in [55].

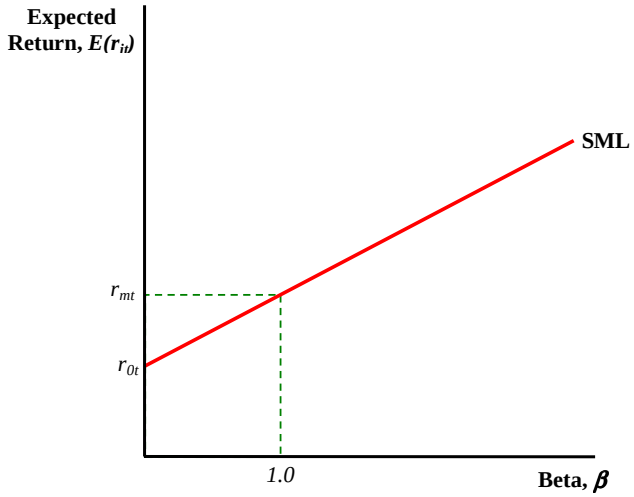


Figure 2.17 The *security market line* (SML) relating the expected returns on assets to their systematic risks.

In the CAPM, Eq (2.33) is also known as the *security market line* (SML). Eq (2.33) implies that the excess return from asset i (in excess of the risk free asset return) is proportional to the market premium (market return in excess of the risk free asset return). The proportionality factor in Eq (2.33) is known as *systematic risk* or the *beta* of an asset. Therefore, mathematically, *beta* of an asset is defined as the ratio of the covariance of the asset's return (r_{it}) with the market return (r_{mt}), to the variance (σ^2_m) of the market return,

$$\beta_i = \frac{Cov(r_{it}, r_{mt})}{\sigma_m^2}. \qquad 2.34$$

If we substitute the value of β from Eq (2.34) in Eq (2.33), then the *security market line* defined by Eq (2.33) can be re-written as,

$$E(r_{it}) = r_{0t} + \beta\left[E(r_{mt}) - r_{0t}\right]. \qquad 2.35$$

Hence the *security market line* relates expected returns on assets to their *systematic risks* – or their *betas*. This is shown in Figure 2.17 where a plot of the expected return versus *beta* produces the *security market line*. As per Eq (2.35) and Figure 2.17 it is clear that the y-intercept is the *riskless* rate of return (r_{ot}) and the slope of the SML is given by [$E(r_{mt}) - r_{0t}$].

The *beta* of an asset can be calculated mathematically using Eq (2.34). However, in empirical finance, the usual estimator for the *beta* of an asset is the ordinary least square (OLS) estimate from the regression defined by Eq (2.36),

$$(r_{it} - r_{0t}) = \alpha + \beta_i (r_{mt} - r_{0t}) + \varepsilon_{it}\,, \qquad 2.36$$

where ε_{it} is a white noise disturbance term known as the non-systematic or idiosyncratic risk, alpha is the intercept of the regression, *beta* is the systematic risk, r_{mt} is the market return, r_{0t} is the *riskless* rate of return, and r_{it} is the rate of return for a risky asset. Figure 2.18 shows how a regression fit is obtained after plotting the excess return (r_{it}-r_{0t}) or *stock premium* versus the *market premium* (r_{mt}-r_{0t}). The slope of the fitted line is equal to the *systematic risk* or *beta* of an asset.

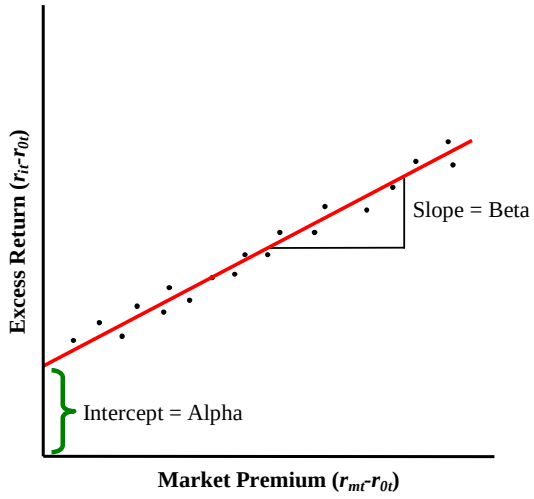


Figure 2.18 Methodology adopted in empirical finance to estimate *beta*.

Early empirical studies on the CAPM were supportive of the fact that average return of high *beta* stocks was higher than the average return of low *beta* stocks [13], [14], [43]. Campbell showed that this relationship is roughly linear but the slope was too flat to strongly support the CAPM [24]. Subsequent studies focused on *beta* estimation issues in more detail. These included testing the stability of *beta* over time [69], effects of borrowing constraints on *beta* [13], estimation of *beta* under structural breaks and regime switches [53], the effect of world markets and volatility on *beta* [10], [9], [68], non-synchronous data issues in the estimation of *beta* [133], impact of investor time horizon on *beta* [93], and the impact of the return interval on *beta* [22], [21], [27], [48], [70], [66], [67].

If we replace r_{it} by r_{mt} in Eq (2.34) we will get a *beta* value of 1, that is, the *beta* of the market is 1. Therefore, for an asset, a value of *beta*, which is greater than one, is considered high and promises higher returns. On the other hand a *beta* value, which is lower than one, is considered low and would potentially indicate lower returns. The accurate estimation of *beta* can help in portfolio management since it can help investors to decide which stocks they want in their portfolio based on how risky the stocks are and what could be the potential returns from these stocks. A detailed procedure for portfolio construction and computation of risk-return tradeoffs at various timescales will be presented in Chapters 3 and 4.

2.3.3 Cointegration

A common property of economic time series is nonstationarity. This means that a variable has no clear tendency to return to a constant value or a linear tend. An important aspect of research in macroeconomics is to estimate relationships amongst variables (i.e. two or more time series). Until the 1980s several large simultaneous equation models were based on the assumption that variables in these models are stationary. However, financial time series are indeed realizations of nonstationary processes. Therefore, statistical inferences associated with stationary simultaneous equation models would not be valid for nonstationary time series. The above difficulty was not particularly well understood by model builders three decades ago until Clive Granger introduced the concept of cointegrated variables in economics.

Until Clive Granger's work on cointegrated economic variables, it was common practice to estimate equations containing nonstationary variables by straightforward linear regression. It was not realized that such standard statistical inferences might lead to totally spurious regressions. In 1974, Clive Granger and his colleague Paul Newbold pointed out that such a regression might falsely suggest a significant relationship amongst variables where no such relationship actually exists [63]. Their method involved generating independent random walks and regressing these series on each other. Their results indicated that though the variables in the regression were independent the null hypothesis of a zero regression coefficient was rejected more frequently than standard theory had proposed until that time. They also observed that the residuals of the estimated equation showed a strong positive autocorrelation. These results proved that existing econometric models that assume significant relationships amongst nonstationary economic variables could be erroneous.

Clive Granger's solution to the above problem can be illustrated by the following regression equation:

$$y_t = \alpha + \beta x_t + \varepsilon_t, \qquad 2.37$$

where y_t is the dependent variable, x_t is an exogenous regressor, and $\{\varepsilon_t\}$ is a white noise zero mean process. With the above equation, Granger proposes, "a simulation of the explanatory right-hand side should produce the major properties of the variable being explained" [60]. This would suggest that if y_t is a seasonal variable then x_t has to be seasonal if ε_t is the white noise. Granger further introduced the concept of *degree of integration* of a variable. That is if a variable z_t can be approximately made stationary by differencing it d times, then the variable is said to be integrated of order d or $I(d)$. This would imply that weakly stationary random variables are $I(0)$. Many macroeconomic variables can be regarded as $I(1)$ variables: if $z_t \sim I(1)$, then $\Delta z_t \sim I(0)$. It is to be noted that $I(1)$ variables dominate $I(0)$ variables – in a linear combination of variables the variation of the former dominates the variation of the latter. For example, if $z_t \sim I(1)$ and $w_t \sim I(0)$, then $z_t + w_t \sim I(1)$.

Let us assume that for Eq (2.37) both $x_t \sim I(1)$ and $y_t \sim I(1)$. This would generally imply that $y_t - \beta x_t \sim I(1)$. There is however one important exception to the above rule which states that if $\varepsilon_t \sim I(0)$, then $y_t - \beta x_t \sim I(0)$, i.e., the linear combination of $y_t - \beta x_t$ has the same statistical properties as an $I(0)$ variable. Therefore we can say that if a linear combination of a set of $I(1)$ variables is $I(0)$, then the variables are *cointegrated*. This is an important concept in the analysis of nonstationary economic time series [60].

The importance of cointegration for modelling nonstationary economic time series is explained by the so-called Granger representation theorem [64]. Let us consider a bivariate autoregressive system of order p described by the following equations:

$$\begin{aligned} x_t &= \sum_{j=1}^{P} \gamma_{1j} x_{t-j} + \sum_{j=1}^{P} \delta_{1j} y_{t-j} + \varepsilon_{1t} \\ y_t &= \sum_{j=1}^{P} \gamma_{2j} x_{t-j} + \sum_{j=1}^{P} \delta_{2j} y_{t-j} + \varepsilon_{2t} \end{aligned} \qquad 2.38$$

where x_t and y_t are $I(1)$ and cointegrated, and ε_{1t} and ε_{2t} are white noise processes. According to the Granger representation theorem, the above system can be written as:

$$\begin{aligned} \Delta x_t &= \alpha_1 (y_{t-1} - \beta x_{t-1}) + \sum_{j=1}^{p-1} \gamma^*_{1j} \Delta x_{t-j} + \sum_{j=1}^{p-1} \delta^*_{ij} \Delta y_{t-j} + \varepsilon_{1t} \\ \Delta y_t &= \alpha_2 (y_{t-1} - \beta x_{t-1}) + \sum_{j=1}^{p-1} \gamma^*_{2j} \Delta x_{t-j} + \sum_{j=1}^{p-1} \delta^*_{2j} y_{t-j} + \varepsilon_{2t} \end{aligned} \qquad 2.39$$

where either $\alpha_1 \neq 0$ or $\alpha_2 \neq 0$. Both equations in Eq (2.39) can be termed as "balanced" that is their right-hand sides and left-hand sides are of the same order of integration: this is because $y_{t-1} - \beta x_{t-1} \sim I(0)$ and both $\Delta x_t \sim I(0)$ and $\Delta y_t \sim I(0)$ since x_t and y_t are $I(1)$.

Let the term $y_t - \beta x_t = 0$ define a dynamic equilibrium relationship between the two economic variables, *y* and *x*. Then $y_t - \beta x_t$ would indicate the degree of disequilibrium. The coefficients α_1 and α_2 represent the strength of the disequilibrium correction, and the system is said to be in *error-correction* form. A system characterized by these two equations is therefore in disequilibrium at any given time, but has a built-in tendency to adjust itself towards the equilibrium. This proves that an econometric model cannot be specified without knowing the order of integration of the variables. Clive Granger shared the 2003 Nobel Prize in economics with Robert Engle (who won it for developing the ARCH model) for his work on cointegration.

Other researchers who proposed the concept of linear combinations of nonstationary variables before Clive Granger include Phillips who coined the term "error correction" [119] and Sargan [132]. The famous consumption equation in Davidson et al. (the so-called DHSY model) [31] also played an important role in the dissemination of the idea of cointegration amongst macroeconomists.

2.4 Discussion

In this chapter, we have introduced the concept of complex systems and explained how patterns of variation over time represent a defining feature of these systems. Various univariate and multivariate time series analysis methods that are used for assessing the state of complex systems have been reviewed and studied. These range from *ad hoc* real world techniques to the more sophisticated, statistically and empirically well-grounded methods for time series analysis.

The take home message though is an important property of complex systems called "universal scaling" or "self organized complexity". This leads us to the concept of *universality* in behaviour of a range of complex systems and hence of time series emanating from them. The power law analysis for complex time series shows that their behaviour is scale invariant – they exhibit universal dynamics at different timescales. From this *universality* stems a need for universal tools that can guide and simplify the study of various complex systems that surround us. Moreover, these universal tools are important for a computer scientist to build robust autonomous systems that can work across many domains.

Multiscale wavelet analysis is a powerful analytical technique originating in signal processing that seems to provide a unity of approach for analyzing "fluctuations" or "rhythms" in complex systems. The wavelet analysis projects a time series onto a collection of orthonormal basis functions (wavelets) to produce a set of wavelet coefficients [118]. These coefficients capture information from the time series at different frequencies or timescales at distinct times. The wavelet analysis can deal well with nonstationary, heterogeneous and transient behaviour, which

makes it quite useful for analyzing a range of complex time series. Moreover, the wavelet analysis is a nonparametric method and hence requires no *a priori* knowledge about the system under study. According to Ramsey, the wavelet analysis has formalized old notions of decomposing a time series into its various components [127]. More detailed and rigorous discussions on wavelets and their usefulness for analyzing complex systems will follow in Chapters 3 and 4.

Motivated by the unity of approach and the systemacity with which wavelets can study complex time serial data, we propose a wavelet framework for analyzing temporal patterns of variation in complex systems. Our wavelet framework has the capability and robustness of analyzing signals across many domains. The core of the framework is the discrete wavelet transform (DWT) and the maximal overlap discrete wavelet transform (MODWT), which is one of the variants of the DWT. The wavelet framework has three main modules:

1. DWT for pattern identification in financial time serial data: This module generates the *Surrey Market Report****** for various financial time series like foreign exchange rates, stock prices, and composite indices. Chief features like cycles, trends, turning points and structural breaks are identified and 'marked up' on the original signal. Moreover, a numerical characterization of these features (like an equation describing the market trend, period of cyclicality etc.) is generated to describe the market dynamics at various timescales.

***A Note on Surrey Market Report**
The Surrey Market Report is an HTML webpage which our wavelet framework generates based on the output from the pattern identification module. The Surrey Market Report displays and reports graphical and numerical summaries of features such as cycles, trends, turning points, and structural changes of the input signal. It also generates a "headline" based on the trend detected in the input signal. The Surrey Marker Report is generated automatically, by filling a pre-specified HTML template with the results for trends, cycles, turning points, and variance changes obtained from the pattern identification module. A sample Surrey Market Report can be viewed at: http://www.saifahmad.com/May6/SurreyMarketReport2.htm We will refer to the output of the pattern identification module under the name of Surrey Market Report throughout the thesis.

2. MODWT and CAPM for financial risk management: Here a multivariate multiscale analysis is performed on financial data based on the MODWT and the CAPM. The module manages portfolios of stocks and generates risk-return tradeoffs at various timescales for these portfolios. Hence this module is useful for risk assessment and risk management in financial markets.

3. MODWT based analysis of variance (ANOVA) for studying power law behaviour: This module analyzes power law behaviour in complex time series based on the computation

of the wavelet variance as a function of timescale. The module assesses the state of a system based on the value of the power law exponent α. Since power laws are a universal feature of complex systems, this module has the capability of analyzing temporal patterns of variation across a variety of domains.

In Figure 2.19 we provide a conceptual outline of our wavelet framework for complex time series analysis. It is interesting to note how the power law analysis breaks the domain barrier to render itself useful across many domains like economics, medicine and physics. A detailed architecture and description of the proposed wavelet framework will be presented in Chapter 4.

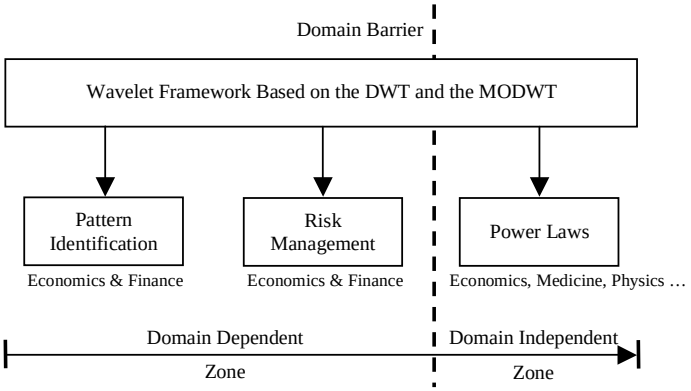


Figure 2.19 Conceptual outline of the proposed wavelet framework for time series analysis in complex systems.

3 Analyzing Complex Time Series

In Chapter 2 we discussed various time series analysis methods that are used to characterize the behaviour of time series. Except a brief description of Fourier transforms and wavelets, all the other methods that were discussed comprised time domain analysis. Time domain analysis usually involves the statistical computation of parameters like the mean and standard deviation to evaluate the behaviour of a time series. Statistical measures of variations or fluctuations in time series are easy to compute and provide valuable information about the series being analyzed. However, time series produced by complex systems are highly nonlinear and exhibit intricate phenomena like time varying volatility, nonstationarity, time changing probability distributions, long memory, and discontinuities [40], [62], [50], [26], [59], [160]. In such a scenario, time domain measures may be susceptible to bias, and they may not be able to reliably distinguish between distinct complex dynamics. There are many potential examples of time series with identical means and standard deviations but with very different underlying fluctuations and dynamics [141]. Therefore, there is a need for more sophisticated time series analysis methods that can cope with complex dynamics of real-world signals.

An alternative to time domain analysis is the frequency domain analysis which is based on the computation of the Fourier transform. According to Fourier theory time series data may be considered a sum of sinusoidal oscillations with distinct frequencies. Based on this assumption, the Fourier transform mathematically converts or transforms a signal from time domain to frequency domain and determines how much of each frequency the signal contains [20]. Such a conversion is termed *spectral analysis* since it provides an evaluation of the power (amplitude) of the contributing frequencies to the underlying signal. However, Fourier analysis alone cannot possibly deal with non-periodic and unobvious temporal patterns that could be contaminated with noise or hidden in the data. The basic premise of Fourier analysis is that the analyzed signal is periodic and stationary. For complex, real-world time series, the periodicity and stationarity conditions are not met.

The wavelet multiscale analysis is a powerful filtering technique originating in signal processing, which seems to provide better insights into the dynamics of complex time series than that provided by current classical statistical and spectral methodology. The main concept that has arisen from wavelet filters is that of timescales and how a signal can be represented by several signals, each characterizing fluctuations of the original signal at a particular timescale. The wavelet transform is also known as a *time-frequency* analysis method since it captures information from the original signal at various frequencies or timescales at distinct times. Since the wavelet

transform decomposes a time series into several sub-series on a timescale basis, it is very close to the philosophy of fractal or scaling behaviour in complex systems (Figure 2.1) and in time series produced by complex systems. Unlike the Fourier transform described above, the wavelet analysis is not restrained by the assumption of stationarity which makes it quite useful for analyzing a range of complex time series. In addition, wavelets are nonparametric tools for decomposing a time series and hence require no *a priori* knowledge about the dynamics of the signal being analyzed. According to Ramsey, wavelets have formalized old notions of decomposing a time series into its various components [127]. We thus conclude that perhaps wavelets are universal tools that can guide and simplify the study of the universal phenomenon of scaling in complex systems.

We begin this chapter with a description of time series and their various components. This is followed by a detailed description of wavelets and the concepts arising from wavelet filtering vis-à-vis decomposition of time series. We also study the computation of statistics like the wavelet variance and wavelet covariance and their significance in assessing complex time series. We go on to study the so called long memory processes in time series and the usefulness of wavelet analysis in dealing with such time series. We conclude this chapter with the problem formulation for analyzing complex time series.

3.1 Components of a Time Series

We begin with a very simple definition of a time series. Let us consider the measurement of unemployment rates, which is a very important measure of the health of an economy. Some figures are gathered by a government agency and each month a new number is announced. Next month there will be another number of the unemployment rate, and so forth. If we string these numbers together on a graph, it will produce a time series. A time series is a sequence of measurements through time.

Mathematically, a time series may be represented as,

$$X = \{x_t, t = 1, \ldots, N\}, \tag{3.1}$$

where t is the time index and N is the total number of observations.

Scientists often speak of a time series X as consisting of different components, namely, *trends*, *seasonal* and *cyclical* fluctuations and the irregular variation or *noise*. Eq (3.1) can thus be re-written as,

$$X = \{(t_t + s_t + \varepsilon_t), t = 1, \ldots, N\}, \tag{3.2}$$

where t_t, s_t, and ε_t are the trend, seasonal and noise components respectively. A synthetic time series with 200 samples ($N = 200$) and its various additive components is shown in Figure 3.1.

This synthetic series has been generated by adding together at each time index *t*, the values of a straight line, a rectified sine wave and a series of normally distributed random numbers with zero mean and unit variance.

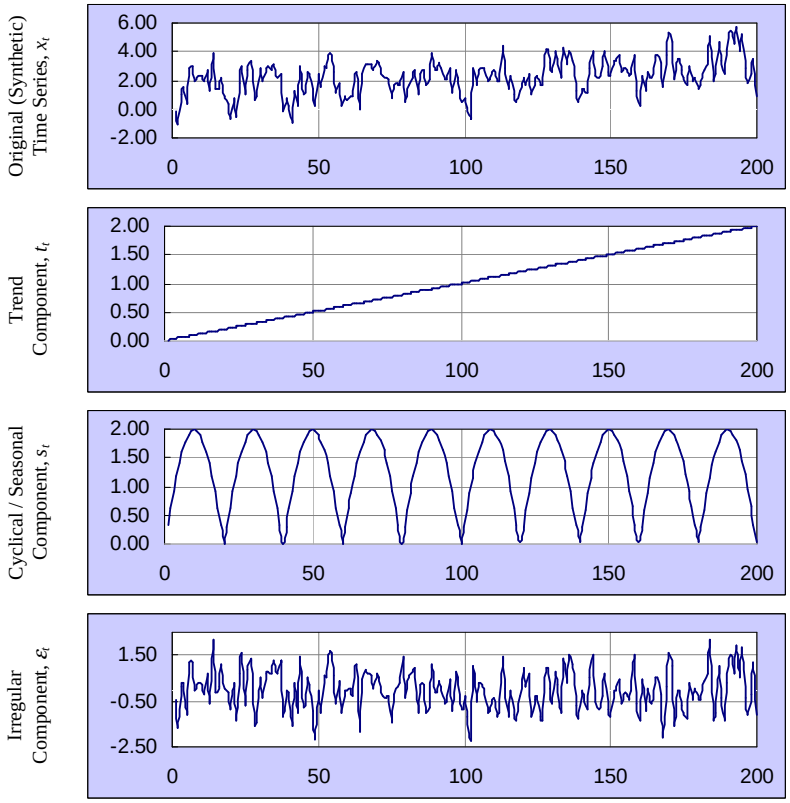


Figure 3.1 A synthetic time series and its additive components.

Mathematically, the synthetic time series of Figure 3.1 can be expressed as,

$$x_t = \left| 2\sin\left(\frac{2\pi t}{40}\right) \right| + 0.01t + \varepsilon_t . \qquad 3.3$$

As per Figure 3.1, we describe below, in more detail, the different sources of variation in a time series:

(1) Seasonal Effect

Many day-to-day time series like temperature readings or sales figures exhibit seasonal variation, which could be annual, biannual and so on. An example could be the systematic rise (in summer) and fall (in winter) in temperature over several years. Once seasonality is removed from a time series we are left with what we call the "deseasonalized" data.

(2) Other Cyclical Changes

Apart from the seasonal effects, a time series may exhibit cyclical changes, which represent a variation at a fixed period. A typical example here would be the daily variation in temperature. In the financial domain, a time series may exhibit oscillations, which do not have a fixed period. For example, economic data are sometimes thought to be affected by business cycles with a period varying between 5 and 7 years.

(3) Trend

Trend is generally defined as "the long-term change in the mean level". However, this definition seems to be constrained by what is meant by "long-term". There could be situations where a cyclical behaviour is observed only over a very long period of time, such as 50 years. Therefore, in such situations if we consider only 20 years' data, then this long-term oscillation would appear to be the trend whereas if several hundred years' data were considered, then the long-term oscillation would be very visible. Granger defines "trend in mean" as comprising all cyclical components whose wavelength exceeds the length of the observed time series [61].

(4) Other Irregular Fluctuations

After trend and cyclical variations have been removed from a time series, we are left with a series of residuals, which may or may not be "random". There may still be some cyclical components in these residuals and a proper spectral analysis might be useful for checking the cyclicality of the residual series.

Just as in complex systems it is important to study the parts and the relationship between the parts to understand the behaviour of the system, in complex time series it is important to study the various parts (components) of a series and the relationship between them to better understand the behaviour of the time series. To be able to do this effectively, the first step is the decomposition of complex time series into simpler units or components. Traditional time series analysis methods involve decomposing a series into trend, seasonal variation, other cyclical changes and the remaining "irregular" fluctuations. However, this decomposition can be unique and effective only if certain assumptions about the underlying phenomena of the process are made, for example nonstationarity.

Several *ad hoc* techniques exist for decomposing time series. For example the trend component can be removed by computing a first difference of the given series ($x_t - x_{t-1}$). Similarly, the annual (12-monthly) seasonal component in a monthly series can be removed by computing a twelfth difference ($x_t - x_{t-12}$). However, these techniques may not give an accurate representation of the actual components of a time series, especially in nonstationary and noisy data. Moreover, such techniques merely offer a transformation rather than a decomposition where various components

(decompositions) can be linearly added to get back the original signal. Wavelet multiresolution analysis (MRA) offers a systematic, linearly additive decomposition of a time series based on timescales. We will show in the following sections how wavelets can uniquely deconstruct a given time series into its various components and how we can study these components separately and the relationships amongst them to better understand the dynamics of the original series and system under study.

3.2 Wavelet Filtering

Wavelet filtering provides a natural platform to deal with the time-varying characteristics of real-world time series and is not restrained by the assumption of stationarity. A number of concepts, for example, nonstationarity, multiresolution, and approximate decorrelation have emerged from wavelet filters [54], [118]. The multiresolution properties of economic, biological, and physical processes can be easily studied using wavelet filters. The wavelet transform decomposes a process into different time horizons (timescales). This ability of the wavelet transform makes it useful in differentiating seasonalities, revealing structural breaks and volatility clusters, and identifying local and global dynamic properties of a process at these timescales. The wavelet transform is closely related to band pass filters with properties similar to those used in business-cycle literature. Moreover, wavelet filters provide a convenient way of dissolving the correlation structure of a process across different timescales. This means that the wavelet coefficients at one level (scale) are not much associated with coefficients at other scales. Equally, they are also not much associated within each scale. This is convenient when performing tasks such as modelling a process, since it is easier to deal with an uncorrelated process as compared to an unknown correlation structure. In addition to the above, wavelets are nonparametric tools for signal analysis and hence no parameters need to be set beforehand for carrying out an analysis using wavelets. The wavelet transform is readily implemented using Mallat's pyramidal algorithm [97] and hence has been termed as a “formal method” for decomposing a time series into its various components by Ramsey [127].

3.2.1 Discrete Fourier Transform

Since wavelet analysis is a *time-frequency* analysis technique, a natural starting point for its study would be the so called *frequency* domain analysis. The Fourier transform is the oldest *frequency* domain analysis available for analyzing time series data and the wavelet analysis builds and improves upon concepts derived from the Fourier transform. The discrete Fourier transform (DFT) or the fast Fourier transform (FFT) can approximate a discretely sampled process (time series) x_t via a linear combination of sines and cosines. Each of these sines and cosines is itself a

function of frequency and hence the DFT may be seen as a decomposition on a frequency-by-frequency basis [54], [20].

According to the Fourier theory, a signal may have several sinusoidal components, possibly with different amplitudes (sizes), different phases, and different frequencies. Hence any infinite sequence or time series x_t, may also be viewed as a combination of an infinite number of sinusoids with different amplitudes and phases,

$$x_t = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(f)e^{i2\pi ft}df\ , \qquad 3.4$$

where $i = \sqrt{-1}$ is an imaginary number and f is the frequency. In Eq (3.4), $X(f)$ is given by,

$$X(f) = \sum_{t=-\infty}^{\infty} x_t e^{-i2\pi ft}\ . \qquad 3.5$$

Eq (3.4) is referred to as the *inverse Fourier transform*, and Eq (3.5) is the *Fourier transform* of x_t. Eqs (3.4 & 3.5) constitute a *Fourier representation* of the sequence x_t, and hence they are called a *Fourier transform pair*. Eq (3.4) is also known as the *synthesis equation* since it represents the original sequence x_t as a linear combination of complex sinusoids infinitesimally close in frequency with $X(f)$ determining the relative weight of each complex sinusoid. Similarly, Eq (3.5) is known as the *analysis equation* which analyzes the original sequence x_t, to determine how much of each frequency component is required to synthesize it [15]. We can write x_t and $X(f)$ as a Fourier transform pair,

$$\sum_{t=-\infty}^{\infty}|x_t|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|X(f)|^2 df\ , \qquad 3.6$$

which is known as *Parseval's theorem*. In Eq (3.6), the left hand side is the total energy in the signal, which may be obtained by integrating the energy per unit frequency $X|(f)|^2/2\pi$ over 2π interval of discrete-time frequencies. The squared magnitude of the Fourier transform, $X|(f)|^2$, is known as the *energy density spectrum* or the *power spectrum* of the signal x_t [114].

If x_t were a finite sequence instead of an infinite one, then its Fourier representation would be given as,

$$x_t = \frac{1}{N}\sum_{k=0}^{N-1} X_k e^{i2\pi f_k t}\ ,\ t = 0, 1, \ldots, N-1, \qquad 3.7$$

and,

$$X_k = \sum_{t=0}^{N-1} x_t e^{-i2\pi f_k t}\ ,\ k = 0, 1, \ldots, N-1, \qquad 3.8$$

where $f_k = k/N$. In this case, the Parseval's relation of Eq (3.6) can be re-written as,

$$\sum_{t=0}^{N-1}|x_t|^2 = \frac{1}{N}\sum_{k=0}^{N-1}|X_k|^2 \,. \qquad 3.9$$

We will now consider an example to demonstrate the computation of the Fourier transform in a discrete time setting. Let us consider a time series generated by Eq (3.10), where ε_t is a normally distributed random variable with zero mean and unit variance.

$$\begin{aligned} x_t &= \cos\left(\frac{2\pi t}{12}\right) + \cos\left(\frac{2\pi t}{20}\right) + \varepsilon_t \\ &= \cos(2\pi f_1 t) + \cos(2\pi f_2 t) + \varepsilon_t \,. \end{aligned} \qquad 3.10$$

This time series also has two cyclical components with time periods of 12 and 20 since $f_1 = 1/12$ and $f_2 = 1/20$.

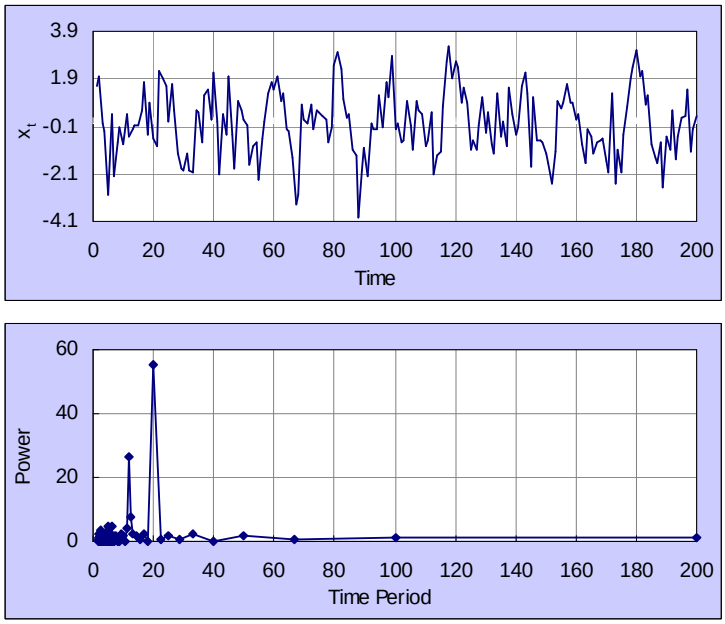


Figure 3.2 Time series described by Eq (3.10) and its Fourier power spectrum.

The top plot of Figure 3.2 shows a sample of the time series described by Eq (3.10) for $N = 200$. Although there are two distinct periodic components in the signal, the random component ε_t makes it difficult to identify these components in the time domain representation of the signal. The bottom plot of Figure 3.2 shows the Fourier power spectrum analysis of the original signal: we can clearly observe two distinct peaks at a time period of 12 and 20 respectively. This indicates that the Fourier analysis has been able to pick up the two cyclical components of the original signal.

From our discussion so far, we can conclude that the Fourier transform is a good alternative to time domain analysis in that it can give us the power (amplitude) of the contributing frequencies

to the underlying signal. However, real-world time series are intrinsically nonstationary and exhibit quite complicated patterns over time (e.g., trends, abrupt changes and volatility clustering). The Fourier transform cannot accurately capture these events since it just transforms the signal from time domain to the frequency domain based on the assumption that the signal is composed of several sinusoids of varying frequencies. Moreover, if the frequency components of a signal are not stationary (that is they appear, disappear and reappear over time), the Fourier analysis may miss them out. The Fourier transform can be seen as summarizing information in the data as a function of frequency and hence it does not preserve information in time. This is quite contrary to how we perceive and observe real-world time series where no frequency resolution exists.

Dennis Gabor realized the shortcoming of the Fourier transform in that it eliminates all time resolution in lieu of frequency resolution. He attempted to achieve a balance between time and frequency by sliding a window across the time series and computing a Fourier transform in each window [51]. This is known as the Gabor transform or Short-time Fourier Transform (STFT). The STFT is simply the same as applying the Fourier transform to several pieces of a time series. The resulting expansion is a function of two parameters, frequency and time shift. The key drawback of the STFT is its fixed window width: this means that it will be unable to resolve nonstationary events if they fall within the width of the window. In such a case the lack of time resolution of the Fourier transform will again come into play. In general, simultaneous time-frequency resolution cannot be achieved because of the Heisenberg's uncertainty principle. Heisenberg, in his uncertainty paper of 1927 said, "The more precisely the position is determined, the less precisely the momentum is known in this instant, and vice versa" [71]. This statement expresses the "uncertainty relation" between the position and the momentum (mass times velocity) of a subatomic particle, such as an electron, in the field of Quantum Mechanics. Philosophically, scientists often refer to the Heisenberg uncertainly principle as the "principle of indeterminacy." For signal processing, this rule translates into the fact that a signal does not simultaneously have a precise location in time and precise frequency. Please see Section 3.2.2 for more details.

3.2.2 Discrete Wavelet Transform

To overcome the fixed time-frequency partitioning of the STFT, a new set of basis functions is needed. The wavelet transform utilizes a basis function called the mother wavelet to capture features that are local in time and local in frequency.

Figure 3.3 introduces this concept by showing the translations and dilations of a square wave function, which is based on the Haar wavelet filter. Figure 3.3a shows the function in its original form while Figure 3.3b shows the same function shifted backward in time. The wavelet filter is

long in time when capturing low-frequency events (Figure 3.3c), and hence a good frequency resolution is achieved. Conversely, the wavelet is short in time when capturing high frequency events (Figure 3.3d) and hence a good time resolution is achieved. By utilizing several combinations of the shifted and stretched mother wavelet, the wavelet transform captures all the pertinent information in a time series and associates it with specific time horizons and locations. This makes the wavelet transform an ideal tool for studying nonstationary and transient time series.

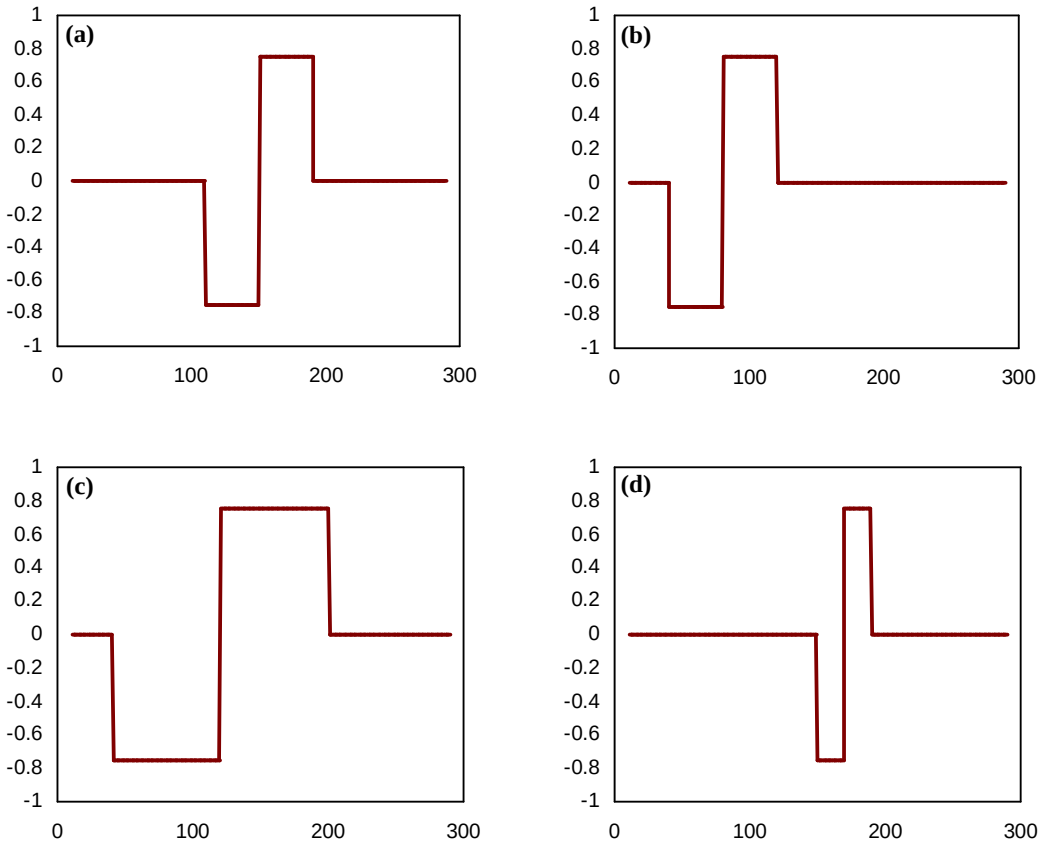


Figure 3.3 Application of translation and dilation to square-wave function.

The Heisenberg uncertainty principle has not been violated, it has just been played with to our advantage. The Fourier transform and the STFT is a function of frequency whereas the wavelet transform is a function of timescale. If the timescale increases, then the wavelet basis is

1 Stretched in time domain,

2 Shrunk in the frequency domain, and

3 Shifted toward lower frequency.

Conversely, if the timescale decreases

4 Time support reduces,

5 Number of frequencies captured increases, and

6 The output shifts towards higher frequencies.

Figure 3.4 shows the partitioning of the time-frequency plane by the Fourier transform, the STFT and the wavelet transform.

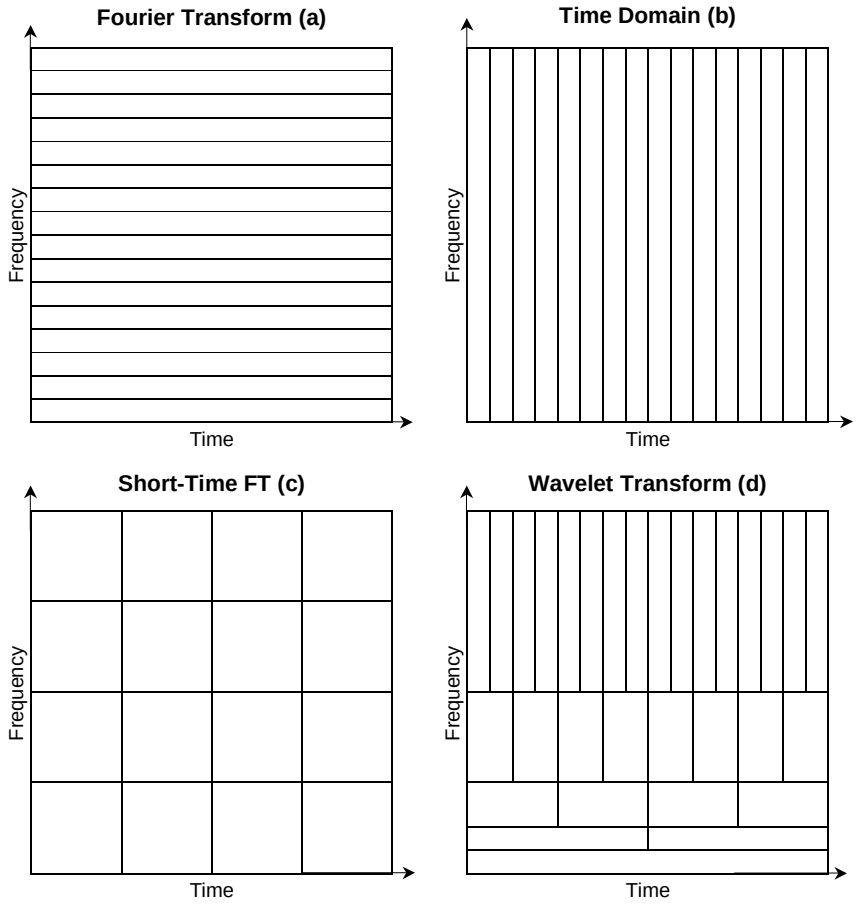


Figure 3.4 Partitioning of the time-frequency plane by different techniques.

In Figure 3.4:

(a) Shows the frequency domain after computing the Fourier transform, representing a perfect frequency resolution and no time resolution.

(b) Shows the time domain representation of the observed time series with perfect time resolution and no frequency resolution.

(c) Shows a balanced resolution between time and frequency by using short-time Fourier transform (STFT).

(d) Shows the wavelet transform, which adaptively partitions the time-frequency plane.

From Figure 3.4d we can see that as the frequency increases, time is more heavily partitioned across longer ranges of frequencies. This is the sort of balance the wavelet transform is able to accomplish to analyze signals. By giving up some frequency resolution, the wavelet transform captures events that are local in time.

The Haar Wavelet

We begin our mathematical discussion on wavelets with the simplest wavelet filter, the Haar wavelet which was introduced in Figure 3.3. The Haar wavelet is a filter of length $L = 2$ which can be precisely described by its *scaling coefficients*,

$$g_0 = g_1 = \frac{1}{\sqrt{2}}, \tag{3.11}$$

or equivalently by its *wavelet coefficients*, $h_0 = 1/\sqrt{2}$ and $h_1 = -1/\sqrt{2}$ through the *quadrature mirror relationship*,

$$h_l = (-1)^l g_{L-1-l}, \text{ for } l = 0,\ldots, L-1. \tag{3.12}$$

In Eqs (3.11 & 3.12), g_l corresponds to the low-pass filter or scaling coefficients and h_l corresponds to the high-pass filter or wavelet coefficients. We will follow this convention throughout this dissertation. The Haar wavelet is special since it is the only compactly supported orthonormal wavelet that is symmetric [30]. The Haar filter also presents the basic properties shared by all Daubechies wavelet filters, for example, orthonormality, and orthogonality to even shifts. Orthonormality can be mathematically expressed as,

$$\sum_{l=0}^{L-1} h_l^2 = 1, \tag{3.13}$$

whereas, orthogonality to even shifts can be expressed as,

$$\sum_{l=0}^{L-1} h_l h_{l+2k} = \sum_{l=-\infty}^{\infty} h_l h_{l+2k} = 0, \tag{3.14}$$

for all non-zero integers k, where by definition $h_l = 0$ for $l < 0$ and $l \geq L$.

Daubechies Families of Wavelet Filters

A wavelet family consists of all wavelet basis vectors, over all scales and translations, derived from a single wavelet filter (or *mother wavelet*). Two wavelet families that we have used extensively in our analyses were developed by Ingrid Daubechies. These are called *extremal phase* and *least asymmetric* wavelets. The Daubechies extremal phase filters are referred to as '*D(L)*' whereas Daubechies least asymmetric filters are referred to as '*LA(L)*'.

Wavelet filter coefficients for the D(4) wavelet, at unit scale, are defined to be,

$$h_0 = \frac{1-\sqrt{3}}{4\sqrt{2}},\ h_1 = \frac{-3+\sqrt{3}}{4\sqrt{2}},\ h_2 = \frac{-3+\sqrt{3}}{4\sqrt{2}},\ \text{and}\ h_3 = \frac{-1-\sqrt{3}}{4\sqrt{2}}. \qquad 3.15$$

Similarly, the scaling coefficients for the LA(8) filter, which have been taken from [118], are given in Table 3.1 below.

Table 3.1 Scaling coefficients for the Daubechies least asymmetric wavelet filter of length $L = 8$.

g_0	=	-0.0757657147893407
g_1	=	-0.0296355276459541
g_2	=	0.4976186676324570
g_3	=	0.8037387518052160
g_4	=	0.2975779560554220
g_5	=	-0.0992195435769354
g_6	=	-0.0126039672622612
g_7	=	0.0322231006040713

We recall that the scaling filter is related to the wavelet filter via the quadrature mirror filter relationship given by Eq (3.12). More details about the scaling coefficients defining Daubechies families of wavelet filters of varying lengths can be found in [30]. Also, the orthonormality (Eq (3.13)) and orthogonality to its even shifts (Eq (3.14)) seen for the Haar wavelet filter is shared by both Daubechies families of wavelet filters discussed here. These can be succinctly expressed using the *squared gain function* of the wavelet filter [54].

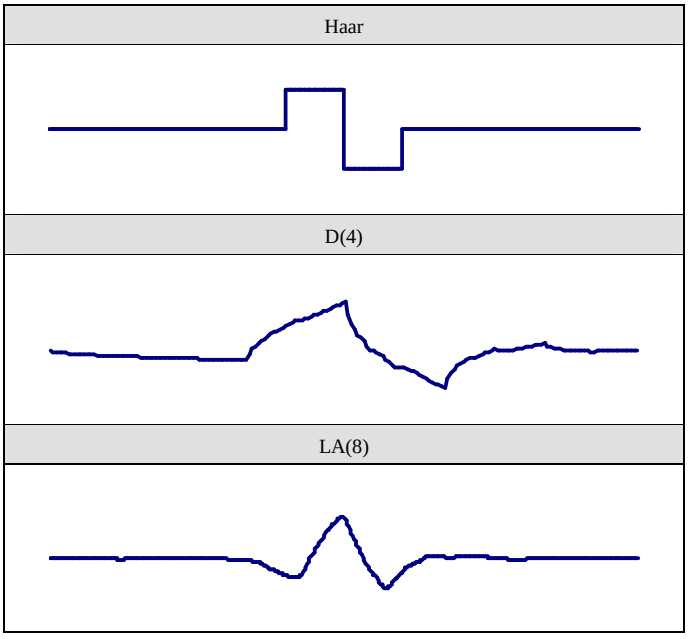


Figure 3.5 Haar, D(4) and LA(8) wavelet filters for scale 6.

Some common wavelet filters or wavelet basis vectors taken from the sixth level of transformation are shown in Figure 3.5. As the length of the wavelet filter increases the smoothness of the basis function increases. However, the increased length, while improving the filters' approximation to an ideal band-pass filter, amplifies boundary effects encountered whenever finite time series are analyzed. From Figure 3.5 we can see that the Haar wavelet filter is a simple square-wave function, the D(4) is quite jagged with a self-similar or fractal-like appearance to it and the LA(8) is quite smooth and very close to being symmetric. When choosing a wavelet filter, we must bear in mind that the filter should agree with the underlying structure of the physical process it is analyzing. For example, if we are analyzing a self-similar fractal time series then we might find the D(4) filter quite suitable for this task.

Implementation of the Discrete Wavelet Transform

In this section we will discuss the implementation of the discrete wavelet transform (abbreviated as DWT) in greater detail. Let $\boldsymbol{x}$ be a dyadic length ($N = 2^J$) vector or discrete time series of observations. The length N vector of discrete wavelet coefficients $\boldsymbol{w}$ is obtained via,

$$\boldsymbol{w} = W\boldsymbol{x}, \tag{3.16}$$

where W is an N × N orthonormal matrix defining the DWT. The vector of wavelet coefficients may be organized into $J + 1$ vectors,

$$\boldsymbol{w} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_J, \boldsymbol{v}_J]^T, \tag{3.17}$$

where $\boldsymbol{w}_j$ is a length $N/2^j$ vector of wavelet coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$ and $\boldsymbol{v}_J$ is a length $N/2^J$ vector of scaling coefficients associated with averages on a scale of length $2\lambda_J = 2^J$.

The matrix W is composed of the wavelet and scaling filter coefficients arranged on a row-by-row basis. Let,

$$\boldsymbol{h_1} = [h_{1,N-1}, h_{1,N-2}, \ldots, h_{1,1}, h_{1,0}]^T, \tag{3.18}$$

be the vector of zero-padded unit scale wavelet filter coefficients in reverse order. That is the coefficients $h_{1,0}, \ldots, h_{1,L-1}$ are taken from an appropriate orthonormal wavelet family of length L, and all values such that $L < t < N$ are defined to be zero. Now h_1 is circularly shifted by factors of two such that,

$$\boldsymbol{h}_1^{(2)} = [h_{1,1}, h_{1,0}, h_{1,N-1}, h_{1,N-2}, \ldots, h_{1,3}, h_{1,2}]^T, \tag{3.19}$$

$$\boldsymbol{h}_1^{(4)} = [h_{1,3}, \ldots, h_{1,0}, h_{1,N-1}, h_{1,N-2}, \ldots, h_{1,5}, h_{1,4}]^T, \tag{3.20}$$

and so on. The $N/2 \times N$ dimensional matrix W_1 is defined to be the collection of $N/2$ circularly shifted versions of $\boldsymbol{h}_1$; that is,

$$W_1 = [\boldsymbol{h}_1^{(2)}, \boldsymbol{h}_1^{(4)}, \boldsymbol{h}_1^{(N/2-1)}, \boldsymbol{h}_1]^T \quad 3.21$$

Let $\boldsymbol{h}_2$ be the vector of zero-padded scale 2 wavelet filter coefficients defined in the same way as Eq (3.18). The matrix W_2 is constructed by circularly shifting the vector $\boldsymbol{h}_2$ by factors of four. This procedure is repeated to construct matrices W_j by circularly shifting the vector h_j (vector of zero padded scale j wavelet filter coefficients) by factors of 2^j. The matrix V_J is simply a column vector whose elements are all equal to $1/\sqrt{N}$ [104]. The structure of the $N \times N$ dimensional matrix W is seen through the sub-matrices W_1, W_2, ..., W_J and V_J via,

$$W = \begin{bmatrix} W_1 \\ W_2 \\ : \\ W_j \\ V_j \end{bmatrix}. \quad 3.22$$

To complete our construction of the orthonormal matrix W, we must be able to explicitly compute the wavelet filter coefficients for scales $1, \ldots, J$. The wavelet filter h_l is associated with unit scale and it satisfies Eqs (3.13 & 3.14). Given the transfer functions of unit scale wavelet and scaling filters, we can define the wavelet filter $h_{j,l}$ for scale $\lambda_j = 2^{j-1}$ as the inverse discrete Fourier transform (DFT) of,

$$H_{j,k} = H_{1,2^{j-1} \bmod N} \prod_{l=0}^{j-2} G_{1,2^l k \bmod N}, \; k = 0, \ldots, N-1. \quad 3.23$$

The modulus operator is required to deal with the boundary of a finite length vector of observations. The resulting wavelet filter associated with scale λ_j has length,

$$L_j = (2^j - 1)(L - 1) + 1. \quad 3.24$$

Also, the scaling filter g_J for scale λ_J is defined as the inverse DFT of,

$$G_{j,k} = \prod_{l=0}^{J-1} G_{1,2^l k \bmod N}, \; k = 0, \ldots, N-1. \quad 3.25$$

In practice, the DWT is implemented via Mallat's pyramidal algorithm [97] that starting with the data x_t, filters a series using h_1 and g_1, subsamples both filter outputs to half their original lengths, keeps the subsampled output from the h_1 filter as wavelet coefficients, and then repeats the above filtering operations on the subsampled output from the g_1 filter. A flow diagram for the first stage of the pyramidal algorithm is shown in Figure 3.6a, where the symbol $\downarrow 2$ means that every other value of the input vector is removed (downsampled by 2).

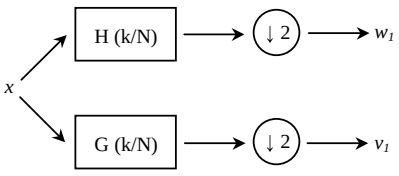


Figure 3.6a Flow diagram illustrating the decomposition of x into the unit scale wavelet coefficients w_1 and the unit scale scaling coefficients v_1.

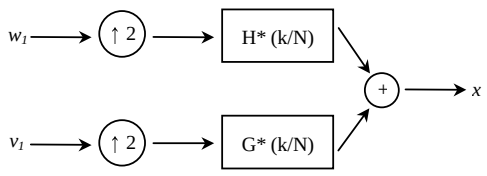


Figure 3.6b Flow diagram illustrating the reconstruction of x from the unit scale wavelet coefficients w_1 and the unit scale scaling coefficients v_1.

For each iteration of the pyramidal algorithm, we require three objects: the data vector $\boldsymbol{x}$, the wavelet filter h_1, and the scaling filter g_1. First iteration of the pyramidal algorithm begins by filtering (convolving) the data with each filter to obtain the following wavelet and scaling coefficients,

$$w_{1,t} = \sum_{l=o}^{L-1} h_l x_{2t+1-l \bmod N} \quad \text{and} \quad v_{1,t} = \sum_{l=0}^{L-1} g_l x_{2t+1-l \bmod N}\,, \tag{3.26}$$

where $t = 0, 1, \ldots, N/2-1$. From Eq (3.26) we note that the downsampling operation has been included in the filtering step through the subscript of x_t. The N length vector of observations has been high- and low-pass filtered to obtain $N/2$ coefficients associated with this information. The second step of the pyramidal algorithm starts by defining the "data" to be the scaling coefficients $\boldsymbol{v}_1$ from the first iteration and applying filtering operations as above to obtain the second level of wavelet and scaling coefficients,

$$w_{2,t} = \sum_{l=0}^{L-1} h_l v_{1,2t+1-l \bmod N} \text{ and, } v_{2,t} = \sum_{l=0}^{L-1} g_l v_{1,2t+1-l \bmod N}\,, \tag{3.27}$$

$t = 0, 1, \ldots, N/4-1$. Keeping all vectors of wavelet coefficients, and the final level of scaling coefficients, we have the following length N decomposition $\boldsymbol{w} = [\boldsymbol{w}_1\ \boldsymbol{w}_2\ \boldsymbol{v}_2]^T$. In the same manner, after the third iteration of the pyramidal algorithm where we apply filtering operations to $\boldsymbol{v}_2$, the decomposition now looks like $\boldsymbol{w} = [\boldsymbol{w}_1\ \boldsymbol{w}_2\ \boldsymbol{w}_3\ \boldsymbol{v}_3]^T$, and so on. The above procedure may be repeated J times where $J = \log_2(N)$, and gives the vector of wavelet coefficients in Eq (3.17).

Inverting the DWT (that is getting back the original signal) is achieved through upsampling the final level of wavelet and scaling coefficients, convolving them with their respective filters (wavelet for wavelet and scaling for scaling) and adding up the two filtered vectors. A flow diagram for reconstructing x from the first level wavelet and scaling coefficient vectors is given in Figure 3.6b. The symbol ↑ 2 means that a zero is inserted before each observation in $\boldsymbol{w}_1$ and $\boldsymbol{v}_1$ (upsampled by 2). Starting with the final level of the DWT, upsampling the vectors $\boldsymbol{w}_J$ and $\boldsymbol{v}_J$ will result in two new vectors,

$$\boldsymbol{w}_J^0 = [0\ w_{j,0}]^T \text{ and } \boldsymbol{v}_J^0 = [0\ v_{j,0}]^T . \tag{3.28}$$

The level *J -1* vector of scaling coefficients $\boldsymbol{v}_{J-1}$ is given by,

$$v_{J-1,t} = \sum_{l=0}^{L-1} h_l w^0_{J,t+l \bmod 2} + \sum_{l=0}^{L-1} g_l v^0_{J,t+l \bmod 2} , \tag{3.29}$$

t = 0, 1. We notice that the length of $\boldsymbol{v}_{J-1}$ is twice that of $\boldsymbol{v}_J$. The next step of reconstruction involves upsampling to produce,

$$\boldsymbol{w}_{J-1}{}^0 = [0\ w_{j-1,0}\ 0\ w_{j-1,1}]^T \text{ and } \boldsymbol{v}_{J-1}{}^0 = [0\ v_{j-1,0}\ 0\ v_{j-1,1}]^T , \tag{3.30}$$

and the level *J -2* vector of scaling coefficients $\boldsymbol{v}_{J-2}$ is given by,

$$v_{J-2,t} = \sum_{l=0}^{L-1} h_l w^0_{J-1,t+l \bmod 4} + \sum_{l=0}^{L-1} g_l v^0_{J-1,t+l \bmod 4} , \tag{3.31}$$

t = 0, 1, 2, 3. The above procedure may be repeated until the first level of wavelet and scaling coefficients have been upsampled and combined to produce the original vector of observations, that is,

$$x_t = \sum_{l=0}^{L-1} h_l w^0_{1,t+l \bmod N} + \sum_{l=0}^{L-1} g_l v^0_{1,t+l \bmod N} , \tag{3.32}$$

t = 0, 1, …, N-1. This is displayed in Figure 3.6b.

We now apply the DWT to the daily IBM stock prices from May 17, 1961 to November 2, 1962. This is a classic dataset exhibiting pronounced changes in its variance structure and has been studied in Box and Jenkins [18]. This series has also been analyzed for changes in variance in [78] and [32]. Before applying the DWT, we compute a return series for the given time series via the first difference of the log-transformed prices – that is, $r_t = ln\ (p_t) - ln\ (p_{t-1})$. Figure 3.7 shows the results of the analysis: the return series is plotted in the upper row while the wavelet coefficient vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_4$ and scaling coefficient vector $\boldsymbol{v}_4$ using the Haar wavelet are shown in the lower rows of the figure. The first scale of the wavelet coefficients $\boldsymbol{w}_1$ are filtering out the high-frequency fluctuations by essentially looking at adjacent differences in the data. The returns are rapidly fluctuating in the interval of 230 to 322. Apart from $\boldsymbol{w}_1$, this interval of rapid fluctuations is also evident in $\boldsymbol{w}_2$ where a change in magnitude of the coefficients is observed between 230 to 322. However, the change in magnitude (between 230 to 322) for $\boldsymbol{w}_2$ is smaller than that observed for the unit scale coefficients, $\boldsymbol{w}_1$. The vector of observations $\boldsymbol{w}_2$ is associated with changes of scale λ_2. Since the IBM series does not exhibit low-frequency oscillations, the higher scale (low-frequency) vectors namely $\boldsymbol{w}_3$ and $\boldsymbol{w}_4$ do not show large variations from zero. The same is true for the scaling coefficients, which are associated with averages of scale $2\lambda_4$ or greater.

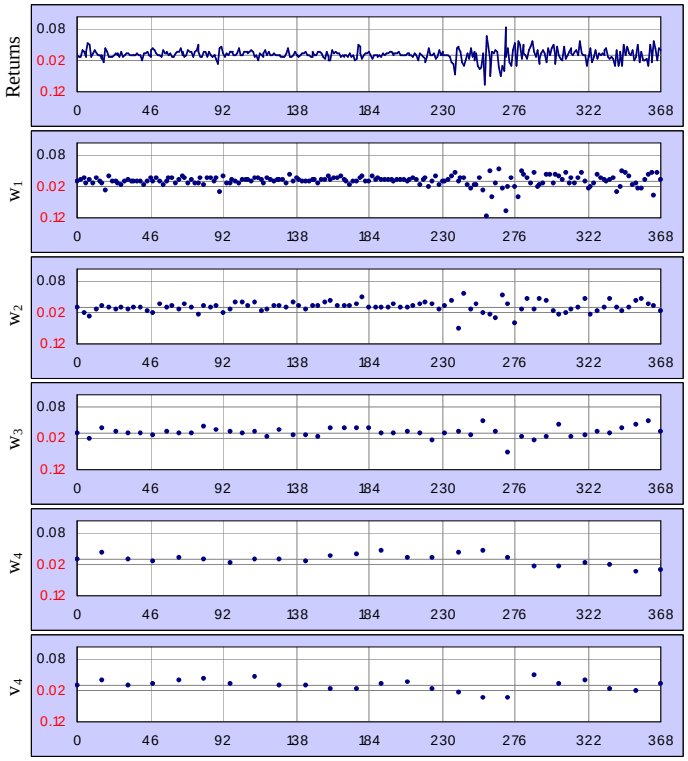


Figure 3.7 Wavelet decompositions of IBM return series ($N = 368$) using Haar wavelet filters. *Negative values are shown in red.*

Multiresolution Analysis

We have seen in the previous section how we can use the DWT to study variations and fluctuations in a given signal at different timescales. However, as we keep on decomposing the subsampled output from the g_1 filter according to Mallat's pyramidal algorithm (Figure 3.6a), we keep losing half the data points at each level of the decomposition (Eq (3.26) and Eq (3.27)) and as such the wavelet coefficient vectors become sparse at higher levels (Figure 3.7). As discussed in Section 3.1, for efficient time series analysis, we would prefer a decomposition where the sub-series have the same number of samples as the original signal and they are linearly additive at each time index t, that is we can add all the sub-series to get back the original signal. The sub-series for an additive decomposition will give us an idea of the exact location of temporal events and fluctuations in a given time series. Moreover, with the above methodology, complex time serial data can be broken down into simpler additive sub-series (with the same number of samples as in the original) which can be studied and interpreted separately and the results combined (aggregated) to throw light on the dynamics of the original signal.

Using the DWT, we may formulate an additive decomposition of a series of observations. Let $\boldsymbol{d}_j = W_j^T \boldsymbol{w}_j$ for $j = 1, \ldots, J$, define the *jth* level *wavelet detail* associated with changes in $\boldsymbol{x}$ at scale λ_j.

The wavelet coefficients $\boldsymbol{w}_j = W_j \boldsymbol{x}$ represent the portion of the wavelet analysis attributable to

scale λ_j, while $W^T_j \boldsymbol{w}_j$ is the portion of the wavelet synthesis attributable to scale λ_j. For a length $N = 2^J$ vector of observations, the final wavelet detail $\boldsymbol{d}_{J+1} = V^T_J \boldsymbol{V}_J$ is equal to the sample mean of the observations.

A wavelet multiresolution analysis (MRA) may now be defined via,

$$x_t = \sum_{j=1}^{J+1} d_{j,t} \text{ , } t = 0, \ldots, N\text{-}1. \qquad 3.33$$

That is, each observation x_t is a linear combination of wavelet detail coefficients. Let $s_j = \sum_{k=j+1}^{J+1} d_k$ define the *jth* level *wavelet smooth* for $0 \le j \le J$, where $\boldsymbol{s}_{J+1}$ is defined to be a vector of zeros. Whereas the wavelet detail $\boldsymbol{d}_j$ is associated with variations at a particular scale, $\boldsymbol{s}_j$ is a cumulative sum of variations and will be smoother and smoother as j increases. In fact $x - s_j = \sum_{k=1}^{j} d_k$ so that only lower scale details (high-frequency features) will be apparent. The *jth* level wavelet rough characterizes the remaining lower-scale details through $r_j = \sum_{k=1}^{j} d_k$ for $1 \le j \le J+1$, where $\boldsymbol{r}_0$ is defined to be a vector of zeros. A vector of observations may thus be decomposed through a wavelet *smooth* and *rough* via,

$$x = \boldsymbol{s}_j + \boldsymbol{r}_j, \qquad 3.34$$

for all j. Percival and Walden use the terminology "detail" and "smooth" to describe additive decompositions from wavelet transforms [118].

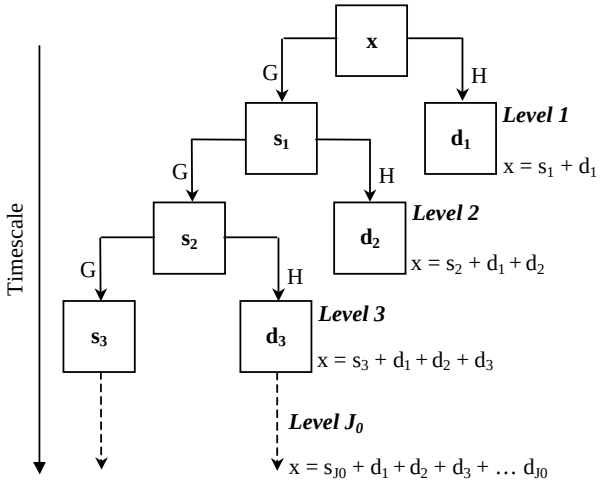


Figure 3.8 Flow diagram of Mallat's pyramidal algorithm for wavelet MRA.

According to Percival and Walden, for a level J_0 decomposition, we may write Eq (3.34) as,

$$x = \sum_{j=1}^{J_0} d_j + s_{J0}\,, \qquad 3.35$$

where, as before, the detail $\boldsymbol{d}_j$ represents changes on a scale of $\lambda_j = 2^{j-1}$ while the smooth $\boldsymbol{s}_{J0}$ represents averages of a scale of $2\lambda_{J0} = 2^{J0}$. Based on the wavelet MRA defined by Eq (3.35), we present a generalized flow diagram of Mallat's pyramidal algorithm in Figure 3.8. Here, *G* and *H* represent a bank of low- and high-pass filters respectively. It is also clear that as we keep on decomposing the output from the low-pass filter banks, the frequency of the resulting signal keeps on decreasing or conversely the timescale keeps on increasing.

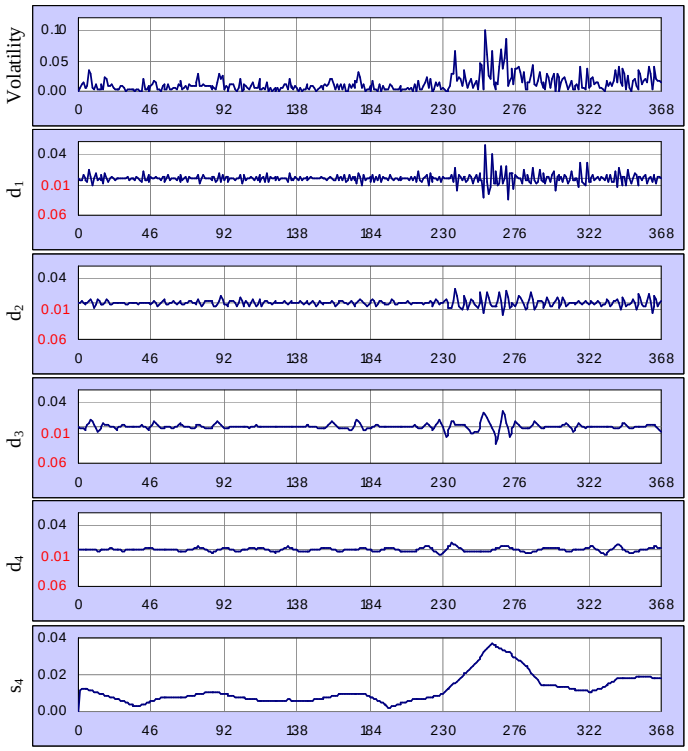


Figure 3.9 DWT MRA of IBM volatility series using D(6) wavelet filters. *Negative values are shown in red.*

We now examine the volatility $v_t = |r_t| = |ln\ (p_t) - ln\ (p_{t-1})|$ of the IBM stock prices through a DWT MRA. The volatility series is plotted in the top row of Figure 3.9 and just as the returns it exhibits a burst in volatility between observation 230 and 322. A DWT MRA using the D(6) wavelet filter is shown below the volatility series in Figure 3.9. The five rows below the input data display the first four wavelet details and wavelet smooth that form an additive decomposition via Eq (3.35). That is at each time index *t*, adding up the wavelet detail coefficients $d_{j,t}$, $j = 1, \ldots, 4$, and the wavelet smooth $s_{4,t}$ will produce the coefficient v_t. All the information contained in the volatility series has been perfectly captured in the MRA and no anomalies have been introduced as a result of the procedure. According to the D(6) MRA, the large fluctuations in volatility occur across the first three scales (d_1, d_2, and d_3) between trading days 230 to 322. There is a lack of

volatility activity at the fourth scale (d_4). However, the broadband feature of the volatility burst is succinctly captured by the wavelet smooth (s_4).

3.2.3 Maximal Overlap Discrete Wavelet Transform

In the previous section we introduced the DWT as an alternative to the DFT for time series analysis. The DWT generates coefficients that are local in both time and frequency. The maximal overlap discrete wavelet transform (MODWT) is a variant of the DWT. The term maximal overlap comes from the relationship of the MODWT with estimators of the Allan variance [2], [117]. In the computation of the MODWT, the filtered output is *not* subsampled as in the computation of the DWT (Figure 3.6a). The MODWT gives up orthogonality in order to gain features that the DWT does not posses. As a result, in the MODWT, the wavelet and scaling coefficients must be rescaled to retain the variance preserving property of the DWT.

Percival and Walden provide the following properties that are important in distinguishing the MODWT from the DWT [118]:

1. The MODWT can handle any sample size, whereas the J_p*th* order DWT restricts the sample size to a multiple of 2^{J_P}.

2. The detail and smooth coefficients of a MODWT MRA are associated with zero phase filters. This means that temporal events and patterns in the original signal are meaningfully aligned with the features in the MRA.

3. The MODWT is invariant to circularly shifting the original time series. Hence, shifting the time series by an integer unit will shift the wavelet and scaling coefficients by the same amount. This property does not hold for the DWT because of the subsampling involved in the filtering process.

4. While both DWT and MODWT can perform an analysis of variance (ANOVA) on a time series, the MODWT wavelet variance estimator is asymptotically more efficient than the same estimator based on the DWT [116].

Let $\boldsymbol{x}$ be an arbitrary length N vector of observations. The length $(J + 1)N$ vector of MODWT coefficients $\tilde{w}$ is obtained via,

$$\tilde{w} = \tilde{W}x, \tag{3.36}$$

where $\tilde{W}$ is a $(J + 1)N \times N$ matrix defining the MODWT. The vector of MODWT coefficients may be organized into $J + 1$ vectors via,

$$\tilde{w} = [\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_J, \tilde{v}_J]^T, \tag{3.37}$$

where $\tilde{w}_j$ is a length *N* vector of wavelet coefficients associated with changes on a scale of length $\lambda_j - 2^{j-1}$ and $\tilde{v}_J$ is a length *N* vector of scaling coefficients associated with averages on a scale of $2\lambda_J = 2^J$, just as with the DWT.

Similar to the orthonormal matrix defining the DWT, the matrix $\tilde{W}$ is also made up of $J + 1$ submatrices, each of them $N \times N$, and may be expressed as,

$$\tilde{W} = \begin{bmatrix} \tilde{W}_1 \\ \tilde{W}_2 \\ : \\ \tilde{W}_J \\ \tilde{V}_J \end{bmatrix}. \qquad 3.38$$

Instead of using the wavelet and scaling filters defined for the DWT, the MODWT utilizes the rescaled filters ($j = 1, \ldots, J$)

$$\tilde{h}_j = h_j / 2^j \text{ and } \tilde{g}_J = g / 2^J, \qquad 3.39$$

To construct the $N \times N$ dimensional submatrix $\tilde{W}_1$, we circularly shift the rescaled wavelet filter vector $\tilde{h}_1$ by integer units to the right such that,

$$\tilde{W}_1 = \left[\tilde{h}_1^{(1)}, \tilde{h}_1^{(2)}, \tilde{h}_1^{(3)}, \ldots, \tilde{h}_1^{(N-2)}, \tilde{h}_1^{(N-1)}, \tilde{h}_1\right]^T. \qquad 3.40$$

The above matrix may be interpreted as interweaving of the DWT submatrix W_1 with a circularly shifted (to the right by one unit) version of itself. The remaining submatrices $\tilde{W}_2, \ldots, \tilde{W}_J$ are formed just like Eq (3.40) where $\tilde{h}_1$ is replaced by $\tilde{h}_j$.

Just like the DWT, a pyramidal algorithm is utilized to compute the MODWT wavelet and scaling coefficients. We start with the data x_t (which is no longer restricted to be of dyadic length) and filter it using $\tilde{h}_1$ and $\tilde{g}_1$ to obtain the length *N* vectors of wavelet and scaling coefficients, $\tilde{w}_1$ and $\tilde{v}_1$ respectively. This is similar to the operation described in Figure 3.6a except the fact that no downsampling is performed.

For each iteration of the MODWT pyramidal algorithm, we require three objects: the data vector $\boldsymbol{x}$, the wavelet filter $\tilde{h}_l$ and the scaling filter $\tilde{g}_l$. The first iteration of the pyramidal algorithm begins by filtering (convolving) the data with each filter to obtain the wavelet and scaling coefficients,

$$\tilde{w}_{1,t} = \sum_{l=0}^{L-1} \tilde{h}_l x_{t-l} \bmod N \text{ and } \tilde{v}_{1,t} = \sum_{l=0}^{L-1} \tilde{g}_l x_{t-l \bmod N}, \quad 3.41$$

where $t = 0, 1, \ldots, N-1$. The length N vector of observations has been high- and low-pass filtered to obtain N (redundant) coefficients associated with this information. The second stage of the MODWT pyramidal algorithm starts by defining the data to be the scaling coefficients $\tilde{v}_1$ from the first iteration and applying the above filtering operations to obtain the second level of wavelet and scaling coefficients,

$$\tilde{w}_{2,t} = \sum_{l=0}^{L-1} \tilde{h}_l \tilde{v}_{1,t-l \bmod N} \text{ and } \tilde{v}_{2,t} = \sum_{l=0}^{L-1} \tilde{g}_l \tilde{v}_{1,t-l \bmod N}, \quad 3.42$$

where $t = 0, 1, \ldots, N-1$. Keeping all vectors of wavelet coefficients and, the final level of scaling coefficients, we have the following length N decomposition: $\tilde{w} = [\tilde{w}_1 \tilde{w}_2 \tilde{v}_2]^T$. Similarly after the third iteration of the pyramidal algorithm where we apply filtering operations to $\tilde{v}_2$ we have $\tilde{w} = [\tilde{w}_1 \tilde{w}_2 \tilde{w}_3 \tilde{v}_3]^T$. The above procedure may be repeated J times where $J = log_2(N)$ to get the MODWT coefficients of Eq (3.37).

Inverting the MODWT is achieved through convolving the final level of wavelet and scaling coefficients with their respective filters (wavelet for wavelet and scaling for scaling) and adding up the two filtered vectors. Starting with the final level of the MODWT, the vectors $\tilde{w}_J$ and $\tilde{v}_J$ are filtered and combined to produce $J-1$ vector of scaling coefficients $\tilde{v}_{J-1}$, given by,

$$\tilde{v}_{J-1,t} = \sum_{l=0}^{L-1} \tilde{h}_l \tilde{w}_{J,t+l \bmod N} + \sum_{l=0}^{L-1} \tilde{g}_l \tilde{v}_{J,t+l \bmod N}, \quad 3.43$$

where $t = 0, 1, \ldots, N-1$. We notice that the length of $\tilde{v}_{J-1}$ is the same as $\tilde{v}_J$. The above procedure may be repeated until the first level of wavelet and scaling coefficients have been combined to produce the original vector of observations; that is,

$$x_t = \sum_{l=0}^{L-1} \tilde{h}_l \tilde{w}_{l,t+l \bmod N} + \sum_{l=0}^{L-1} \tilde{g}_l \tilde{v}_{l,t+l \bmod N} \quad 3.44$$

where $t = 0, 1, \ldots, N-1$.

Let us consider an example to illustrate the computation of the MODWT. Figure 3.10 shows a level $J = 4$ MODWT performed on the IBM return series using the Haar wavelet filters. We notice that there are N wavelet coefficients at each scale because the MODWT does not subsample after filtering. The first scale of wavelet coefficients $\tilde{w}_1$ contain the DWT coefficients of Figure 3.7, scaled by $1/\sqrt{2}$, and also the DWT coefficients applied to $\boldsymbol{x}$ circularly shifted by

one. This means that the MODWT coefficients are correlated and will appear smoother than the DWT coefficients. This is more noticeable when we move to higher scales (lower frequencies).

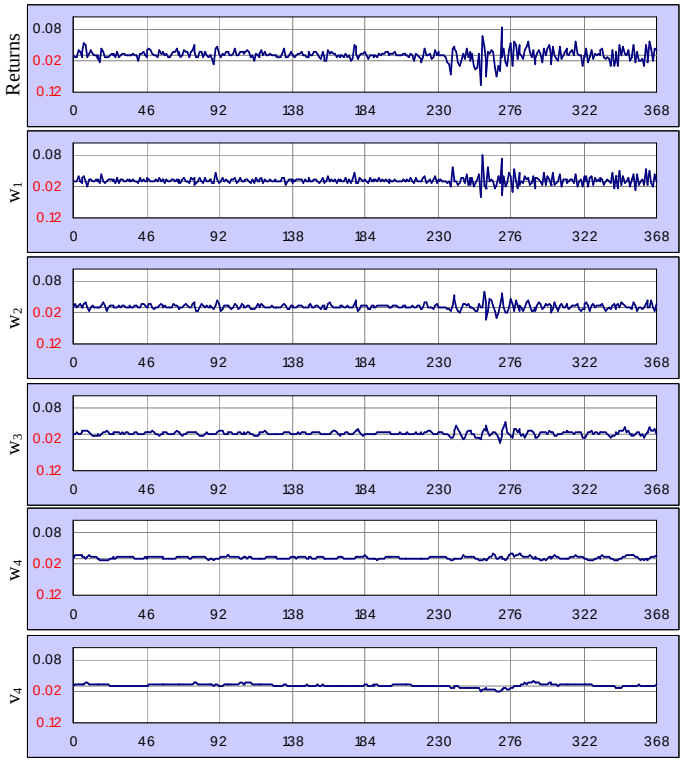


Figure 3.10 MODWT decomposition of the IBM return series (*N = 368*) using the Haar wavelet filters. *Negative values are shown in red.*

Multiresolution Analysis

An analogous MRA to that of the DWT may be performed using the MODWT via,

$$x_t = \sum_{j=1}^{J+1} \tilde{d}_{j,t} \text{ , } t = 0, \ldots, N\text{-}1, \qquad 3.45$$

where $\tilde{d}_{j,t}$ is the tth element of $\tilde{d}_j = \tilde{W}_j^T \tilde{w}_j$ for $j = 1, \ldots, J$. We may also define the MODWT-based wavelet smooths and roughs to be,

$$\tilde{s}_{J,t} = \sum_{k=J+1}^{J+1} \tilde{d}_{k,t} \text{ and } \tilde{r}_{j,t} = \sum_{k=1}^{j} \tilde{d}_{k,t} \text{ , } t = 0, \ldots, N\text{-}1, \qquad 3.46$$

respectively. Walden and Percival define the level J_0 MODWT MRA as [118],

$$x = \sum_{j=1}^{J_0} \tilde{d}_j + \tilde{s}_{J_0} \text{ , } \qquad 3.47$$

where $\tilde{d}_j$ is the jth level detail while $\tilde{s}_{J_0}$ is the J_0th level smooth. A key feature to the MODWT MRA is that the wavelet details and smooth are associated with zero-phase filters. Thus,

interesting features in the wavelet details and smooth may be perfectly aligned with the original time series. This property is not possessed by the DWT since it subsamples the output of its filtering operations (Figure 3.6a).

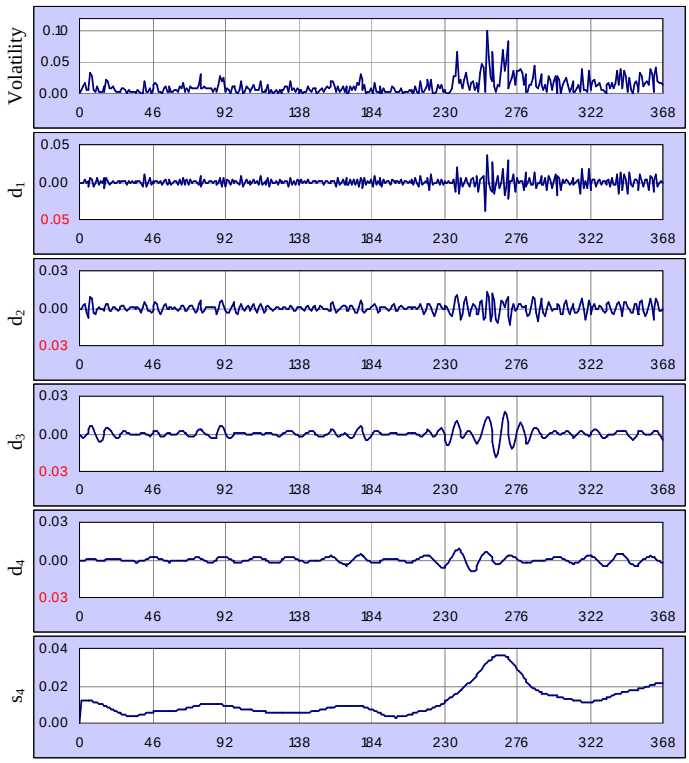


Figure 3.11 MODWT MRA of IBM volatility series using D(6) wavelet filters. *Negative values are shown in red.*

Figure 3.11 shows a MODWT MRA performed on the IBM volatility series using the D(6) wavelet filter. When compared with Figure 3.9, there does not seem to be any obvious "contamination" by the wavelet basis function. The correlation between coefficients is effectively smooth over features specific to the wavelet filter. This implies that even if we use non-smooth filters like Haar we will observe reasonably smoother wavelet details and smooths. Since $\tilde{d}_j$ and $\tilde{s}_j$ are associated with zero phase filters, features in the original time series are meaningfully aligned with the MRA. Thus the burst in volatility between 230 and 322 is perfectly captured in time by the MODWT wavelet details. The wavelet smooth shows a significant bump in the same region, indicating that this volatility change is a broadband feature of the volatility series.

3.3 Wavelet Variance and Covariance

The classical method of decomposing the variance of a stationary time series x_t is through the *spectral representation theorem* [125]. The spectral representation theorem allows us to express a discrete stationary process as an infinite sum of complex exponentials with random amplitudes

and phases, which is called the integrated spectrum *I (f)*. If *I (f)* is differentiable everywhere then its derivative *S(f)* is defined to be the *spectrum* or *spectral density function* (SDF). If the autocovariance sequence (ACVS*) of x_t is square summable, that is $\sum \gamma_{x,\tau}^2 < \infty$, then the SDF of x_t may be expressed as the Fourier transform of its ACVS via,

$$S_x(f) = \sum_{\tau=-\infty}^{\infty} \gamma_{x,\tau} e^{-i2\pi f\tau}, \text{ for } |f| \le 1/2. \qquad 3.48$$

Eq (3.48) implies that the spectrum exhibits large coefficients when the sinusoid associated with a particular frequency *f* is highly correlated with $\gamma_{x,\tau}$.

***A Note on Autocovariance Sequence (ACVS)**

Let $\{X_t : t = \ldots, -1, 0, 1, \ldots\}$ be a discrete parameter real-valued stochastic process, which by definition is a sequence of random variables indexed over the integers. When the independent variable *t* is instead taken to vary over the entire real axis, the stochastic process is said to have a continuous parameter.

- *Stationarity and autocovariance sequence*

The process $\{X_t\}$ is said to be second order stationary if it satisfies the following two properties:

(1) $E\{X_t\} = \mu_X$ for all integers *t*; i.e., the expected value of the t^{th} component X_t of the process $\{X_t\}$ is equal to a finite constant μ_X which does not depend on *t*; and

(2) $\text{Cov}\{X_t, X_{t+\tau}\} = s_{X,\tau}$ for all integers *t* and τ; i.e., the covariance between any two components X_t and $X_{t+\tau}$ of the stationary process $\{X_t\}$ is equal to a finite constant $s_{X,\tau}$ which depends only on the separation τ between the indices *t* and $t + \tau$ of the components.

The sequence $\{s_{X,\tau} : \tau = \ldots, -1, 0, 1, \ldots\}$ is called the *autocovariance sequence* (ACVS).

The wavelet transform has proven to be very effective in studying local features in a time series. The wavelet decomposition offers a good local representation of a time series since it efficiently partitions the time-frequency plane by using short basis functions for high-frequency oscillations and long basis functions for low-frequency oscillations (Figure 3.3). An important characteristic of the DWT and the MODWT is their ability to decompose (analyze) the variance of a stochastic process.

In the previous section we discussed that a vector of wavelet coefficients is associated with changes at a particular timescale. For example, the level *j* wavelet coefficients, $w_{j,t}$, are associated with changes on a scale of $\lambda_j = 2^{j-1}$. Applying the DWT to a stochastic process produces a decomposition on a scale-by-scale basis. By decomposing a given time series into wavelet sub-series, we now look at the computation of the variance of a stochastic process on a scale-by-scale basis.

Let x_t be a real-valued stochastic process, which may not have a zero mean or be stationary. The *time-varying wavelet variance* for x_t is defined to be the variance of the scale λ_j wavelet coefficient $w_{j,t}$, given by,

$$\sigma_{x,t}^2(\lambda_j) = \frac{1}{2\lambda_j} Var(w_{j,t}). \quad 3.49$$

If we assume that the wavelet variance is independent of *t*, that is x_t is a stationary process, then we may write the time-dependent wavelet variance of Eq (3.49) as the *time-independent wavelet variance* or simply the *wavelet variance*,

$$\sigma_x^2(\lambda_j) = \frac{1}{2\lambda_j} Var(w_{j,t}). \quad 3.50$$

While the SDF decomposes a process on a frequency-by-frequency basis, the wavelet variance decomposes the variance of x_t on a scale-by-scale basis via [116],

$$\sum_{j=1}^{\infty} \sigma_x^2(\lambda_j) = Var(x_t). \quad 3.51$$

The scale λ_j is associated with the frequency interval $[1/2^{j+1}, 1/2^j]$, and we can use this property to obtain an approximate relationship between the wavelet variance and the SDF of x_t via,

$$\sigma_{x,t}^2(\lambda_j) \approx 2\int_{1/2^{j+1}}^{1/2^j} S_x(f)df \ . \quad 3.52$$

The factor 2 is needed because the spectrum is an even function of frequency over the interval $[-1/2, 1/2]$.

Estimation of the Wavelet Variance

Let us consider a dyadic length $N = 2^J$ realization $\boldsymbol{x} = (x_0, x_1, \ldots, x_{N-1})$ of the stochastic process x_t and apply the DWT of order $J_p \le J$ to produce the length N vector of wavelet coefficients $\boldsymbol{w}$. An unbiased estimation of the wavelet variance is based on the MODWT using,

$$\hat{\sigma}_x^2(\lambda_j) = \frac{1}{2\lambda_j \hat{N}_j} \sum_{t=L_j'}^{N/2^{j-1}} w_{j,t}^2 \ , \quad 3.53$$

where $L_j' = \left| (L-2)(1-2^{-j}) \right|$ is the number of DWT coefficients computed using the boundary and hence $\hat{N}_j = N/2^j - L_j'$ is the number of wavelet coefficients at scale λ_j that are unaffected by the boundary.

If we relax the requirement of a dyadic sample size ($N = 2^J$), we may compute the MODWT of order $J_p < log_2(N)$ to produce the length $(J_p + 1)N$ vector of wavelet coefficients $\tilde{w}$. An unbiased estimation of the wavelet variance based on the MODWT is given by,

$$\tilde{\sigma}_x^2(\lambda_j) = \frac{1}{\tilde{N}_j} \sum_{t=L_j-1}^{N-1} \tilde{w}_{j,t}^2 , \qquad 3.54$$

where $L_j = (2^j - 1)(L - 1) + 1$ is the length of the scale λ_j wavelet filter and $\tilde{N}_j = N - L_j + 1$ is the number of coefficients that are unaffected by the boundary. The renormalization with $2\lambda_j$ is not required in Eq (3.54) since filters $\tilde{h}_l$ associated with the MODWT are rescaled versions of the wavelet filters h_l (Section 3.2.3).

Estimation of the SDF

We have studied how to compute the wavelet variance of a stochastic process x_t (Eqs (3.53 & 3.54)). We now want to be able to compute the SDF based on the wavelet variance. Let us reconsider Eq (3.52) which relates the wavelet variance to the SDF. Under the assumption that L, the width of the wavelet filter, is selected such that Eq (3.52) is a reasonable approximation, we can estimate $S_x(.)$ using a function $\bar{S}_x(.)$ that is piecewise constant over each interval $[1/2^{j+1}, 1/2^j]$ for $j = 1, 2, \ldots, J_0$. That is, we assume,

$$\bar{S}_x(f) = C_j \text{ when } \frac{1}{2^{j+1}} < f \le \frac{1}{2^j} \qquad 3.55$$

where C_j is a constant defined such that,

$$\int_{1/2^{j+1}}^{1/2^j} S_x(f)df = \int_{1/2^{j+1}}^{1/2^j} \bar{S}_x(f)df = \frac{C_j}{2^{j+1}} \qquad 3.56$$

Substituting the value of Eq (3.56) in Eq (3.52), we have,

$$\sigma_{x,t}^2(\lambda_j) \approx \frac{C_j}{2^j} . \qquad 3.57$$

Hence we can use,

$$\hat{C}_j \equiv 2^j \hat{\sigma}_x^2(\lambda_j) , \qquad 3.58$$

to estimate the spectral density levels or the SDF based on the wavelet variance.

Estimation of the Wavelet Covariance

We have studied the computation of the wavelet variance for univariate (one-variable) time series. Here, we will investigate a multivariate generalization, called *wavelet covariance* for bivariate time series. We will examine how wavelet methodology can be extended to study two concurrently recorded time series.

Suppose,

$\boldsymbol{X} = (X_0, X_1, \ldots, X_{N-1})$

$= ((x_{1,0}, x_{2,0}), (x_{1,1}, x_{2,1}), \ldots, (x_{1,N-1}, x_{2,N-1}))$

is a length N realization of a bivariate stochastic process X_t. Applying a MODWT of order $J_p < log_2 (T)$ to each univariate process $x_{1,t}$ and $x_{2,t}$ would yield J length N vectors of MODWT coefficients,

$$\begin{aligned}\tilde{W}_j &= (\tilde{W}_{j,0}, \tilde{W}_{j,1}, \ldots, \tilde{W}_{j,N-1}) \\ &= \left((\tilde{w}_{1,j,0}, \tilde{w}_{2,j,0}), (\tilde{w}_{1,j,1}, \tilde{w}_{2,j,1}), \ldots, (\tilde{w}_{1,j,N/2^j-1}, \tilde{w}_{2,j,N/2^j-1})\right)\end{aligned}$$

and length N vector of MODWT scaling coefficients,

$$\begin{aligned}\tilde{V}_J &= (\tilde{V}_{J,0}, \tilde{V}_{J,1}, \ldots, \tilde{V}_{J,N-1}) \\ &= \left((\tilde{v}_{1,J,0}), (\tilde{v}_{2,J,0}), \ldots, (\tilde{v}_{1,j,N/2^J-1}, \tilde{v}_{2,j,N/2^J-1})\right).\end{aligned}$$

An unbiased estimation of the wavelet covariance based on the MODWT is given by,

$$\tilde{\gamma}_X(\lambda_j) = \frac{1}{\tilde{N}_j} \sum_{l=L_j-1}^{N-1} \tilde{w}_{1,j,l}, \tilde{w}_{2,j,l} \,. \qquad 3.59$$

where $\tilde{N}_j = N - L_j + 1$. This estimate does not include any coefficients affected by the boundary. We can also construct a biased estimation of the wavelet covariance by simply including the MODWT wavelet coefficients affected by the boundary and a renormalization term in Eq (3.59).

A MODWT-based estimation of covariance is generally preferred over the DWT-based estimation of covariance. This is because the DWT-based estimation $\gamma_X(\lambda j)$ depends on the lag between $x_{1,t}$ and $x_{2,t}$, whereas the variance of the MODWT-based estimation is invariant to lag between two series. Since the lag between two time series is generally not known, it is important to have an estimator that is invariant to circular translation and hence the MODWT-based estimation of covariance [54].

3.3.1 Power Laws

An introduction to power law behaviour in time series was provided in Section 2.2.5. A power law behaviour was demonstrated for the S&P 500 time series using frequency domain analysis, more specifically the Fourier power spectrum analysis (Figure 2.15). We also discussed that a Fourier-based power law analysis might not be suitable for studying nonstationary time series. In this section we provide mathematical details for the computation of the power law. We investigate, with specific examples, how we can use wavelet-based ANOVA to study power law behaviour in complex time series.

Power law behaviour may be described by the following general equation,

$$f(x) = \beta x^{\alpha}, \tag{3.60}$$

where β and α are constants. Taking a logarithmic on both sides of Eq (3.60), a straight line (graph *log f [x]* vs *log x*) of slope α and intercept *log* β is obtained,

$$\log f(x) = \log(\beta x^{\alpha}) = \log \beta + \log x^{\alpha} = \log \beta + \alpha \log x, \tag{3.61}$$

where α is the power law exponent. From Eq (3.61) it implies that the power law behaviour is scale invariant, that is, if a variable x is replaced by Ax', where A is a constant, then the fundamental power law relationship remains unaltered. In frequency domain analysis, if dynamics follow a power law, then a log-log representation of the Fourier power spectrum (log power versus log frequency) reveals a straight line which is always within a defined range consistent with the size and duration of the system. We saw this in Figure 2.15 for the S&P 500 index.

One way to study power law behaviour in nonstationary time series using the wavelet analysis instead of the Fourier analysis is to plot the log-log representation of the wavelet variance as a function of timescale (log variance versus log timescale). A straight line fit to the log variance versus log timescale plot would suggest the existence of a power law behaviour. Such an analysis, as we will see, can help us assess the state and stability of complex systems. A perfect power law behaviour across many scales (a straight line fit to the log variance-timescale plot) might indicate that the system is in the optimum state whereas a break in power law behaviour (different straight line fits to different sections of the log variance-timescale plot) may point towards anomalies in the system.

As discussed in Section 3.3 the classical method for studying the variance of a time series x_t is through the SDF, which is defined by Eq (3.48). Theory suggests that a process $\{x_t\}$ is a pure power law (PPL) if its SDF has the form,

$$S_x(f) = C_S |f|^{\alpha}, \; -1/2 \le f \le 1/2 \tag{3.62}$$

where $C_S > 0$. These processes can be divided into two categories corresponding to stationary and nonstationary processes [118]. The first corresponds to the case $-1 < \alpha$, which means that the process is stationary. Also, when $-1 < \alpha < 0$, a PPL process is said to be a stationary long memory process. The second category applies when $\alpha \le -1$: such a PPL process $\{x_t\}$ can be interpreted as a nonstationary process that can be turned into a stationary process through an appropriate differencing operation. PPL processes are thus well defined for any α on the real axis and these processes are sometimes called "*1/f processes*". Hence long memory processes are 1/f-type processes with $-1 < \alpha < 0$, while *1/f-processes* with $\alpha \le -1$ are sometimes called nonstationary long memory processes.

The wavelet variance is a useful alternative to the SDF for studying PPL processes, since it provides a way of "regularizing" the SDF. We have studied that the wavelet coefficients at scale λ_j are associated with the frequency interval $[1/2^{j+1}, 1/2^j]$, and we can use this property to obtain an approximate relationship between the wavelet variance and the SDF of x_t via Eq (3.52). The wavelet variance thus summarizes the information in the SDF using just one value per octave frequency band and is particularly useful when the SDF is relatively featureless within each octave band. Suppose, for example, that $\{x_t\}$ is a PPL process, i.e. its SDF is given by $S_x(f) = C_S |f|^{\alpha}$. If we use the approximation of Eq (3.52), we see that,

$$\sigma_x^2(\lambda_j) \propto \lambda_j^{-\alpha-1}. \qquad 3.63$$

Linear variation on a plot of $\log(\sigma_x^2(\lambda_j))$ versus $\log(\lambda_j)$ thus indicates the existence of a power law process and the slope of the line can be used to deduce the exponent α of the power law. The estimation of the power law exponent α through the wavelet variance rather than the SDF is useful because for non-periodic and nonstationary processes whose Fourier power spectra are relatively featureless, the wavelet variance is an attractive alternative characterization that is easy to interpret and estimate.

Let us now examine a 'time' series of vertical ocean shear measurements that has been studied in [118]. This data was collected by an instrument that is dropped over the side of a ship and descends vertically into the ocean. As the probe descends, it collects measurements of the ocean shear as a function of the depth. Thus the ordering variable of this 'time' series is depth. The shear measurements are made every 0.1 meters, hence the sampling rate $\Delta t = 0.1$. This 'time' series is shown in Figure 3.12.

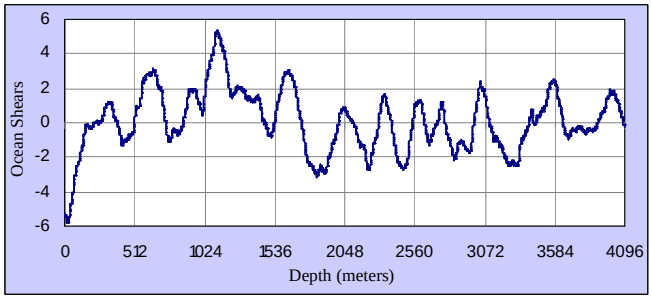


Figure 3.12 Series of vertical ocean shear measurements.

We notice that the sample size of our series is 4096, hence a wavelet transform of level $J_0 = 12$ ($2^{12} = 4096$) would be appropriate. We now compute the unbiased MODWT Haar wavelet variance estimates $\tilde{\sigma}_x^2(\lambda_j)$ for levels $j = 1$ to 12 using Eq (3.54).

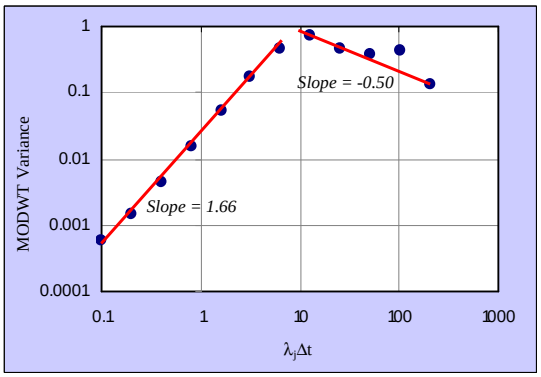


Figure 3.13 MODWT variances for the vertical ocean shear series using the Haar wavelet filter.

Since coefficients at *level 1* correspond to changes on a scale of $\lambda_1 = 2^{1-1}$ and coefficients at level *2* correspond to changes on a scale of $\lambda_2 = 2^{2-1}$ and so on, and the sampling rate $\Delta t = 0.1$, the physical scales for the MODWT variance will range from $\lambda_1\Delta t = 0.1$ up to $\lambda_{12}\Delta t = 204.8$ meters. Figure 3.13 shows the *log-log* plot of the MODWT variance $\tilde{\sigma}_x^2(\lambda_j)$ versus the physical scales $\lambda_j\Delta t$. As discussed earlier, regions of linearity on a plot such as Figure 3.13 indicate the presence of a power law process over a particular region of frequencies, with the power law exponent being related to the slope of the line. We see from Figure 3.13 that the Haar wavelet variance estimates for the smallest seven scales fall perfectly along such a straight line. The line drawn through them was calculated via linear least squares and has a slope of 1.66 = 5/3. Since $\tilde{\sigma}_x^2(\lambda_j)$ varies approximately as $\lambda_j^{-\alpha-1}$ for power law processes for exponent α (Eq (3.63)), the Haar wavelet variance plot suggests the presence of a power law process over physical scales 0.1 to 6.4 meters with an exponent of $\alpha = -2.66$. More specifically, we can summarize the power law behaviour in the vertical ocean shear series as,

$$\tilde{\sigma}_x^2(\lambda_j) \approx \begin{cases} -1.64\lambda_j^{+1.66} & , \quad j = 1, 2, \ldots, 7; \\ +0.43\lambda_j^{-0.50} & , \quad j = 8, 9, \ldots, 12. \end{cases} \qquad 3.64$$

The above translate into,

$$S_x(f) \propto \begin{cases} |f|^{-2.66} & , \quad 1/256 < |f| \leq 1/2; \\ |f|^{-0.50} & , \quad 1/8192 < |f| \leq 1/256. \end{cases} \qquad 3.65$$

A point to note here is that we have used the Haar wavelet filter to demonstrate power law behaviour in the ocean shear series. Similar results have also been observed with Duabechies filters like the LA(8) and D(4) and D(6) [118].

3.3.2 Multiscale CAPM

The CAPM of Sharpe and Lintner [136], [94] was introduced in Section 2.3.2. We recall that the risk of a project is measured by the *beta* of the cash flow with respect to the market return. According to Eq (2.34), the *beta* of a company is defined as,

$$\beta_i = \frac{Cov(r_{it}, r_{mt})}{\sigma_m^2}, \qquad 3.66$$

where *Cov* (r_{it}, r_{mt}) is the covariance between the return r_{it} on investing in project *i* and market return r_{mt}. The variance of the market return is denoted by σ_m^2. A low *beta* for a company (or project) would indicate a relatively low risk, whereas a *beta* greater than one would indicate a risky investment.

Wavelet covariance and wavelet variance can be utilized to estimate company *betas* at different timescales [55]. Specifically, a zero-lag wavelet covariance (Eq (3.59)) between daily return for an asset and the daily return in the market at a particular detail (or level) can be divided by the wavelet variance (Eq (3.54)) of the market return at that specific detail (or level) to obtain the *beta* of a company at a particular *timescale*. Mathematically, we may express Eq (3.66) as,

$$\beta_{ij}^w = \frac{Cov(\tilde{w}_{ij}, \tilde{w}_{mj})}{\sigma_{mj}^2}, \qquad 3.67$$

to obtain wavelet *betas*, β_{ij}^w, at different scales *j*. In the right-hand side of Eq (3.67), $\sigma_{mj}^2 = Var(\tilde{w}_{mj})$ is the wavelet variance of the market return at the wavelet scale *j* and $Cov(\tilde{w}_{ij}, \tilde{w}_{mj})$ is the wavelet covariance at wavelet scale *j* between individual stock return and the market return.

In Table 3.2 we present wavelet-based *betas* for six different companies: Coca Cola, General Electric (GE), General Motors (GM) International Business machines (IBM), Proctor & Gamble (P&G), and Boing which belong to the composite index Dow Jones Industrial Average (DJIA). We obtained the daily prices from Yahoo Finance ® for the sample period of January 4, 1999 to November 8, 2005. The sample size is 1718 days, roughly seven years. Market return was taken as the daily return of the DJIA for the specified time period. If we recall Eq (2.36), in empirical finance, we must calculate the *beta* of an asset based on the difference of the return of an asset and the *riskless* rate of return ($r_{it} - r_{0t}$) which is called the *stock premium* and the difference of the market return and the *riskless* rate of return ($r_{mt} - r_{0t}$) which is called the *market premium*. In our calculation, the *riskless* rate of return r_{0t} has been assumed to be the *10-Year Treasury Note* [45], the data for which is also available at Yahoo Finance ®.

Table 3.2 MODWT-based estimates of *betas* for different companies.

	Coca Cola	**GE**	**GM**	**IBM**	**P&G**	**Boing**
Raw Data	0.85	1.09	1.00	1.00	0.91	0.95
λ_1	0.84	1.07	1.02	1.06	0.89	0.94
λ_2	0.90	1.08	0.97	0.98	0.92	0.96
λ_3	0.79	1.18	0.98	0.91	1.00	0.89
λ_4	0.82	1.12	1.06	0.77	0.91	1.17
λ_5	0.87	1.17	1.05	1.01	0.83	1.07
λ_6	0.89	1.24	1.23	1.19	0.74	1.14
λ_7	1.25	0.99	0.62	0.95	0.96	0.69

The daily return of each stock is calculated as the log price difference $r_{it} = log\ P_{it} - log\ P_{it-1}$, where P_{it} is the price of asset i at day t. The daily market return, r_{mt}, is taken as the log difference of the DJIA index $r_{mt} = log\ D_t - log\ D_{t-1}$, where D_t is the index value at day t. Finally the daily *riskless* rate of return r_{0t} is taken as the log difference of the 10-Year Treasury Note $r_{0t} = log\ T_t - log\ T_{t-1}$, where T_t is the treasury note's value at day t. The values of the *market premium* ($r_{mt} - r_{0t}$) and the *stock premium* ($r_{it} - r_{0t}$) are plugged into Eq (3.67) to compute the wavelet-based *multiscale betas*.

In Table3.2, the first detail λ_1 captures oscillations with a period of 2 to 4 days whereas the last detail λ_7 captures oscillations with a period length of 128 to 256 days. Wavelet *betas* at different scales reveal important information about the *riskiness* of a particular asset. For example, raw data (averages of all timescales) for Coca Cola indicates that the stock has relatively low risk (0.85). However, wavelet analysis, especially λ_7 reveals a greater risk for the Coca Cola stock (1.25). An investor operating at scale seven would be subject to high risk by holding a Coca Cola share in his portfolio – that is if an investor has a Coca Cola stock and he intends to sell it after 2^7 (= 128-256) days, he is likely to lose money since the price of Coca Cola is expected to fall considerably by that time. For General Motors, the *betas* are high for the raw data as well as for the first six wavelet scales. However, for the seventh scale the *beta* is relatively lower (0.62). An investor, who operates at timescale λ_7 for General Motors shares, can safely keep this stock in his portfolio. This information would not have been available to an investor if he just looked at the raw *beta* value which is high (1.00). We can put forward a similar argument for IBM stocks which have a low *beta* at scale four (0.77), Proctor & Gamble stocks which have a low *beta* at scale six (0.74), and Boing stocks which have a low *beta* at scale seven (0.69). On the other hand, *betas* for General Electric are consistently higher at all wavelet scales as well as for the raw data. We conclude that there are obvious fluctuations in the *riskiness* of stocks over different timescales and enough care should be taken by investors about the stocks they want to keep in their portfolios.

3.4 Testing for Homogeneity of Variance

Recently, attention has been given to identifying and modelling the so-called long memory processes. In such processes, the correlations between variables do not decay at a sufficiently fast rate. For example in a financial time series, observations separated by great periods of time would still exhibit significant correlation. Such time series are said to be generated by long-memory or long-range dependent processes and require different approaches to modelling than the so-called short-memory processes (for example, ARIMA models). In Section 3.3.1 we saw how power law analysis (more specifically the value of α) can be used to determine whether a time series exhibits long memory or not.

We mentioned earlier, that the wavelet transform has been shown to approximately decorrelate time series with long memory structure. In fact the discrete wavelet transform (DWT) of a long memory process produces sub-series, which are approximately white noise sequences. Features that differ from this long memory structure, such as sudden changes of variance, are retained in certain sub-series of wavelet coefficients.

Whitcher et al. showed that this approximate "decorrelation" property of the DWT and the cumulative sum of squares method (applied to the volatility of a time series) can be used to test for homogeneity of variance, on a scale-by-scale basis, for such long memory processes [160]. This provides a statistically sound technique of testing for nonstationary features without knowing the exact nature of the correlation structure in a given time series. In other words, the DWT can be used to locate a variance change or nonstationarity in a given time series. Sometimes, these nonstationarities or variance changes are also referred to as structural breaks.

Locating a Variance Change

According to Whitcher et al., the normalized cumulative sum of squares (NCSS) test statistic D can be used to detect an unknown change-point k_c in the variance (or volatility $v_t = |r_t| = |ln\ (x_t) - ln\ (x_{t-1})|$) of a given time series x_t [160]. Let the ratio P_k be defined as,

$$P_k = \frac{\sum_{t=L_j-1}^{k} w_{j,t}^2}{\sum_{t=L_j-1}^{N-1} w_{j,y}^2}, \; k = L_j - 1, \ldots, N-2, \qquad 3.68$$

where N is the total number of samples in the volatility series v_t, L is the length of the filter used for the DWT analysis, and $w_{j,t}$ is the *level-j* DWT of the volatility series v_t. Let us define,

$$D = max\ (D^+, D^-), \qquad 3.69$$

where,

$$D^{+} = \max_{k}\left(\frac{k - L_j + 2}{N - L_j} - P_k\right) \text{ and } D^{-} = \max_{k}\left(P_k - \frac{k - L_j + 1}{N - 1}\right). \qquad 3.70$$

The estimated location of the variance change k_c is the point at which D is achieved. In other words, the change-point in the variance k_c is the abscissa of the peak (maximum) value of *max* (D^{+}, D^{-}).

We now examine the change in variance for the daily IBM prices from May 17, 1961 to November 2, 1962. We have studied this series in Section 3.2.2 where Figure 3.9 shows a DWT MRA using D(6) wavelet filters for the volatility of the IBM prices. We want to locate a single change in variance at an unknown time in this series using the NCSS test statistic *D*, described above. There is an obvious increase in volatility in the latter half of the IBM volatility series. We can observe this in Figure 3.9 which shows an obvious flow of increased volatility in the wavelet details (d_1 to d_3) between observations 230 and 322. Moreover, the wavelet smooth (s_4) corresponding to the low frequency trends and the sample mean, shows a significant bump in this region.

For a time series $X \equiv [x_0, x_1, \ldots, x_{N-1}]$ with *N* samples, the null hypothesis H_0 [161], [74], [26] states that,

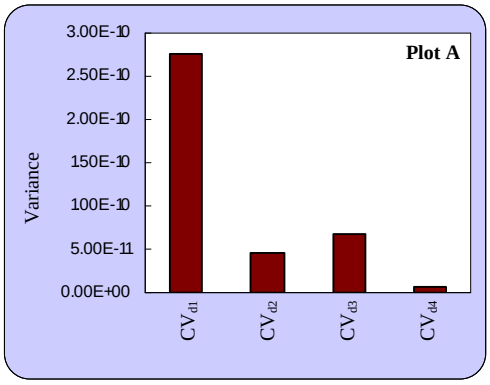
H_0 : var $\{x_0\}$ = var $\{x_0, x_1\}$ = …= var $\{x_0, x_1, \ldots, x_{N-1}\}$,

Based on the above, we can define a cumulative variance (CV) statistic as,

$CV \equiv [\text{var}\{x_0\}, \text{var}\{x_0, x_1\}, \ldots, \text{var}\{x_0, x_1, \ldots, x_{N-1}\}]$.

Ideally, under H_0, the variance of the CV time series should be equal to zero. To test for the null hypothesis, we compute the CV statistic for wavelet details d_1, d_2, d_3, and d_4. We then compute the variances of the CV time series for each wavelet detail d_1 to d_4 which are 2.76E-10, 4.58E-11, 6.81E-11, and 5.56E-12 respectively.

The plots of the variances of the CV time series for the wavelet details are shown in Plot A to the right. As expected the CV_{d4} time series shows the lowest variance, indicating that d_4 constitutes a stationary process as compared to d_1, d_2, and d_3. The *p* values for a one-tailed *F-test* [154] performed between the CV time series of different wavelet details statistically confirm our hypothesis.



Measure	p-value
ftest (CV_{d1}, CV_{d4})	1.26E-206
ftest (CV_{d3}, CV_{d4})	5.37E-104
ftest (CV_{d2}, CV_{d4})	4.83E-078
ftest (CV_{d1}, CV_{d2})	2.57E-059
ftest (CV_{d1}, CV_{d3})	2.98E-038
ftest (CV_{d2}, CV_{d3})	1.52E-004

Figure 3.14a Rejecting the null hypothesis for wavelet details d_1, d_2, and d_3.

Therefore, this change in volatility is a very broadband feature of the IBM volatility series. In Figure 3.9 we explore the extent of this nonstationary feature with the help of the DWT MRA. We observe that the increase in volatility is confined to the first three wavelet details (d_1, d_2 and d_3) only. If we look at d_4, we find no apparent difference in volatility throughout its length. Hence d_4 may be thought of as "stationary" for practical purposes. We can thus reject the null hypothesis for homogeneity of variance for the first three scales (d_1, d_2, and d_3) of IBM stock volatility. That

is nonstationarities or variance changes do exist in the first three wavelet details d_1, d_2 and d_3. Please refer to Figure 3.14a where we present a statistical test for rejecting the null hypothesis for wavelet details d_1, d_2, and d_3 as compared to d_4.

After rejecting the null hypothesis, we now perform a formal test to locate a variance change in the IBM volatility using the NCSS test statistic D defined by Eq (3.69). The results of the analysis are shown in Figure 3.14b.

The NCSS test statistic D successfully picks up the location of the variance change, which occurs at observation number 237. In Figure 3.14b, the volatility series is shown in the top panel while below it is *max* (D^+, D^-) plotted for levels *1* to *3* of the DWT of the IBM volatility series. The peak value of the *level-1* plot for *max* (D^+, D^-) detects the location of the variance change k_c, which is shown by a red vertical line. Since *level-1* wavelet coefficients are associated with highest frequencies (second row from top in Figure 3.14b), its maximum is a good estimate of the time of variance change. Here the wavelet transform allows for a rigorous test of homogeneity of variance, only with mild assumptions on the underlying spectrum of the process.

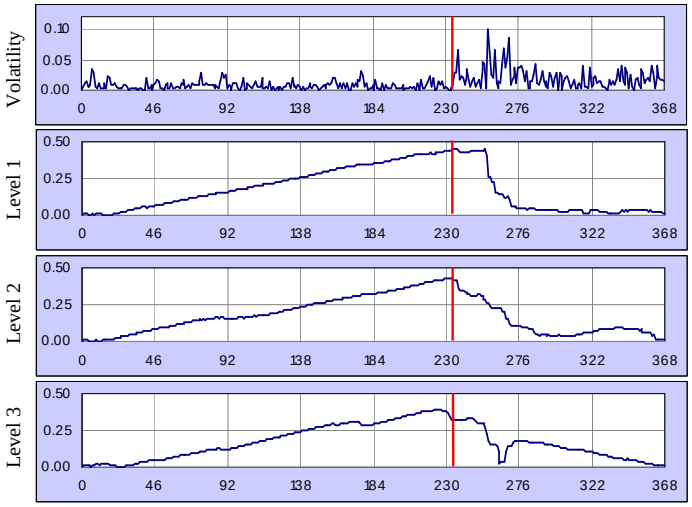


Figure 3.14b IBM stock price volatility (top panel) along with the NCSS statistic *max* (D^+, D^-) for its DWT coefficients for scales 1 to 3.

3.5 Problem Formulation

In Chapter 2 we introduced the concept of "universal scaling" or "self-organized complexity" in complex systems. In this Chapter we described in more detail how the wavelet multiscale analysis can be used to assess time series produced by complex systems. We also showed the existence of power laws and scaling in complex time series and how the DWT can be used to study such phenomena. If we draw an analogy between scaling in complex systems and the manner in which the DWT analyzes scaling in complex time series, we find striking similarities between the two. Let us revisit Figure 2.1 which shows the characteristics of complex systems and Figure 3.8 which

shows the implementation of Mallat's pyramidal algorithm for decomposing complex time series. Figure 2.1 represents a complex system with many *interacting* subunits, each comprising a particular "scale" at a particular level. Similarly, Figure 3.8 represents a complex time series that has been decomposed into many subunits (components). Each wavelet decomposition or subunit describes the behaviour of the original signal at a particular "timescale" or level. It is indeed possible to study how the various wavelet decompositions (sub-series) are *related* to one another, and how the original signal evolves from these decompositions. We have shown this in the power law analysis of time series using wavelets where we study the relationship between variances of different wavelet sub-series. Such an analogy proves the efficacy of wavelets for the analysis of time series emanating from complex systems. Moreover, since the wavelet analysis is a non-parametric method (no parameters need to be set for model building) for analyzing signals in myriad fields, it can be used as a universal tool for studying complex time series. As discussed earlier, according to Ramsey, the wavelet analysis has formalized old notions of decomposing a time series into its various components [127].

In this section we will develop the problem of analyzing complex time series in different domains (more specifically in the financial and medical domain) and then give a brief description of our proposed wavelet-based system for analyzing such time series. We will discuss how the proposed system for time series analysis can be regarded as a *multidisciplinary time series summarization system*.

3.5.1 Pattern Identification

Cycles, trends, and turning points are endemic to financial forecasting [29]. In Section 3.1, we discussed in some detail various components of a time series like cycles, trends and the irregular or random component. Regular patterns like cycles and trends help us to study the so-called "stable features" of an otherwise stochastic time series. Turning points may be defined as critical points on a time series, which might indicate a trend reversal, sudden changes or interesting events. The WOODSONWAVEREPORT shown in Figure 2.6 is a good example of studying turning points in financial data. In Section 3.4 we discussed the detection of variance changes or structural breaks in a time series. Detection of a structural break is important in that it tells us exactly where in time a nonstationarity has occurred, which suggests a total change in the dynamics and behaviour of the series under study. The vertical line marked as *T* in Figure 2.3 is an example of a structural break in financial time series data. A similar example is also presented in Figure 3.14b.

We can regard features like cycles, trends, turning points and structural breaks as patterns in a time series which need to be identified efficiently to understand its behaviour. By looking at these

patterns, traders often make “investment” decisions. Figure 3.15 shows a simplified synthetic time series with cycles, trends, turning points, and a structural break (variance change) marked on it.

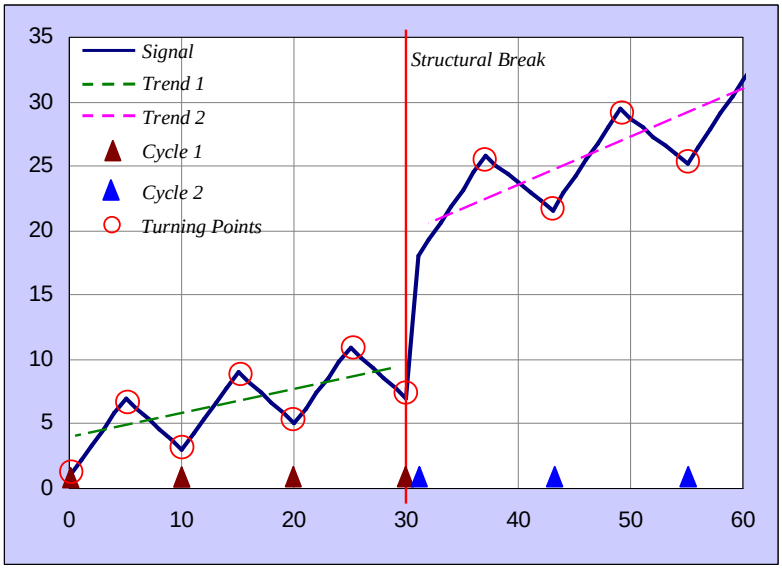


Figure 3.15 Synthetic time series with cycles, trends, turning points and a structural break.

The original signal is the solid dark blue line. The turning points are marked with red circles, whereas the structural break is indicated with a red vertical line. We note that this synthetic time series comprises two trends and two cycles. The slope of the trend (dashed green line) before the structural break is lower than the slope of the trend (dashed cyan line) after the structural break. Similarly the time period of the cycle (dark red triangle) before the structural break is smaller than the time period of the cycle (blue triangle) after the structural break.

We have studied how the wavelet analysis can effectively analyze local and global features at various timescales in complex time series. We will discuss in subsequent sections how our wavelet framework can identify patterns like cycles, trends, turning points and structural breaks in complex time series data to describe financial market dynamics at various timescales.

3.5.2 Risk Management

We have studied the CAPM and the multiscale CAPM in Sections 2.3.2 and 3.3.2 respectively. According to the CAPM, high *beta* (*risk*) stocks promise a high *return*. Traders often build portfolios of stocks according to the risk of each stock. For example, a trader might have 3 portfolios of 3 stocks each, that is a total of 9 stocks. He would want to keep a track of his returns from each portfolio. According to the CAPM, the portfolio with the highest average risk would promise the highest average return and the portfolio with the lowest average risk would promise the lowest return. A trader would therefore, want to study the risk-return tradeoffs for the

portfolios he has with an aim to get higher returns for lower risks. This would give him an indication as to which stocks he might keep in his portfolios and which he might discard.

Suppose a trader has a total of 9 stocks for which he has assessed the *betas* or risks ($\beta_1, \beta_2, .., \beta_9$). He can make 3 portfolios (P_1, P_2, and P_3) by putting 3 stocks in each portfolio. He can then monitor the return from each portfolio. Let us suppose further that $\beta_1 < \beta_2 < \ldots < \beta_9$. One way to make the portfolios is to put the first three stocks with the lowest *betas* in the first portfolio, the next three stocks with the next highest *betas* in the second portfolio and so on. This is shown in Figure 3.16 where 3 portfolios contain 3 stocks each in ascending order of their *betas*.

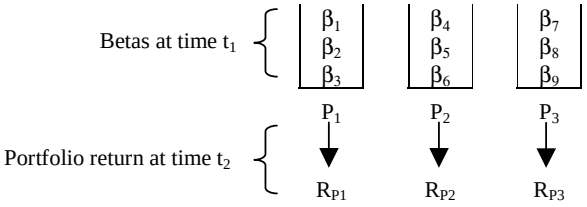


Figure 3.16 A method to study risk-return tradeoffs for portfolios of stocks.

According to the CAPM, since portfolio P_3 has the highest risk stocks, its return should be maximum, followed by P_2 and then P_1. For *betas* computed during time period t_1, we can in fact compute the average return of each portfolio for time period t_2 to test our hypothesis (Figure 3.16) – that is $R_{P3} > R_{P2} > R_{P1}$ at time t_2 for $\beta_{P1} < \beta_{P2} < \beta_{P3}$ at time t_1.

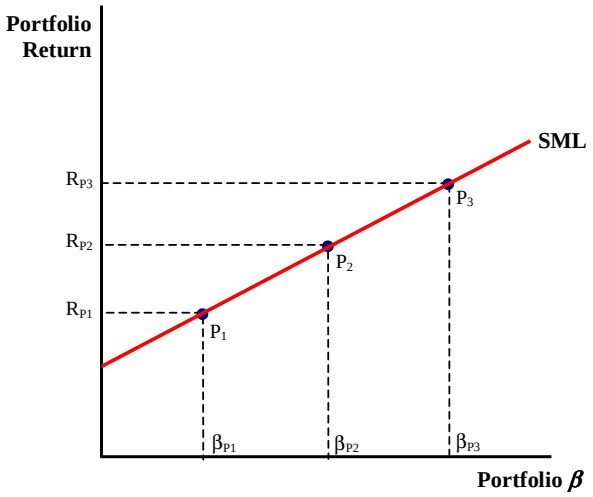


Figure 3.17 SML for studying risk-return tradeoffs of portfolios of stocks.

If for each portfolio, we plot the average return at time t_2 versus the average *beta* at time t_1, we should get a straight line, which is also called the *security market line* (SML) and was described in Figure 2.17. Figure 3.17 shows an exemplary SML for the portfolios described above. The *betas* can be computed using Eq (2.34) whereas the average returns can be computed

using $\frac{1}{N}\sum_{t=1}^{N} r_t$, where $r_t = log\ P_t - log\ P_{t-1}$ and N is the total number of samples in the time series of stock prices P_t. The slope and a straight line fit to the portfolio return versus portfolio *beta* plot would determine the validity and effectiveness of the CAPM (Figure 3.17).

A high slope of the SML and a perfect straight line fit would indicate an ideal risk-return tradeoff where higher returns can be obtained for lower risks. However, as discussed earlier, real world data is complex and exhibits scaling behaviour. In Section 3.3.2, with the help of wavelet analysis we have seen how *betas* or risks fluctuate over timescales (Table 3.2). Therefore, we would expect the slope and straight line fit of the SML of Figure 3.17 to vary over timescales. We can study the slope and straight line fit of the SML at different timescales using the wavelet analysis. For example, if we perform a level six wavelet decomposition on our data we would get six SML plots (similar to Figure 3.17) at six different timescales, each with a different slope and goodness of fit. In this manner we can study risk-return tradeoffs at various timescales, thus throwing more light on the dynamics of the CAPM. Such an analysis can help traders decide which stocks to keep in their portfolios and at which timescale to operate to maximize their profits or returns. We will discuss in subsequent sections how our wavelet framework can be used for efficient financial risk management.

3.5.3 Medical Diagnosis

Medical experts often visually examine numerical data recorded from the human organism to make a diagnosis. A typical example is a heart-rate tracing which is popularly known as an electrocardiogram (ECG). Medical experts extensively study ECGs to diagnose heart diseases such as congenital heart disease in infants and myocardial infarction and myocarditis in adults. Experts visually examine the peak-to-peak intervals, popularly known as the R-R intervals, in an ECG tracing to make a diagnosis. Is it possible for a machine to make a similar diagnosis using sophisticated time series analysis techniques?

The answer to the above question is yes – if we take into account the scaling behaviour of heartbeats in the complex human organism. Since the duration of each R-R interval is crucial for distinguishing between a normal and an abnormal heart, the first step towards analyzing an ECG tracing as a 'time' series is to convert it into a series of intervals between adjacent heartbeats τ_i. Such a series is called an R-R 'time' series. This is shown in Figure 3.18, where (a) represents a synthetic ECG tracing and (b) represents a synthetic R-R 'time' series.

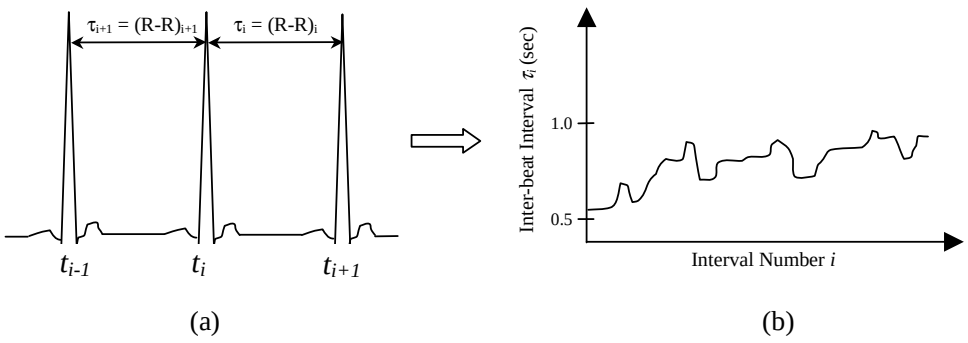


Figure 3.18: (a) A synthetic ECG tracing, (b) A synthetic R-R 'time' series.

We notice that the *x-axis* in Figure 3.18b is not *time* but, in fact, the *interval number* or *beat number*, that is, the ordering variable of the R-R 'time' series is the *beat number*. In cardiology, the R-R time series is thought to result from a complex superposition of multiple physiological processes at their respective characteristic timescales [8]. Empirical results show that the *log-log* plot of the wavelet variance versus timescale (discussed in Section 3.3.1) for a R-R recording from a normal patient produces a perfect straight line whereas the same plot for a congestive heart failure (CHF) patient does not produce such a straight line [152], [3]. Moreover, variability of a normal R-R recording is higher than the variability of a CHF R-R recording.

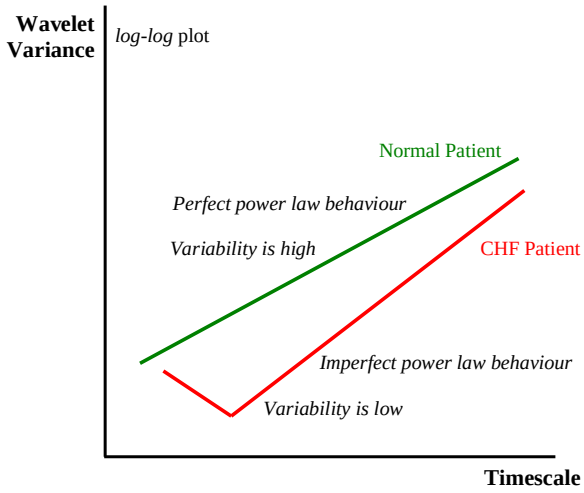


Figure 3.19 An exemplary wavelet variance versus timescale plot for a normal patient and a CHF patient for R-R recordings.

Figure 3.19 shows an exemplary plot of the wavelet variance versus timescale for a healthy patient and a CHF patient where a perfect power law behaviour (green line) suggests normalcy while an imperfect power law behaviour (red line) suggests congestive heart failure. That is the value of the power law exponent α remains constant across all timescales (frequencies) for normal patients and the value of the power law exponent α fluctuates across timescales (frequencies) for CHF patients when we examine the wavelet variance versus timescale plots for R-R recordings. In

Figure 3.19, we also observe that the overall variance (variability) of R-R recordings of normal patients is higher than that of CHF patients at all scales.

We will discuss in subsequent sections, how our wavelet framework uses the above empirical results to distinguish between CHF patients and normal patients for an unknown dataset of R-R recordings.

3.5.4 Time Series Summarization

In Sections 3.5.1, 3.5.2, and 3.5.3 we have formulated the problem of pattern identification, risk management and medical diagnosis respectively in complex time series. We can view the results or outputs of each of these analyses as a *summary* of the input data. In Section 2.2.3 we discussed time series summarization systems where the system generates an 'intelligent' *linguistic description* of the time series being analyzed. To us, summarization means to 'intelligently' *characterize* the dynamics and behaviour of complex time series based on well grounded mathematical techniques and analyses.

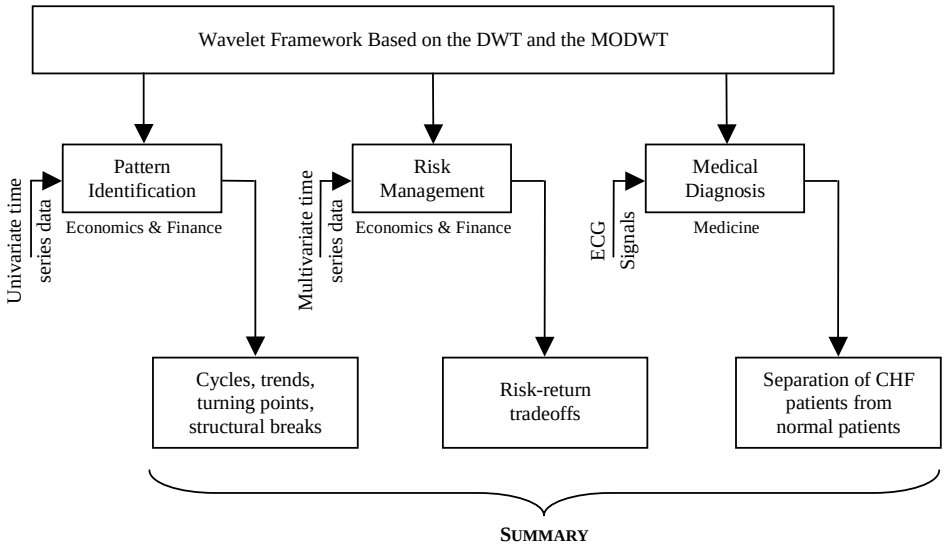


Figure 3.20 Wavelet framework for accomplishing time series summarization.

In Section 2.4 we presented a conceptual outline of our system. In Figure 3.20 we present a more detailed view of the architecture and functionality of our wavelet framework for analyzing complex time serial data. The output from our system could be graphical where identified patterns like cycles, trends, and turning points are 'marked up' on a time series or numerical where numbers 'characterize' the behaviour of the time series, for example, an equation describing the gradient of the market trend or the numeric value of the risk (*beta*) of a stock at various time horizons. Our framework, which is based on the wavelet analysis, has the robustness to deal with high-frequency and nonstationary data that may exhibit different dynamics at different timescales.

Our framework has both univariate and multivariate capabilities. The framework can analyze time series data from different domains, for example stock prices from economics, and ECG signals from medicine. The core of the framework is a DWT and MODWT engine. The system has three main modules namely, pattern identification, risk management, and medical diagnosis. The input to the pattern identification module is univariate time series data like stock prices, composite indices, and foreign exchange rates. The pattern identification module uses a combination of the DWT and the FFT to generate the *Surrey Market Report* where patterns like cycles, trends, turning points, and structural breaks are 'marked up' on financial time serial data and equations and numbers characterize the behaviour of the market at different time horizons. The input to the risk management module comprises multivariate data, for example the concurrently recorded time series for the DJIA and the share prices of the 30 companies within the DJIA. The risk management module builds portfolios of stocks to generate risk-return tradeoffs at different time horizons by making use of the MODWT-based analysis of variance and covariance. The risk-return tradeoffs are presented graphically where different slopes and fits of the SML are shown for different time horizons and numerically where equations characterize the risk-return tradeoffs at different time horizons. Finally, the input to the medical diagnosis module consists of ECG signals of normal and CHF patients. The medical diagnosis module separates out healthy patients from CHF patients by employing a power law analysis based on the wavelet variance. A statistic to measure the fluctuation of α over timescales (time horizons) is computed, which is then numerically and graphically presented to distinguish CHF patients from normal patients.

All the outputs from our framework comprise a graphical or numerical *summary* of the input data. Hence our framework can be regarded as a *multidisciplinary time series summarization system* where the summary output 'intelligently' characterizes the state and behaviour of the complex system under study. A detailed description of our framework including architectures, algorithms, case studies and evaluation will be presented in Chapter 4.

4 System Description

In this chapter, we outline a wavelet-based method that can be used for analyzing complex time series in economics and medicine by describing various modules of a *multidisciplinary time series summarization system*. The results of our analyses would be ready for use by both theoreticians and practitioners in economics and medicine.

Firstly, we re-iterate our motivation to build a wavelet framework for analyzing complex time series across different domains. A key message from Chapter 2 was the phenomenon of universal scaling in complex systems and in time series produced by complex systems. In Chapter 3, we established that wavelet analysis is a universal tool that can systematically guide and simplify the study of the phenomenon of scaling in complex time series. For a system designer, both universal principles (phenomena) and tools are important to develop robust systems that can render themselves useful across many disciplines. We chose to build a wavelet-based system because we believe that scaling is one of the universal principles in complex time series data and wavelet analysis is one of the universal tools that can help us study this universal principle of scaling. Since wavelet analysis is a nonparametric formal (systematic) method for analyzing complex time series data, it is a strong candidate for readily building a robust autonomous system that can work well across many domains.

We begin this chapter by providing details of the architecture of our wavelet framework for complex time series analysis (Section 4.1). Specific examples of real-world data from economics and medicine are considered to explain the full functionality of our framework. This is followed by Section 4.2 where we provide detailed descriptions of the algorithms we use to accomplish our analyses. In Section 4.3 we undertake three comprehensive case studies (two in economics and one in medicine) and show the various outputs and results produced by our system. In Section 4.4, we provide details of how our system was evaluated. Finally, Section 4.5 concludes this chapter with comments and discussion on the work reported in the last three chapters.

4.1 Architecture

In Section 3.5.4, we provided a general architecture of our wavelet framework for time series analysis which comprises three modules, namely, pattern identification, risk management and medical diagnosis. We discussed how the outputs from each of these modules can be regarded as a *summary* of the input data. In Figure 4.1, we again present the general architecture of our wavelet-based time series analysis system. The core of our wavelet framework is a DWT and a

MODWT engine. Our system has been implemented in Matlab®: the DWT analysis is accomplished using the Matlab Wavelet Toolbox [103], while the MODWT analysis is accomplished using the WMTSA (Wavelet Methods for Time Series Analysis) Wavelet Toolkit for Matlab [103], [118].

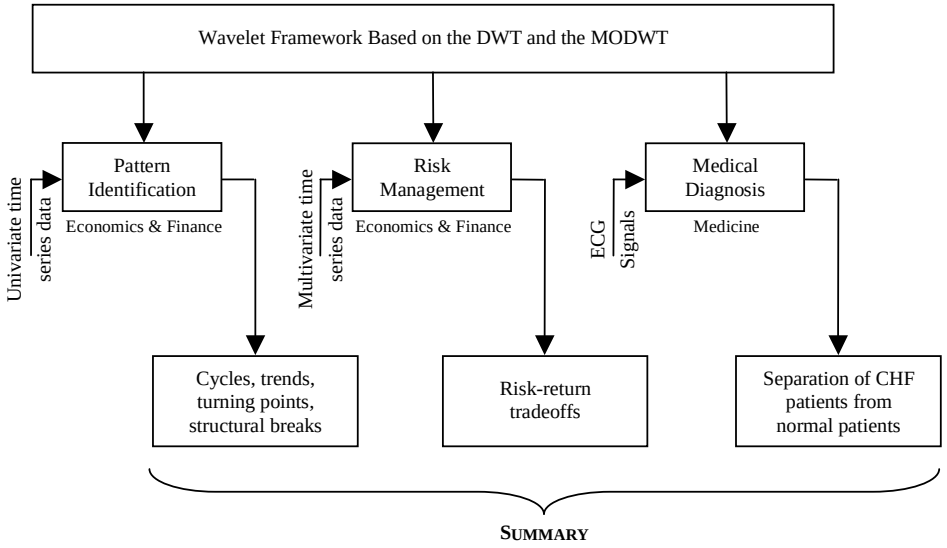


Figure 4.1 Wavelet framework for accomplishing time series summarization.

As discussed in Section 3.5.4 our wavelet framework has both univariate and multivariate time series analysis capabilities and can analyze signals in economics and medicine. The pattern identification and the risk management modules analyze univariate and multivariate financial time series data respectively while the medical diagnosis module analyzes ECG signals. More specifically, the pattern identification module identifies temporal patterns like trends, cycles and turning points in univariate financial time series data; the risk management module generates risk-return tradeoffs at various timescales by analyzing multivariate financial time series data; and the medical diagnosis module separates out congestive heart failure (CHF) patients from normal patients by analyzing ECG signals. In this section we discuss in more detail the functionality of each of the three modules with specific examples.

4.1.1 Pattern Identification

The pattern identification module uses a combination of the DWT and the FFT to identify temporal patterns like cycles, trends, turning points and structural breaks in univariate time serial data. The main output of the pattern identification module is the *Surrey Market Report*, which is an HTML page, displaying the results of the analysis performed by the module. We will describe the *Surrey Market Report* in more detail in Section 4.3.1.

A block diagram detailing the functionality of the pattern identification module is presented in Figure 4.2. The first component performs the *highest-level* DWT MRA on the original signal. The level of the DWT analysis depends on the length of the signal. Typically, for 300 samples, a level-8 DWT analysis is appropriate. Results of the DWT component are fed into two components - one for performing Fast Fourier Transform (FFT) and the other for detecting trends and turning points in the data. The second component computes the volatility of the original signal, which is again fed to a DWT component and then to the Normalized Cumulative Sum of Squares (NCSS) component. The results of the Fourier analysis and NCSS components are fed to two components for detection of inherent cycles and variance change respectively. Finally, numerical values outlining the trend, turning points, inherent cycles and variance change are fed to the numerical characterization component. This component performs tasks such as fitting a *trendline* to the wavelet smooth by ordinary least square (OLS) method and thereafter producing the *Surrey Market Report* that characterizes the market dynamics at various timescales.

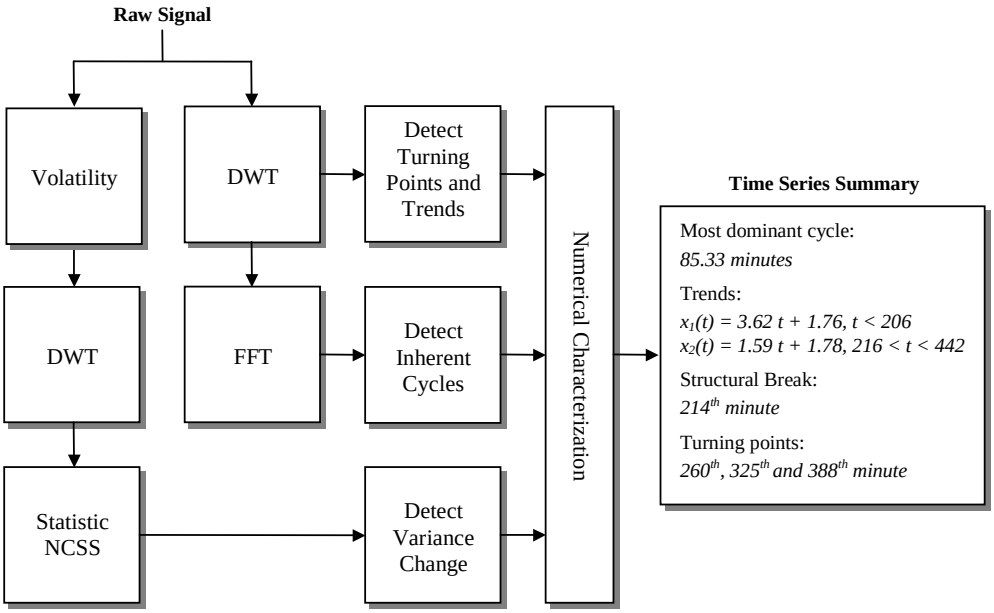


Figure 4.2 Block diagram of the pattern identification module.

We recall that the DWT MRA represents a signal as a sum of approximations or smooths (s_{J0}) and details (d_j) that are localized in time and frequency (Eq (3.35)). These individual approximations and details are obtained by recursively convolving the signal with a bank of low and high pass filters according to Mallat's pyramidal algorithm (Figure 3.8). The pattern identification module employs Daubechies six-coefficient wavelet filters, D(6), to perform a DWT MRA.

Cycle Detection

A *level-7* ($J_0 = 7$) DWT MRA using D(6) filters and a corresponding FFT performed on the daily IBM stock prices from May 17, 1961 to November 2, 1962 is shown in Figure 4.3. The original series is plotted in the top row on the left while below it from top to bottom are the wavelet approximation (smooth) and details (s_7, d_8, …, d_1). The right panel shows the corresponding Fourier power spectra of the signals on the left: the original series (IBM), the wavelet approximation (s_7) and wavelet details (d_1 to d_8).

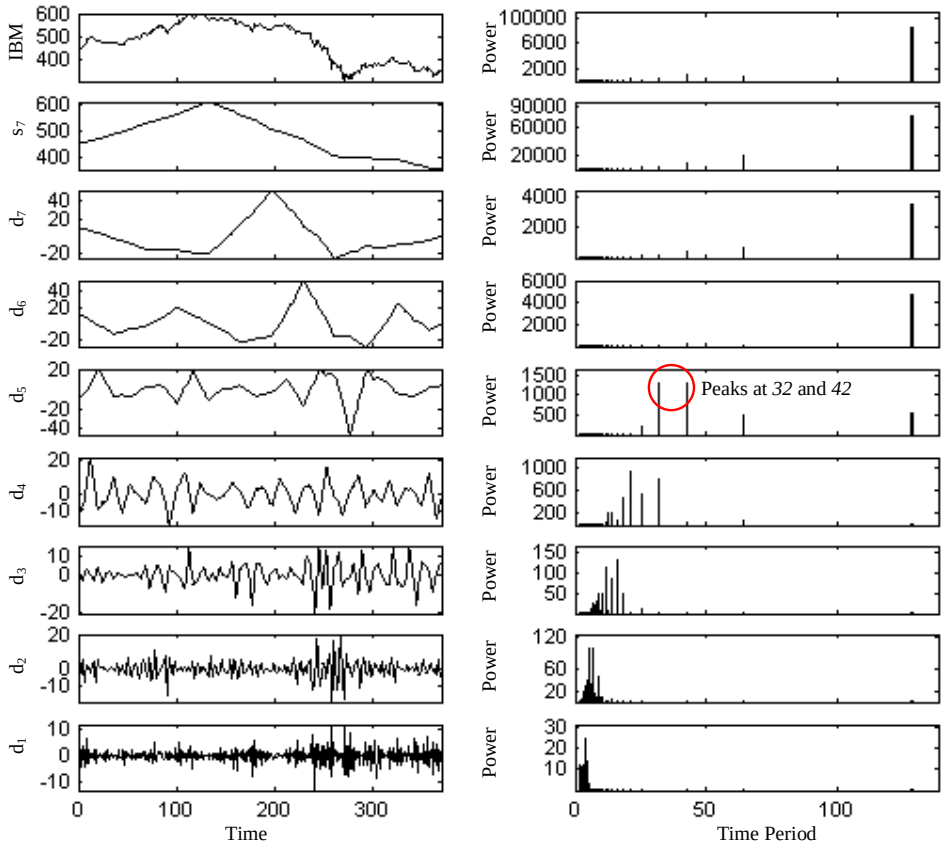


Figure 4.3 DWT MRA using D(6) filters and the FFT of daily IBM stock prices.

By performing an FFT (Fast Fourier Transform) analysis on each of the wavelet decompositions of a signal, we are able to extract both time and frequency information. The FFT power spectrum peaks indicate the frequencies (cycles) while the 'outstanding' peaks and troughs of the wavelet decompositions (extrema) indicate where in time these frequencies (cycles) are located in the original signal. In Figure 4.3, if we ignore the Fourier power spectra for the original signal (IBM), s_7, d_7 and d_6, which seem to be relatively featureless except large peaks at 128 (which is the total number of points for which the FFT was computed), we observe that the detail d_5 represents the most dominant cycle in the IBM series since its Fourier power spectrum has the highest power as compared to the spectra of d_4, d_3, d_2, and d_1.

In our cycle detection algorithm, we ignore the FFT peaks at 128 for $d_5 - d_7$, s_7 and IBM since these peaks arise due to the very low frequency characteristics of these signals and cannot be considered cycles as such. For example, we get a big peak at 128 for s_7 which is essentially a pair of straight lines with zero cyclicality. If we ignore these spurious peaks then what is left are the meaningful frequencies in the FFT spectra of each signal. We also note that we do not get spurious peaks at 128 for higher frequency periodic signals like $d_1 - d_4$. This implies the FFT works more efficiently for higher frequency periodic functions.

The pattern identification module compares the peaks of the Fourier power spectra of the wavelet details, and picks up the detail with the highest power as the dominant cycle in the signal. In this particular example, the wavelet detail d_5 will be picked up as the dominant cycle (since it has the highest power in its Fourier power spectrum), which according to the relation, 2^{j-1}, corresponds to changes on a scale of $2^{5-1} = 16$ days. Hence we can say that the dominant cycle in the IBM prices is in the range of *16* to *32* days and is characterized by the wavelet detail d_5. The Fourier power spectrum for d_5 also shows large peaks at *32* and *42* respectively, which have been marked by a red circle in Figure 4.3.

It is interesting to note that the second FFT peak of 42 lies outside the 16-32 days interval characterized by the wavelet detail d_5. This again shows that the FFT is very sensitive to the stationarity and periodicity of the signal being analyzed. Wavelet detail d_5 is periodic and stationary for most part but shows a significantly higher negative trough of about -40 in amplitude towards the end. It is possible that this may have caused a peak at 42 (which is outside the 16-32 days DWT d_5 interval) to appear in the FFT power spectrum for d_5.

Trend Detection

Traditional smoothing techniques like linear segmentation and moving average might not give the most accurate picture about the trend. The so-called outliers, noise and transients in the data might induce errors in the emergence of a trend through such techniques. The wavelet analysis on the other hand systematically removes noise, cyclical components and fluctuations in each level of decomposition, thus giving the purest ('de-noised') representation of a signal x_t at the *highest-level* of decomposition. If we recall Eq (3.35), for a level J_0 wavelet decomposition, a time series x_t, may be represented as,

$$x_t = s_{J0} + d_1 + d_2 + \ldots + d_{J0}. \qquad 4.1$$

Eq (4.1) shows how a signal x_t can be decomposed using the DWT MRA where d_1 is the highest frequency wavelet decomposition while d_{J0} is the lowest frequency wavelet decomposition of the original signal. Manipulating Eq (4.1) gives us,

$$s_{J0} = x_t - (d_1 + d_2 + \ldots + d_{J0}). \qquad 4.2$$

The wavelet details d_j can be looked upon as noise, cyclical components and other fluctuations of x_t and thus removing (subtracting) them from x_t would give a de-noised (smooth) signal (s_{J0}). As discussed in Chapter 3, s_{J0} is called the highest-level wavelet approximation or smooth while d_1, d_2, …, d_{J0} are called the wavelet details of the original signal x_t. The highest level wavelet smooth s_{J0} conforms well to the notion of a long-term trend in the signal since it represents averages of a scale of $2\lambda_{J0} = 2^{J0}$. In Figure 4.3, s_7 represents the long-term trend in the IBM time series x_t.

According to Mallat's pyramidal algorithm (Figure 3.8), we may decompose a given signal x_t into as many wavelet decompositions as we want. However, a point to note here is that the level of wavelet decompositions (J_0) required for an appropriate analysis, especially of the trend in the signal, depends on the length (N) of the original signal x_t. In our analysis, we use the following formula to compute the number of levels of decompositions J_0,

$$J_0 = floor\ [log\ (N) / log\ (2)] - 1 \qquad 4.3$$

The *floor* operation in Eq (4.3) ensures that the number J_0 is always an integer. In case the quantity [log (N) / log (2)] is a decimal number, the *floor* operation picks up the integer number to the left of the decimal point as J_0. The IBM series of Figure 4.3 has 369 samples, hence we arrive at a value of $J_0 = 7$ via Eq (4.3).

Turning Point Detection

Marr, in his research in edge detection in computer vision, showed that the Gaussian operator and its derivatives are a good model for the human visual system's detection of edges [102]. System TREND uses the cubic spline mother wavelet (which is a close approximation to the first derivative of the Gaussian) to detect edges in a signal [19]. Wavelet analysis is a multiresolution signal processing technique that adaptively smoothes a time series. The time-scale properties of the DWT MRA can be used to extract true features in a noisy signal by detecting local maxima of wavelet detail coefficients [56], [98]. We use the DWT's multiscale edge detection properties to detect turning points in a financial time series. We observe that the highest and lowest peaks (extrema) of wavelet details d_1 to d_{J0} successfully detect most of the turning points. The underlying hypothesis behind this approach is that the market is highly 'volatile' somewhere in the neighbourhood of these 'turning points'. This high 'volatility' or sharp change is picked up well by the wavelet decompositions d_1 to d_{J0}. Moreover, most turning points coincide with major edges in the signal.

Locating a Variance Change

To locate a variance change or a structural break in a given time series x_t, we adopt the methodology discussed in Section 3.4. We use the NCSS statistic described by Eqs (3.68 to 3.70) to locate a single change-point k_c in the variance (volatility) v_t, of a given time series x_t. Volatility

may be defined as the relative rate at which the price of a security moves up and down. If the price of a stock moves up and down rapidly over short time periods, it has high volatility. Conversely, if the price almost never changes, it has low volatility. Mathematically, volatility may be defined as a measure of the absolute value of the *first difference* of a time series. First, a return series is computed via the *first difference* of the log-transformed prices,

$$r_t = log(x_t) - log(x_{t-1}) \qquad 4.4$$

Then, the volatility is measured by the absolute return,

$$v_t = |r_t| \qquad 4.5$$

The NCSS computations to locate a variance change (Section 3.4) are carried out on the volatility series v_t described by Eq (4.5).

Numerical Characterization

The numerical characterization component performs tasks such as determining the slopes and intercepts of the trends represented by the wavelet smooth s_{J0} and filling numerical values outlining the trends, cycles, turning points and variance changes into an HTML template, to generate the *Surrey Market Report*. This component also generates a *headline* for the *Surrey Market Report* based on the trend information extracted from the wavelet smooth s_{J0}. As an example, let us consider how the numerical characterization module extracts information about the trend from the wavelet smooth s_{J0}. Figure 4.4a shows the wavelet smooth s_7 (blue plot) for the DWT MRA of Figure 4.3.

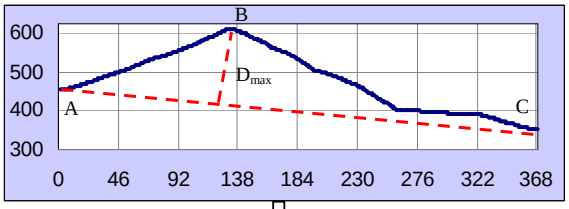


Figure 4.4a Procedure to locate an inflexion point in the wavelet smooth s_7.

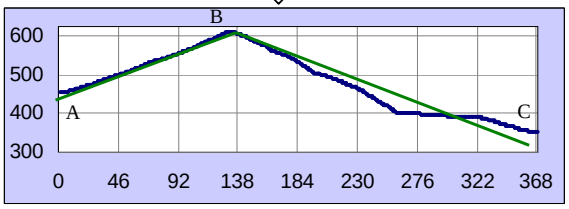


Figure 4.4b Two trends detected in the wavelet smooth s_7.

The numerical characterization module is programmed to look for either one or two trends in the wavelet smooth s_{J0}. This is done by drawing a straight line between the endpoints of s_{J0}, and computing the maximum distance D_{max} between the drawn straight line and s_{J0}. This is shown in Figure 4.4a. An inflexion point (B) in s_{J0} is assumed where D_{max} occurs. Two slopes ($Slope_{AB}$ and $Slope_{BC}$) are calculated using an ordinary least square (OLS) estimate for the two segments (*AB*

and *BC*) of s_{J0} where an inflexion point is assumed. If the ratio of the two slopes is greater than a predefined *tolerance*, then *B* is confirmed to be an inflexion point, otherwise, it is established that s_{J0} has only one trend, namely *AC*. That is,

if ($Slope_{AB}$ / $Slope_{BC}$) > *tolerance*,
then the two trends are AB and BC
else there is one trend AC
end;

In this particular example, there are indeed two trends in s_{J0}, namely *AB* and *BC*. The numerical characterization component computes the trends through the OLS method and presents this information in the form of an equation of a straight line of the form $y = mx + c$, where m is the slope and c is the *y-intercept*. In this example, the information about the trends can be summarized as,

$$x_1(t) = 1.21\,t + 445.56,\ t < 138$$
$$x_2(t) = -1.08\,t + 576.62,\ 138 < t < 369 \qquad 4.6$$

For Eqs (4.6) characterizing the trends in s_{J0}, the numerical characterization module will produce the following headline: *The market falls after initial rise*. Initial work on our pattern identification module has been published in *Lecture Notes in Computer Science* (LNCS) [1].

4.1.2 Risk Management

The risk management module builds portfolios of stocks to generate risk-return tradeoffs at different time horizons by making use of the MODWT-based analysis of variance and covariance. A block diagram of the risk management module is shown in Figure 4.5. The input to the risk management module comprises multivariate time series data, namely, a composite market index like the DJIA or the S&P 500, the ten year Treasury note and the companies within the composite index. The first two components of the risk management module are called Multiscale CAPM and CAPM respectively. As the names suggest, the Multiscale CAPM component helps in the generation of risk-return tradeoffs at different time horizons, whereas the CAPM component helps in the generation of risk-return tradeoffs at just one time horizon (that is the horizon at which the input data was recorded).

Let us discuss the functionality of the Multiscale CAPM branch first. The Multiscale CAPM component computes wavelet-based *betas* at time t_1 for each company by making use of the methodology described in Section 3.3.2. The Build Portfolio component builds portfolios of stocks at each timescale, according to the total number of companies being analyzed and how many companies does the user want per portfolio. For example, if there are a total of *30* companies that are being analyzed, the Build Portfolio component can build *10* portfolios with *3* companies each or *5* portfolios with *6* companies each at each timescale.

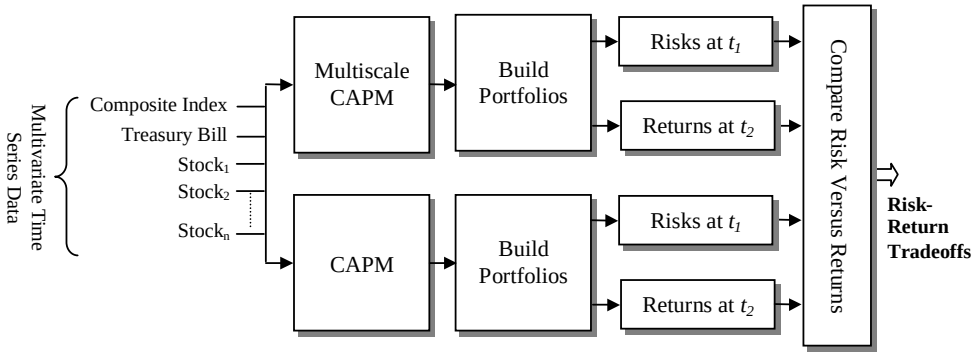


Figure 4.5 Block diagram of the risk management module.

The portfolios are labelled P_1, P_2, … and so on at each timescale and stocks are arranged in these portfolios in increasing order of their *betas*. The Risks at t_1 component computes the average portfolio risk at time t_1 while the Returns at t_2 component computes the average portfolio return at time t_2. The Compare Risks Versus Returns component plots the average return for each portfolio versus the average *beta* for each portfolio at different timescales and fits a straight line to each plot by an OLS method to generate risk-return tradeoffs. A similar procedure is repeated for the CAPM branch, where the only difference is that portfolios are built at just one timescale (the sampling rate at which data was recorded) and hence it generates risk-return tradeoffs at a single timescale.

We will now consider an example to elaborate the functionality of the risk management module. Our dataset comprises multivariate returns for the DJIA, the 30 companies within it, and the 10-Year Treasury Bill from January 3, 2000 to December 31, 2004. All time series were collected from Yahoo Finance ® and were converted into returns using the formula $r_t = log\ x_t - log\ x_{t-1}$. The sample size of each return series is 1245. Using our risk management module, we want to study the risk-return tradeoffs at different wavelet scales (λ_1, λ_2, …, λ_{J0}) by building ten portfolios which have three companies each. Further, out of a total sample size of 1245, we want to compute multiscale risks for the first 741 samples and want to study the returns for the remaining 541 samples. That is, $t_1 = 741$ and $t_2 = 541$. A point to note here is that the user specifies the number of portfolios he wants to build, the number of stocks per portfolio he wants and the time period t_1, and t_2. In this example, the total number of companies is 30, so three companies per portfolio will generate a total of ten portfolios. For simplicity, we will only show how the first three portfolios are built for the first two wavelet scales namely λ_1 and λ_2. The remaining seven portfolios at higher wavelet scales are built in the same manner.

Figure 4.6 shows an excerpt of our dataset where three return series, namely the DJIA, Coca Cola and the Treasury Bill are shown. The vertical red line shows the intervals for which we intend to study risk-return tradeoffs at different timescales: we want to compute risks for time period t_1 and

study the returns for time period t_2. The first step involves the computation of company *betas* for the first two scales λ_1 and λ_2 using the methodology outlined in Section 3.3.2. We recall that before computing multiscale risks using the Eq (3.67) we must subtract the *riskless* rate of return (i.e. the Treasury Bill) from each stock (company) to get the *stock premium* ($r_{it} - r_{0t}$) and we must also subtract the *riskless* rate of return (i.e. the Treasury Bill) from the market return (i.e. DJIA) to get the *market premium* ($r_{mt} - r_{0t}$). Now using Eq (3.67), we compute risks at time t_1, for each company at scale λ_1 and arrange them in increasing order in portfolios as per Table 4.1. Similarly for the next scale, we compute risks at time t_1 for each company at scale λ_2 and arrange them in increasing order in portfolios as per Table 4.1.

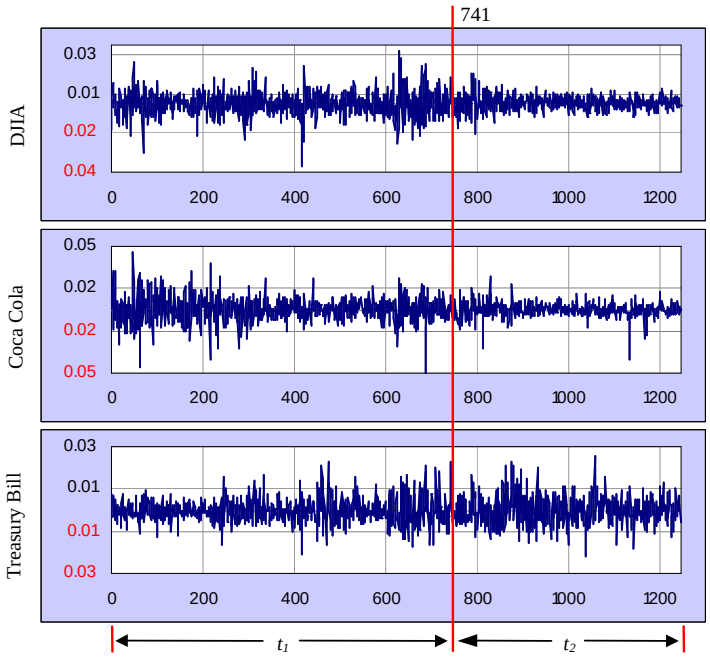


Figure 4.6 Returns for DJIA, Coca Cola and the Treasury Bill.

From Table 4.1, we observe that since we always arrange companies in portfolios in increasing order of their *betas*, the contents of the portfolios change at each timescale owing to different values of *betas* at different timescales. The average return R at time t_2 for each stock is computed using the formula,

$$R = 1000 * \frac{1}{N_{t_2}} \sum_{t=t_1}^{t_2} r_t \ , \qquad 4.7$$

where $r_t = log\ x_t - log\ x_{t-1}$ and N_{t2} is the total number of samples in the return time series of stock prices x_t in the interval t_2.

The value of R for a stock is put alongside that stock in each portfolio as shown in Table 4.1. Hence, in each portfolio, we have multiscale *betas* at time t_1 and the corresponding returns at time

t_2. We are now ready to compute the average *beta* $\overline{\beta}$, at time t_1, and the average return $\overline{R}$, at time t_2 for the portfolios to study risk-return tradeoffs at various timescales. The values for average portfolio *betas* at time t_1, and the average portfolio returns at time t_2, are shown in Table 4.1 for three portfolios P_1, P_2, and P_3, for the first two wavelet scales λ_1 and λ_2.

Table 4.1 Procedure for building portfolios by arranging companies in each portfolio in increasing order of their *betas* at each timescale.

Portfolio	Scale 1 (λ_1)					Scale 2 (λ_2)				
	Company	β	R	$\overline{\beta}$	$\overline{R}$	Company	β	R	$\overline{\beta}$	$\overline{R}$
P_1	Pfizer Inc	0.55	4.87	0.65	4.39	Pfizer Inc	0.69	4.87	0.74	5.09
	Johnson and Johns DC	0.67	4.46			McDonalds CP	0.76	6.55		
	Exxon Mobil CP	0.73	3.84			Exxon Mobil CP	0.77	3.84		
P_2	Walt Disney-Disney C	0.76	6.69	0.77	6.53	Altria Group Inc	0.78	5.63	0.80	5.61
	SBC Communications	0.77	6.80			Merck Co Inc	0.79	5.09		
	Verizon Commun	0.78	6.09			Verizon Commun	0.82	6.09		
P_3	Coca Cola Co The	0.79	4.15	0.80	5.44	Coca Cola Co The	0.84	4.15	0.87	3.86
	McDonalds CP	0.80	6.55			Johnson and Johns DC	0.87	4.46		
	Altria Group Inc	0.81	5.63			Procter Gamble Co	0.89	2.98		

We can in fact plot the average portfolio returns versus the average portfolio *betas* at each timescale and fit a straight line to our plots using an OLS estimate. This straight line will correspond to the *security market line* (SML) described in Sections 2.3.2 and 3.5.2, and its slope and intercept will be a measure of the efficiency of the CAPM and hence of the risk-return tradeoff.

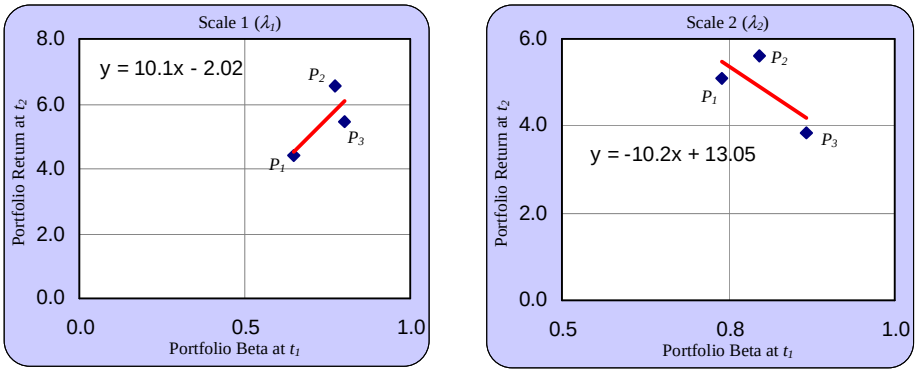


Figure 4.7 Risk-return tradeoffs at scales one and two for portfolios constructed in Table 4.1. The red straight line fit is the *security market line* (SML).

Figure 4.7 shows the risk-return tradeoffs at wavelet scales λ_1 and λ_2 for the three portfolios built in Table 4.1. We observe that the predictions of the CAPM are more meaningful at scale λ_1 than at scale λ_2 for this particular example although we note that we need to plot more portfolio values

before we can actually fit a straight line through them with a certain degree of confidence – this example is just for illustrating our methodology and a full study is reported in Section 4.3.2. For scale λ_1 a high slope and a good fit of the SML is observed while for scale λ_2 the slope of the SML is negative and does not conform to the predictions of the CAPM. From this analysis, we can conclude that the time horizon of $\lambda_1 = 2$ *to* 4 days, has been most profitable for companies in portfolios P_1, P_2, and P_3 since high returns were obtained for low risks (Table 4.1), while maintaining the validity of the CAPM at this time horizon.

We can regard the output of the risk management module as a kind of *summary* that characterizes the risk-return tradeoffs at various time horizons for multivariate time series data. In Section 4.3.2, we will study the same data in greater detail where risk-return tradeoffs will be generated for six wavelet scales for all the ten portfolios. We will also show how we can change the number of companies per portfolio and the time periods t_1 and t_2 to study risk-return tradeoffs at various timescales in multivariate data.

4.1.3 Medical Diagnosis

The medical diagnosis module separates out congestive heart failure (CHF) patients from normal patients by carrying out a power law analysis based on the wavelet variance. As discussed in Section 3.5.3, empirical results show that the scaling exponent α, for R-R recordings of normal patients is scale invariant or in other words is stable across timescales. On the other hand, the scaling exponent α for R-R recordings of CHF patients is unstable across timescales [152]. The medical diagnosis module makes use of this empirical result to discriminate between normal and CHF patients by analyzing their ECGs.

A block diagram of the medical diagnosis module is presented in Figure 4.8. The input to the medical diagnosis module comprises ECG signals of normal and congestive heart failure patients. The ECG-RR Converter component converts each ECG recording into an R-R 'time' series, as detailed in Section 3.5.3. The Wavelet Variance Based SDF component estimates the spectral density levels in each R-R time series as a function of timescale λ_j using methodology described in Section 3.3 and Eq (3.58). As the name suggests, the Power Law Analysis component examines power law behaviour in the multiscale spectral density levels of each R-R time series by calculating the scaling exponent α, using methodology developed in Section 3.3.1. Finally, the Scaling Instability Index component computes the scaling instability index ΔS to measure the extent of the fluctuation of the scaling exponent α over timescales for each R-R time series. It is obvious that for CHF patients the value of ΔS will be consistently higher (since α fluctuates over timescales for CHF patients) as compared to the value of ΔS for normal patients (since α is stable

over timescales for CHF patients). The medical diagnosis module plots the value of ΔS for all ECGs examined as its output and hence separates out CHF patients from normal patients.

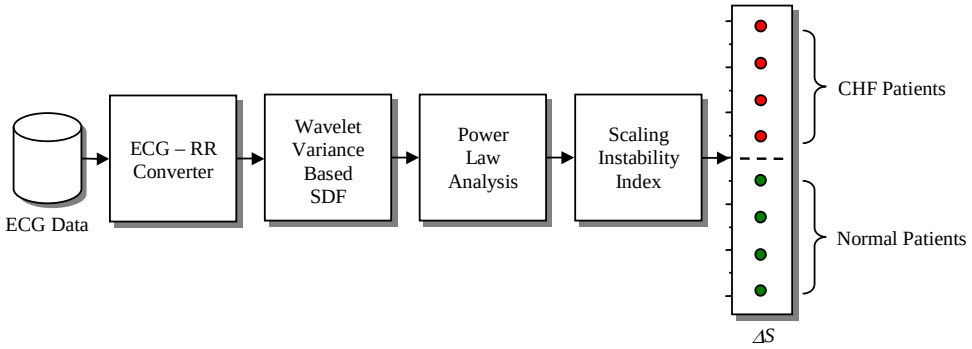


Figure 4.8 Block diagram of the medical diagnosis module.

We will now consider an example to elaborate the functionality of the medical diagnosis module. Figure 4.9 shows two R-R time series, one for a normal patient (green) and the other for a CHF patient (red). The normal R-R series has 100210 samples while the CHF series has 75542 samples. The green plot is for the normal patient while the red plot is for the CHF patient. By visually examining the two signals we can say that the normal R-R time series (green plot) has certain trends whereas the CHF R-R time series (red plot) appears to be more like a random process. In other words the red plot appears to posses more complexity (and variability), which is a characteristic of healthy complex systems, as compared to the red plot. Using methodology described in Section 3.3 and Eq (3.58), we compute the Haar wavelet variance based spectral density levels (SDF) as a function of timescale for the two R-R recordings. For both R-R recordings, we use a *level-10* MODWT decomposition ($J_0 = 10$) to estimate the wavelet variance. The result of this analysis is presented in Figure 4.10 which shows the SDF versus timescale plot on a *log-log* graph, for the two R-R recordings. As discussed in Section 3.3.1, the slope of the SDF versus timescale plot on a *log-log* graph will give us the value of the scaling exponent α.

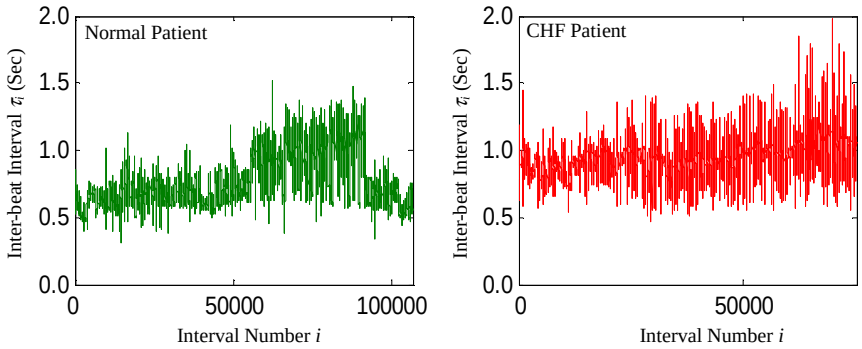


Figure 4.9 R-R time series for a normal patient (green) and a CHF patient (red).

To check for the stability of α over timescales (i.e. scaling stability) we examine its value over two regions, namely between scales 1 and 3, and between scales 3 and 10. For this purpose, we have fitted two separate straight lines between scales 1 and 3, and between scales 3 and 10 for each of the two SDF versus timescale plots using the ordinary least square (OLS) method (Figure 4.10). The two lines for the normal patient are A'B' and B'C' while the two lines for the CHF patient are AB and BC. By visually examining the plots in Figure 4.10, we can say that the difference in the slope of A'B' and B'C' is significantly lower than the difference in the slope of AB and BC. That is, the value of the scaling exponent α is stable over timescales for the normal patient whereas its value is unstable over timescales for the CHF patient. This leads us to define a scaling instability index ΔS given by,

$$\Delta S = \left|(\alpha_{1-3}) - (\alpha_{3-10})\right|, \qquad 4.8$$

where the suffix of α represents the scales for which it is computed.

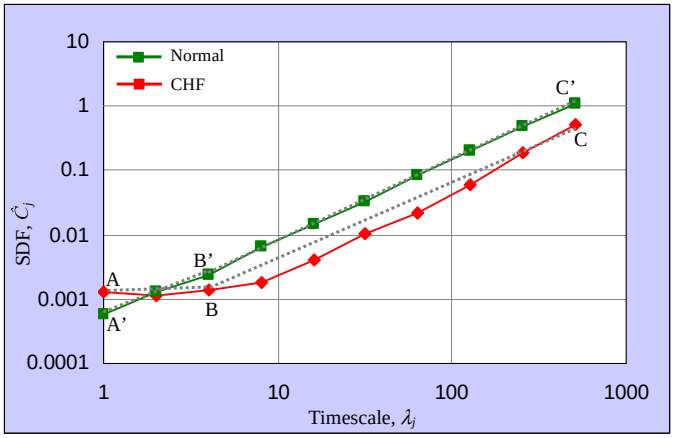


Figure 4.10 Wavelet variance based SDF versus timescale plot on a log-log graph for the normal patient (green) and the CHF patient (red) of Figure 4.9.

From Eq (4.8) and Figure 4.10, it is clear that the value of ΔS (scaling instability index) for CHF patients is higher than its value for normal patients, that is,

$$|(Slope_{AB}) - (Slope_{BC})| > |(Slope_{A'B'}) - (Slope_{B'C'})|. \qquad 4.9$$

In this particular example, the value of ΔS for the CHF patient is 1.23, and that for the normal patient is 0.22. In Section 4.3.3, we will analyze a dataset of 45 ECG signals where 22 ECGs are of normal patients and the other 23 are of CHF patients.

4.2 Algorithms

In the last section, we discussed in detail the architecture of our wavelet framework for time series analysis and summarization. In this section, we present three generalized algorithms, namely

pattern identification, risk management, and medical diagnosis to complete the description of our wavelet framework.

Figure 4.11 shows the first algorithm, which is called the *pattern identification algorithm*. Input to the pattern identification algorithm comprises financial *tick* data and the *compression* period for which the tick data is to be analyzed. The output of the pattern identification algorithm is the Surrey Market Report. With the advent of the Internet and online data vendors, more and more financial time series data is being recorded, supplied and stored online. In financial markets, traders both 'bid', price at which they are prepared to buy and 'ask', price at which they will sell. This system of bid / ask pricing ensures the sale / purchase of instruments without any delay. A day's trading (comprising almost 24 hours) for such data could generate 25,000 to 30,000 *ticks* per day per instrument.

Financial tick data is not a time series as such – ticks are reported as and when a transaction takes place, and hence values in time are not equally spaced. The pattern identification algorithm has the capability to *compress* tick data to a desired time period and render it as a time series of equally spaced intervals. The *compression* acts as a surrogate for the original: the maxima (High) and minima (Low) of the data over a fixed interval (typically, one minute).

Algorithm 1: Pattern Identification

Input : (1) Financial Tick Data
(2) Compression Period

Output : Surrey Market Report describing the market dynamics at different time horizons

I. Compress the tick data to get open (o_t), high (h_t), low (l_t), close (c_t), values for a given input compression period (for example one minute or five minutes).

II. Calculate the level J_0 of the DWT MRA needed, based on the number of samples N in c_t,
$$J_0 = floor\ [log\ (N) / log\ (2)]\ -1.$$

III. Perform a *level-J_0* DWT MRA on c_t, using Eqs (3.16, 3.17, 3.26, 3.27, 3.33, 3.34, & 3.35), to get the wavelet details and smooth,
$$d_j, \text{ and } s_{J0}, j = 1, \ldots, J_0.$$

III-1. Compute the **trend(s)** by checking for an inflexion point in s_{J0} and using linear regression.

III-2. Extract **cycle** (seasonality) by performing a Fourier power spectrum analysis (using Eqs (3.7 to 3.9)) on the wavelet details, d_j, and choosing the detail with the maximum power as the cyclical component, $d_{cyclical}$,
$$d_{cyclical} = \max\left(\sum_{t=0}^{N-1} \left|d_{j,t}\right|^2\right),\ j = 1, \ldots, J_0.$$

III-3. Extract **turning points** by determining the abscissa of the extremas of d_j,
$$turning\ points = abscissa\ [max\ (d_j), min\ (d_j)],\ j = 1, \ldots, J_0.$$

IV. Compute the volatility of c_t via,
$$v_t = |log\ (c_t) - log\ (c_{t-1})|.$$

IV-1. Locate a **single variance change** k_c in the volatility series v_t by using the NCSS statistic described by Eqs (3.68 to 3.70),
$$k_c = abscissa\ (D) = abscissa\ [max\ (D^+, D^-)].$$

V. Fill the **Surrey Market Report** HTML template with outputs of Steps III-1, III-2, III-3, IV, and IV-1 to generate a numerical and graphical summary.

Figure 4.11 Pattern Identification Algorithm.

The value at the start (Open) and at the end (Close) of the minute acts as the surrogate for other data during the minute. Data *compression* essentially yields four new time series: Open, High, Low, and Close data values. For instance, the Intraday trading of an exchange rate instrument GBP-USD (£/$) comprising over 25,000 *ticks* can be compressed into 1-minute slices resulting in 1440 data points per day. Once the data has been compressed, the pattern identification algorithm analyzes it with the DWT MRA and the FFT to generate the Surrey Market Report. The output from our pattern identification module can be regarded as a kind of *summary* that characterizes market dynamics at different time horizons.

Algorithm 2: Risk Management

Input : (1) One time series of the composite index, x_{mt}
(2) n time series of companies within composite index, x_{it}, $i = 1, \ldots, n$
(3) One time series of the treasury note, x_{0t}
(4) Period for which risks are to be computed, t_1
(5) Period for which returns are to be computed, t_2
(6) Number of portfolios to be made, k,
(7) Number of time horizons to be examined, J_0

Output : Risk-return tradeoffs at different time horizons (1 *to* J_0) and raw risk-return tradeoffs

I. Compute **returns** for all input time series,

$r_{it} = log\ (x_{it}) - log\ (x_{it-1})$, $i = 1, \ldots, n$. *// company returns*
$r_{mt} = log\ (x_{mt}) - log\ (x_{mt-1})$. *// market return*
$r_{0t} = log\ (x_{0t}) - log\ (x_{0t-1})$. *// risk-free rate of return*

I-1. Convert returns into **premium returns** for all company returns and market return,

$r_{it} = r_{it} - r_{0t}$, $i = 1, \ldots, n$. *// company premiums*
$r_{mt} = r_{mt} - r_{0t}$. *// market premium*

II. For time period t_1, compute **multiscale *betas*** (Eq (3.67)) for each company using results of Step I-1,

$$\beta_{ij}^{w} = \frac{Cov(\tilde{w}_{ij}, \tilde{w}_{mj})}{\sigma_{mj}^2},\ j = 1,..., J_0;\qquad i = 1,..., n$$

II-1. For time period t_1, compute **raw *betas*** (Eq (3.66)) for each company using results of Step I-1,

$$\beta_i = \frac{Cov(r_{it}, r_{mt})}{\sigma_m^2}, i = 1,...., n$$

III For time period t_2, compute **average returns** for all companies, using Eq (4.7),

$$R = 1000 * \frac{1}{N_{t_2}} \sum_{t=t_1}^{t_2} r_{it}$$

IV. **Build** k **portfolios** at each **timescale** j, and fill each portfolio with n/k companies in increasing order of their *betas* at time t_1 (output of Step II), along with their corresponding average returns at time t_2 (output of Step III),

$P_{c,j}$, $c = 1, \ldots, k$; $j = 1, \ldots, J_0$.

IV-1. **Build** k **portfolios** at the sampling frequency (for **raw** data), and fill each portfolio with n/k companies in increasing order of their *betas* at time t_1 (output of Step II-1), along with their corresponding average returns at time t_2 (output of Step III),

P_c, $c = 1, \ldots, k$.

V For each portfolio $P_{c,j}$, at each **timescale** j, compute the average **portfolio *beta*** $\beta_{Pc,j}$ and average **portfolio return** $R_{Pc,j}$,

$[\beta_{Pc,j}, R_{Pc,j}]$, $c = 1, \ldots, k$; $j = 1, \ldots, J_0$.

V-1 For each portfolio P_c, at the sampling frequency (for **raw** data), compute the average **portfolio *beta*** β_{Pc} and average **portfolio return** R_{Pc},

$[\beta_{Pc}, R_{Pc}]$, $c = 1, \ldots, k$.

VI. Plot average portfolio *betas* versus average portfolio returns and fit straight lines through OLS estimates at different scales (output of Step V) and at the sampling frequency (output of Step V-1) to study **risk-return tradeoffs** at different time horizons.

Figure 4.12 Risk Management Algorithm.

The second algorithm, called the *risk management algorithm*, is shown in Figure 4.12. This algorithm manages portfolios of stocks and generates risk-return tradeoffs at different time horizons. Inputs to the risk management algorithm comprise multivariate time series data, time periods for which risk-return tradeoffs are to be studied, number of portfolios to be built, and the number of time horizons to be examined. Based on the CAPM, the risk management algorithm computes multiscale *betas* (risks) in one time period for all companies inside a composite index and computes the returns in the next time period. The algorithm then builds portfolios of stocks at each wavelet timescale according to the value of *beta* for each company. Finally, at each timescale, average portfolio *betas* and average portfolio returns are computed and plotted.

A linear regression (OLS estimate) is performed to fit straight lines to the risk-return plots at each timescale. These straight-line fits represent the *security market line* (SML) of the CAPM (discussed in Section 2.3.2) and their slopes and intercepts give us a measure of the performance of our portfolios. We are able to determine, as to what companies gave higher returns for lower risks and at what timescales. The output from our risk management module can be regarded as a kind of *summary* that characterizes risk-return tradeoffs at different time horizons.

The third algorithm is called the *medical diagnosis algorithm*, and is shown in Figure 4.13. Inputs of the medical diagnosis algorithm are ECG signals of healthy (normal) and CHF patients.

Algorithm 3: Medical Diagnosis

Input : (1) ECG signals of CHF and normal patients
(2) Number of time horizons to be examined, J_0

Output : Separation / classification of CHF and normal patients

I. **Convert** all ECG signals into **R-R 'time' series**.

II. For each R-R time series, compute the **wavelet variance** based spectral density levels (**SDF**) as a function of timescale λ_j using Eq (3.54 & 3.58),

$$\hat{C}_j \equiv 2^j \hat{\sigma}_x^2(\lambda_j),\ j = 1,\ldots J_0$$

II-1. For each R-R time series, **plot *log-log*** representation of the SDF $\hat{C}_j$ as a function of timescale λ_j (output of Step II) to study **power law** behaviour.

II-2. For each SDF versus λ_j plot of Step II-1, determine the **scaling instability index** ΔS by computing slopes (or the value of scaling exponent α) in two regions, namely *1* to r_1 and r_1 to J_0,

$$\Delta S = \left|(\alpha_{1-r1}) - (\alpha_{r1-J0})\right|.$$

III Plot the value of the scaling instability index ΔS for each R-R recording to **separate** out **CHF** patients (who will have higher values of ΔS) from the **normal patients** (who will have lower values of ΔS).

Figure 4.13 Medical Diagnosis Algorithm.

The algorithm first converts all ECG signals into R-R time series and then examines power law behaviour in each one of them by computing the SDF as a function of timescale. Finally, the value of the scaling exponent α is computed in two frequency ranges to check for its stability over these frequencies. In other words, a scaling instability index is defined and computed. Based on empirical results, which suggest that the value of the scaling instability index is higher for CHF

patients, and lower for normal patients, the algorithm is able to classify (separate) the two categories of patients.

4.3 Case Studies

Now that we have discussed our system architecture and algorithms in detail, we are ready to perform case studies to explain the kind of outputs our wavelet framework produces. First, we will demonstrate the generation of the Surrey Market Report by the pattern identification module for a British Pound - US Dollar (GBP-USD) exchange rate time series. This will be followed by a study of risk-return tradeoffs in the DJIA using our risk management module. Finally, we will examine ECG signals of healthy and CHF patients with our medical diagnosis module.

4.3.1 Surrey Market Report

In this case study, we examine how our pattern identification module generates the Surrey Market Report to describe the market dynamics at various time horizons for a given time series. Our input data comprises *tick data* for the British Pound – US Dollar (GBP-USD) foreign exchange rate on May 6, 2004. There are approximately 25,000 ticks for the GBP-USD exchange rate data on this particular day. Our intention is to study this data on a minute-by-minute basis hence we perform a one-minute compression on it before analysis. A point to note here is that our pattern analysis module has data compression capability, and we can compress tick data into different compression periods, for example 5 minutes, 30 minutes, etc. Here, the compression operation results in an equally spaced time series of 1409 minutes or samples, in which we will study cycles, trends, turning points, and structural changes.

The front page of the Surrey Market Report for the GBP-USD *tick* data is shown in Figure 4.14. The pattern identification module generates a *headline* based on the characterization of the trend in terms of its slope and intercept. This headline is displayed on the front page of the Surrey Market Report. The front page also displays menus like *Trend Analysis* and *Cycle Analysis*, which the user can click to go to that particular section of the analysis. The full report for the GBP-USD exchange rate *tick* data generated by the pattern identification module can be viewed at [145].

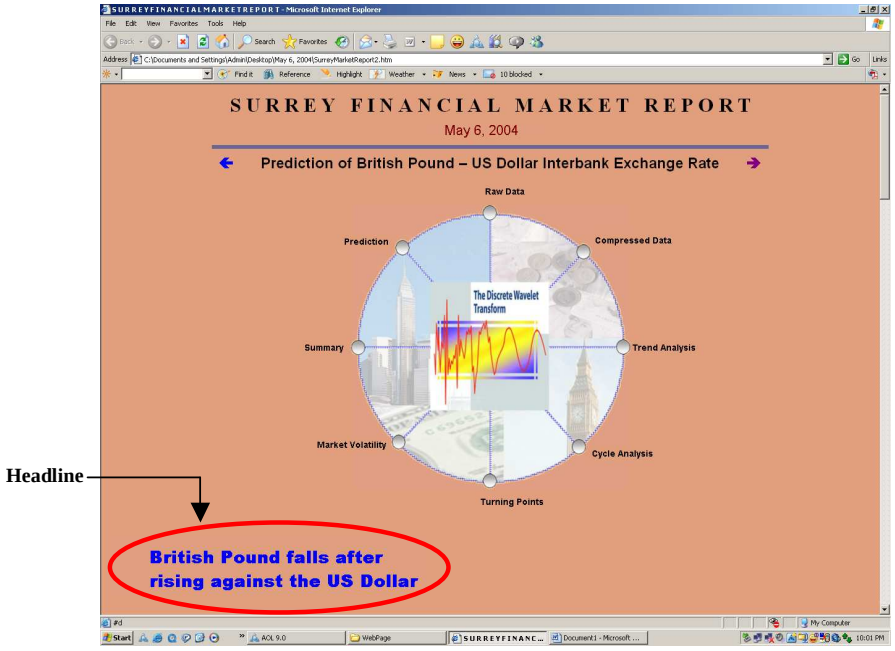


Figure 4.14 Surrey Market Report generated by the pattern identification module for the GBP-USD minute-by-minute exchange rate on May 6, 2004. The full report can be viewed at http://www.saifahmad.com/May6/SurreyMarketReport2.htm.

Figure 4.15 shows the graphical summary generated by the pattern identification module for the GBP-USD exchange rate of May 6, 2004. The original data is shown as a grey coloured plot. We observe that there are two trends in the data (blue plot). The first trend is an *uptrend* while the second trend is a *downtrend*. This is what is displayed in the headline of the Surrey Market Report (Figure 4.14), which says, "The British Pound falls after rising against the US dollar". The graphical summary of Figure 4.15 also shows the cycle identified in the data (green plot), the turning points (open red circles) and the variance break (dark red vertical line). In this particular example, the wavelet detail d_7 represents the cyclical component in the original signal and hence the cyclicality in the data is 64 - 128 minutes. We observe that the cyclical component d_7 (green plot) coincides with major fluctuations of the input time series (grey plot). If we look at the structural or variance change (dark red vertical line), we observe that the amplitude of fluctuations in the original signal (grey plot) has reduced after this point in time. In other words the volatility of the time series has reduced after the 854^{th} minute which represents a change in variance of the series.

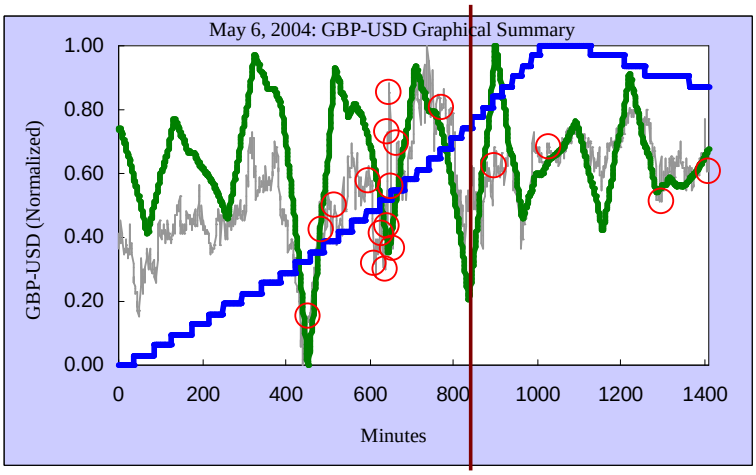


Figure 4.15 Graphical summary of the GBP-USD generated by the pattern identification module, showing the trend (blue plot), cycle (green plot), turning points (red circles), and variance change (dark red vertical line).

The numerical summary generated by the pattern identification module for the GBP-USD time series is shown in Table 4.2.

Table 4.2 Numerical summary of the GBP-USD generated by the pattern identification module.

Trend	*1st Phase*	$x_1^{Trend} = 2.85e-06t+1.79$, $t < 1028$
	2nd Phase	$x_2^{Trend} = -9.73e-08t+1.80$, $1028 < t < 1409$
Turning Points		452, 484, 516, 596, 612, 628, 636, 642, 644, 648, 653, 656, 666, 772, 900, 1028, 1297, and 1409.
Variance Change		854^{th} minute.
Cycle		64 - 128 minutes.

The trend has been characterized by two equations of straight lines which tell us about the gradients, intercepts and locations of the two trends detected. In addition, the numerical summary comprises the value of the period of cyclicality (64 - 128 minutes), and the location in time of the structural break (854^{th} minute) and turning points.

The graphical and numerical summaries produced by our pattern identification module are solely based on the *details* and *smooths* of the DWT MRA. In this case study, the trend is the wavelet smooth s_9, the cycle is the wavelet detail d_7 and the turning points are extremas of wavelet details d_1 to d_9. Hence our pattern identification module has been able to describe the market dynamics at various timescales for the given dataset.

In Section 4.4, we will undertake a human, as well as a statistical evaluation of our pattern analysis module to test the robustness with which it can identify trends, cycles, turning points, and variance breaks in time serial data. In the human evaluation section (Section 4.4.1) real-world time series will be given to experts who will mark features on these time series with hand. The same time series will then be analyzed by our pattern identification module. The results produced by our system will be compared with the experts' opinion with regards to various patterns in the time series examined. In the statistical evaluation section (Section 4.4.2) artificial time series will be created with predefined properties. Noise will then be added to the artificial time series to investigate what level of noise still allows for the identification of features like trends and variance breaks. Outputs produced by our system using different wavelet filters (*mother wavelets*) will be compared with other approaches like the well known moving average (MA) technique, and the null hypothesis testing technique.

4.3.2 Risk-Return Tradeoffs in the DJIA

In this case study we perform a multivariate multiscale analysis using our risk management module to study risk return tradeoffs in the DJIA. The input data comprises the daily values of the DJIA, the daily prices of the 30 companies within the DJIA, and the daily values of the *Treasury note* for the sample period of January 3, 2000 to December 31, 2004, a total of 1246 data samples.

We note here that since all prices will be converted into returns by the risk management module, the sample size will reduce to 1245 from 1246. We wish to make 10 portfolios for the 30 companies in the Dow and want to study the risk-return tradeoffs at six wavelet scales. In addition, we want to analyze risks for a period of 1000 samples ($t_1 = 1000$) while we want to examine the returns for the remaining 246 samples ($t_2 = 245$).

The output of the risk management module for the above dataset is presented in Figure 4.16, which shows the risk-return tradeoffs of 10 portfolios for the raw data and for wavelet scales 1 to 6. Since the number of portfolios is 10, each portfolio contains three companies each. Full details of the portfolios built by the risk management module for the raw data and the six wavelet timescales are shown in Tables 1a - 1g in *Appendix A.*

In Figure 4.16, we observe that different slopes and intercepts of the SML are obtained for raw data and for the various wavelet timescales. The optimum risk-return tradeoff is achieved at scale 6 (λ_6), since the slope of the SML is highest at this time horizon. If we look at Table 1g in *Appendix A* which shows the construction of portfolios for this particular time horizon (λ_6), we observe that risk values therein are consistently lower as compared to risk values for raw data (Table 1a, *Appendix A*) and for scales λ_1 to λ_5 (Tables 1b - 1f, *Appendix A*). This is exactly what an investor is looking for – *high returns for lower risks*. For the dataset and time periods

examined, our risk management module tells us that a trader operating at timescale 6 (64-128 days) is likely to make *maximum profit* for a *minimum risk* by building his portfolios as per Table 1g (*Appendix A*). The outputs of Figure 4.16 and Tables 1a to 1g (*Appendix A*) can be regarded as a *summary* of the multivariate input data that characterizes risk-return tradeoffs for raw data and for different time horizons in the data.

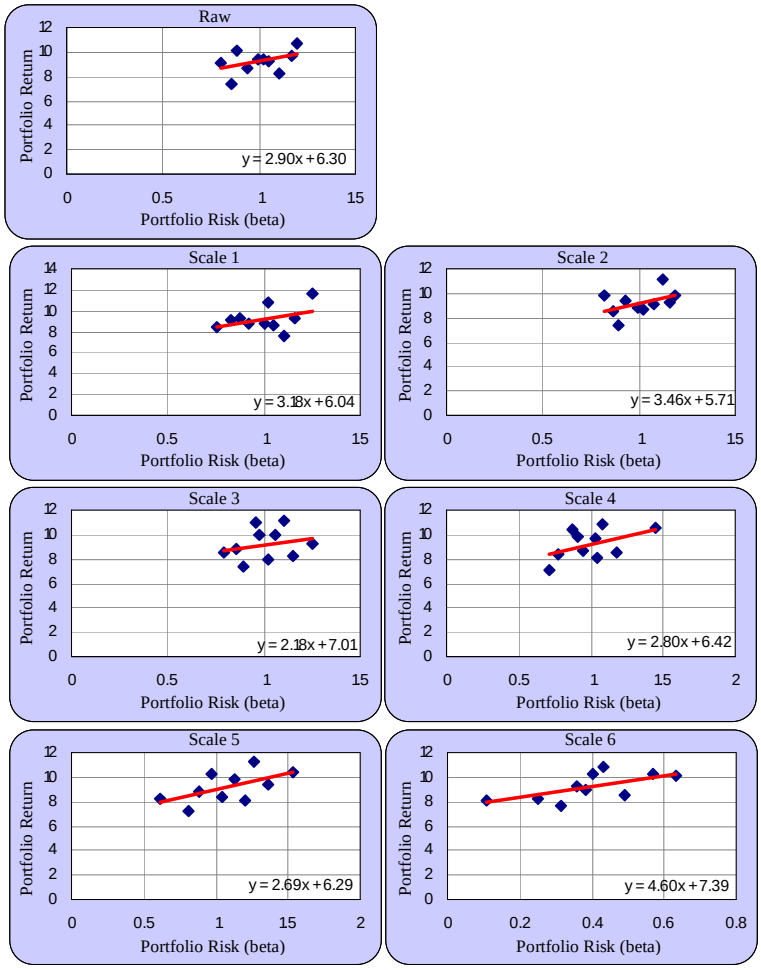


Figure 4.16 Output of the risk management module: Risk-return tradeoffs in the DJIA for 10 portfolios (with 3 companies each) for raw data, and scale 1 (i.e. 2 – 4 days) to scale 6 (i.e. 64 – 128 days). See *Appendix A* for more details.

4.3.3 Separating CHF and Normal Patients

In this case study we examine ECG signals of 45 subjects using our medical diagnosis module. A similar study but with a smaller dataset of 27 subjects (15 CHF and 12 normal) has been reported by Thurner et al. in *Physical Review Letters* [152]. For their data source, the authors refer to the Physionet website [120]. However, no mention is made as to which of the several databases that exist on Physionet have actually been used for the study. For our analysis, we have collected 45 ECG signals from the Physionet website, the details of which are as follows:

Table 4.3 Details of the ECG signals examined.

Database Name	Source	Description	Names of Signals Used	Number of Signals Used	References
The BIDMC Congestive Heart Failure Database	Physionet	ECG recordings from 15 subjects (11 men, aged 22 to 71, and 4 women, aged 54 to 63) with severe congestive heart failure.	CHF01 to CHF15	15	[150]
Congestive Heart Failure RR Interval Database	Physionet	29 long-term ECG recordings of subjects aged 34 to 79, with congestive heart failure.	CHF201 to CHF208	8	[28]
Normal Sinus Rhythm RR Interval Database	Physionet	54 long-term ECG recordings of subjects in normal sinus rhythm (30 men, aged 28.5 to 76, and 24 women, aged 58 to 73).	NSR001 to NSR022	22	[111]

As per Figure 4.8 and algorithm of Figure 4.13, all ECG recordings are converted into R-R 'time' series by our medical diagnosis module before analysis. The number of samples in the 45 R-R recordings range from 75,000 to 100,000. We perform a *level-10* ($J_0 = 10$) MODWT decomposition on the 45 R-R recordings. The output of the medical diagnosis module for the above data is presented in Figure 4.17.

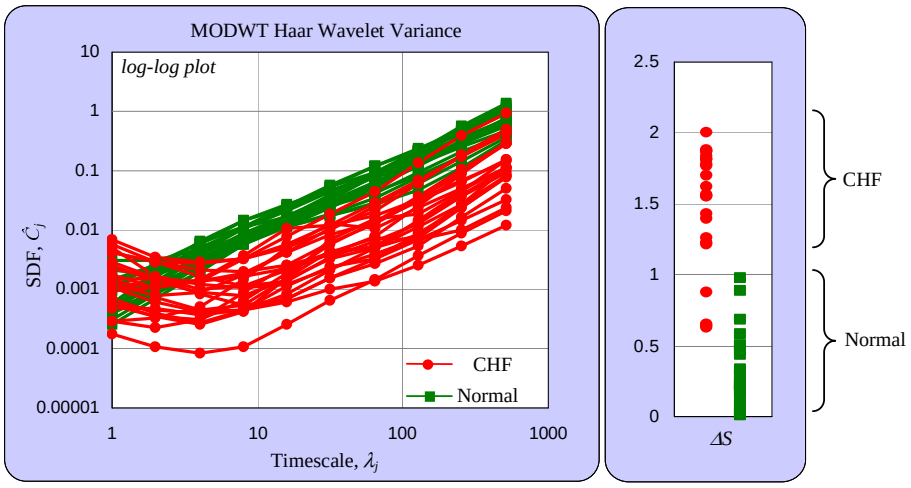


Figure 4.17 Output of the medical diagnosis module: the left-hand plot shows the variation of the MODWT SDF ($\hat{C}_j$) against timescale (λ_j) on a log-log graph while the right-hand plot shows the values of the scaling instability index (ΔS) for all the 45 patients examined.

The plots of CHF patients are shown in red and the plots of normal patients are shown in green to better visualize the separation of the two datasets. The left-hand plot in Figure 4.17 shows the variation of the SDF ($\hat{C}_j$) against timescale (λ_j) on a log-log graph while the right-hand plot shows the values of the scaling instability index (ΔS) for all patients examined. For all 45 signals, the value of ΔS is computed based on Step II-2 of Algorithm 3 of Figure 4.13 using scale 3 as the inflexion point. As discussed earlier in Sections 3.5.3 and 4.1.3, the wavelet variance (or SDF) plots for normal patients are roughly linear across all timescales (constant value of the scaling

exponent α). On the other hand, the wavelet variance (or SDF) plots for CHF are not perfectly linear across all timescales (instable value of the scaling exponent α). This is clearly depicted in the left-hand plot of Figure 4.17, where all the green lines tend to be straight whereas the red lines tend to be bended.

Table 4.4 Values of scaling instability index (ΔS) for all the 45 subjects using the MODWT.

MODWT				
S.No.	Normal		CHF	
	Signal	ΔS	Signal	ΔS
1	NSR001	0.2223	CHF01	1.2285
2	NSR002	0.0446	CHF02	2.0018
3	NSR003	0.3000	CHF03	1.8126
4	NSR004	0.2398	CHF04	1.2577
5	NSR005	0.0749	CHF05	0.8752
6	NSR006	0.8862	CHF06	1.8708
7	NSR007	0.6855	CHF07	1.8316
8	NSR008	0.3369	CHF08	0.6397
9	NSR009	0.5048	CHF09	1.3962
10	NSR010	0.0198	CHF10	1.5605
11	NSR011	0.1848	CHF11	1.7697
12	NSR012	0.0069	CHF12	1.6257
13	NSR013	0.1918	CHF13	1.8764
14	NSR014	0.4441	CHF14	1.7845
15	NSR015	0.2962	CHF15	0.6585
16	NSR016	0.4705	CHF201	1.5529
17	NSR017	0.2168	CHF202	0.6281
18	NSR018	0.5874	CHF203	1.429
19	NSR019	0.5122	CHF204	1.7008
20	NSR020	0.9769	CHF205	1.2200
21	NSR021	0.1208	CHF206	1.8113
22	NSR022	0.2251	CHF207	1.5509
23	-	-	CHF208	1.8652

If we carefully examine the wavelet variance or SDF plots in Figure 4.17 (left-hand plot), we can see a *separation* of healthy and CHF patients at wavelet scales 4 and 5. At all other scales, the values of the SDF for the two datasets seem to be intermingled. As pointed out by Thurner et al. in their *Physical Review Letter*, such a separation can be attributed to the ability of the wavelet analysis to expose a scale window between scale 4 ($2^4 = 16$ heartbeat intervals) and scale 5 ($2^5 = 32$ heartbeat intervals) for which the variances of normal and CHF R-R recordings fall into disjoint sets [152]. We also observe that the overall variability (values of SDF) of normal patients is higher than the overall variability (values of SDF) of CHF patients at most timescales – the green plots tend to be above the red plots at most scales.

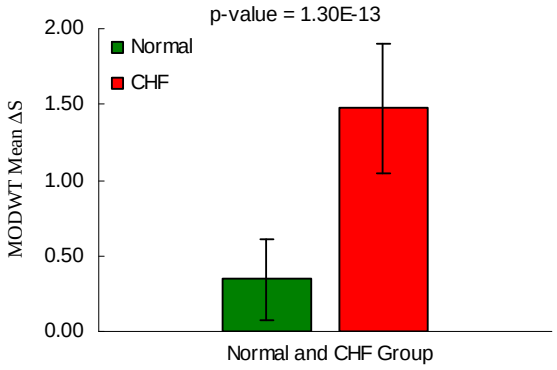


Figure 4.18 Bar plot showing average values of the MODWT-based scaling instability index (ΔS) of the 23 CHF and 22 normal subjects along with the *p value* obtained from a student *t-test*. The positive and negative error bars show the standard deviation of each group.

The right-hand plot in Figure 4.17 shows the values of the scaling instability index ΔS for the two datasets. In most cases, the value of ΔS is higher for CHF patients as compared to normal patients. There are a total of four instances where a CHF patient (red dot) has been wrongly classified as normal (green dots). That is, our system has been able to separate out (classify) CHF and normal patients with a high degree of accuracy where only four signals out of the 45 that were analyzed were classified wrongly. In Table 4.4, we provide the values of ΔS for each of the 45 subjects studied. We plot the average values of ΔS for the CHF and normal group as bars in Figure 4.18. We observe a clear distinction between the red and the green bar in Figure 4.18 which is further confirmed by an extremely low p value of 1.30E-13 for a double tailed, two-sample equal variance student's *t test* [142], [144]. This is further confirmed by the positive and negative error bars (representing the standard deviation of each group), which are significantly apart from each other for the two groups.

Thurner et al. report that 13 out of the 15 CHF subjects that they studied were classified correctly by their system thus giving rise to a *sensitivity* of 87%, and a *specificity* of 100% since their system classified all the 12 normal subjects correctly. Measures of *sensitivity* and *specificity* are based on a two by two contingency table of the type shown in Table 4.5 [135], where,

$$\textit{sensitivity} = A / (A+C),$$

$$\textit{specificity} = D / (B+D), \text{ and,} \qquad 4.10$$

$$\textit{accuracy} = (A+D)/(A+B+C+D).$$

By looking at Eq (4.10) and Table 4.5 we note that in *information retrieval* (IR) systems, the statistic *precision* is analogous to *sensitivity*, whereas *recall* is analogous to *specificity* [124], [147]. We also note from Eq (4.10) that the *accuracy* statistic is a combined measure of the *sensitivity* and *specificity*.

Table 4.5 Contingency table for measuring *specificity, sensitivity and accuracy*.

	Patients with disease	**Patients without disease**
Test is positive	**A**	**B**
Test is negative	**C**	**D**

It is not clear whether Thurner et al. use the MODWT or DWT for their analysis; although the ΔS values that they report are similar to the ΔS values we obtain using the MODWT suggesting that they do use the MODWT. We will later show that ΔS values obtained using the DWT are significantly different than those obtained using the MODWT.

Although Thurner et al. report a *sensitivity* of 87%, if we look at Figure 3 (d) on page 1546 in their letter, they plot ΔS (or Δ, as they call it) values for only 25 out of a total of 27 subjects that were actually studied. Moreover, we can observe from Figure 3 (d) that two ΔS dots for CHF do lie in the normal region. The question that remains unanswered is that whether the reported *specificity* and *sensitivity* corresponds to the entire dataset of 27 subjects or a smaller dataset of 25 subjects as shown in Figure 3 (d) of their paper. On a comparative basis, for a bigger dataset of 45 subjects, the *sensitivity* of our system is 83% since 19 out of 23 CHF signals are classified correctly, whereas the *specificity* of our system is 100% since all the 22 normal signals are classified correctly.

Based on our results, we conclude that heartbeat intervals of the complex human organism (which is a complex system) exhibit scaling behaviour. For a system under normal conditions, the scaling is universal across timescales (fractal behaviour), whereas for an abnormal system this universality is disturbed.

Despite encouraging results produced by our medical diagnosis module, we believe that the classification problem of CHF and normal subjects needs to be further statistically examined. Therefore, in Section 4.4.2, we will put our medical diagnosis module to a more rigorous test, so as to evaluate its robustness in classifying CHF and normal subjects. This would involve classification based on a "leave one out" methodology. Thresholds for scaling instability index (ΔS) and inflexions in the SDF curves (to determine critical scale r1) will be computed automatically each time during training for testing the classification of the "left out" (or unknown) signal. For a more critical evaluation of the performance of our medical diagnosis module, values of *sensitivity* (precision), *specificity* (recall), and *accuracy* will be computed based on the "leave one out" methodology described above. In addition, we will also compare the performance of the MODWT versus the performance of the DWT for classification of CHF and normal subjects, again, based on a "leave one out" methodology.

4.4 Evaluation

In this section we evaluate the performance and robustness of our wavelet framework for complex time series analysis. Two of the three modules namely pattern identification, and medical diagnosis, are put to rigorous and extensive testing. The evaluation of our system has been divided into two parts, namely *human* evaluation, and *statistical* evaluation.

In the human evaluation section (Section 4.4.1) we compare the patterns identified by our pattern identification module with patterns identified by experts for real-world time serial data. In this section, we report the results of a questionnaire based study for the evaluation of our pattern identification module.

The statistical evaluation section (Section 4.4.2) has three subsections – the first two subsections (*Identification of Trends* and *Detecting Variance Breaks*) deal with the statistical evaluation of the pattern identification module whereas the third subsection (*Medical Diagnosis Module*) deals with the statistical evaluation of the medical diagnosis module.

In the first two subsections (*Identification of Trends* and *Detecting Variance Breaks*) of the statistical evaluation section we assess the performance and robustness of our pattern identification module by creating artificial time series with predefined properties. Noise is then added to these time series to determine what level of noise still allows for the identification of features such as trends and variance changes. We use our pattern identification module with different mother wavelets to identify patterns in artificial time series with increasing levels of noise. We then compare the outputs from our system with existing techniques like moving average and cumulative variance statistic.

In the third subsection (*Medical Diagnosis Module*) of the statistical evaluation section we evaluate the performance of our medical diagnosis module. This involves testing of the classification capability of the medical diagnosis module based on a "leave one out" methodology with automatic computation of the ΔS thresholds and SDF inflexions (or critical scale r1) during training. A comparison is made between MODWT and DWT for the classification of CHF and normal subjects based on contingency tables and the computation of statistics such as *sensitivity*, *specificity*, and *accuracy*.

4.4.1 Human Evaluation

A major part of the human evaluation exercise was devoted to assessing the performance of the pattern identification module for identifying trends, turning points and variance changes in complex time serial data. A comprehensive questionnaire was prepared which was then forwarded

to experts for their opinion. The questionnaire comprised two major sections. In the first section, experts were presented with various time series plots and were asked to mark features like trends, turning points, and changes in behaviour with pencil. The second section of the questionnaire showed outputs of our system for various time series and the experts were asked to rate the identified patterns on a scale of minus five, through zero, to plus five. We have prepared an evaluation document that contains details of the questionnaire and the responses we received from various experts from around the world. The evaluation document can be downloaded from: www.saifahmad.com/evaluation.pdf [146]. In *Appendix B*, we show some excerpts of the evaluation we undertook to assess our system.

There were a total of eight experts who actually took time out to complete our questionnaire. Details about the eight experts who evaluated our system can be found in Table 4.6.

Table 4.6 Rating of system output for Q8 to Q12.

Expert #	Name & Details	Trend	Var. Chng.	Turning Points		All Three	Avg. Score
		Q8	Q9	Q10	Q11	Q12	
1	**Dr. Jo Evans** Lecturer, Economics University of Surrey	4.0	3.0	-2.0	-2.0	3.0	**1.2**
2	**Dr. Shah Jalaal Sarker** Department of Public Health Sciences King's College London	4.0	4.0	1.0	3.0	3.0	**3.0**
3	**Prof. Philip Hans B.F. Franses** Econometric Institute, & Department of Marketing & Organization Erasmus University Rotterdam	-5.0	-2.0	-5.0	-1.0	1.0	**-2.4**
4	**Prof. Martin Crowder** Department of Maths / Stats Imperial College	-2.0	-3.0	-5.0	-5.0	-5.0	**-4.0**
5	**Prof. John Nankervis** Department of Economics University of Essex	-3.0	-3.0	1.0	2.0	-2.0	**-1.0**
6	**Mr. Florian Wieners** Trader JRC Berlin	-4.0	-5.0	3.0	4.0	3.0	**0.2**
7	**Dr. A Heen** Researcher London Hedge Fund	2.0	-2.0	0.0	2.0	1.0	**0.6**
8	**Dr. Steve Pollock** Reader, Economics Queen Mary University of London	0.0	5.0	-3.0	-5.0	-3.0	**-1.2**

From the feedback we got from the experts, it is difficult to conclude whether the pattern identification capability of our system is “good” or “bad”. The experts seemed to have a very subjective opinion when marking patterns on a given time series. Therefore, generally every expert marked a different feature on the same time series. This is quite evident from Tables 2a and 2b (*Appendix B*) where there seems to be limited consensus on the location of turning points, variance changes, and trends amongst experts. Moreover, since the experts marked the graphs by hand, it was a difficult exercise to pinpoint exactly which time-point they were referring to. In Figure 2b (*Appendix B*), we show the system output for the identification of a variance change in

the EUR-USD exchange rate series along with the opinion of three experts for the same feature. There seems to be a disparity in opinion amongst the experts as to the exact location of the variance change. One, expert (Expert #3) thinks that there is no change in the variance of the signal.

For turning point identification, we found that a visual comparison is more useful for drawing conclusions about the performance of our system. If we look at Table 2a in *Appendix B*, there seems to be little agreement in the time points identified by the system and the experts as turning points. However, on the other hand, if we look at Figure 4.19, which shows the turning points identified by the system and the experts for S&P 500 index, we observe that the system has been able to pick up 'key' turning points identified by the experts, apart from picking up a few others which the experts did not mark.

If we visually examine the system output and expert opinion for all time series analyzed in the evaluation document [146], we find that the system is actually doing well in identifying various features in a time series.

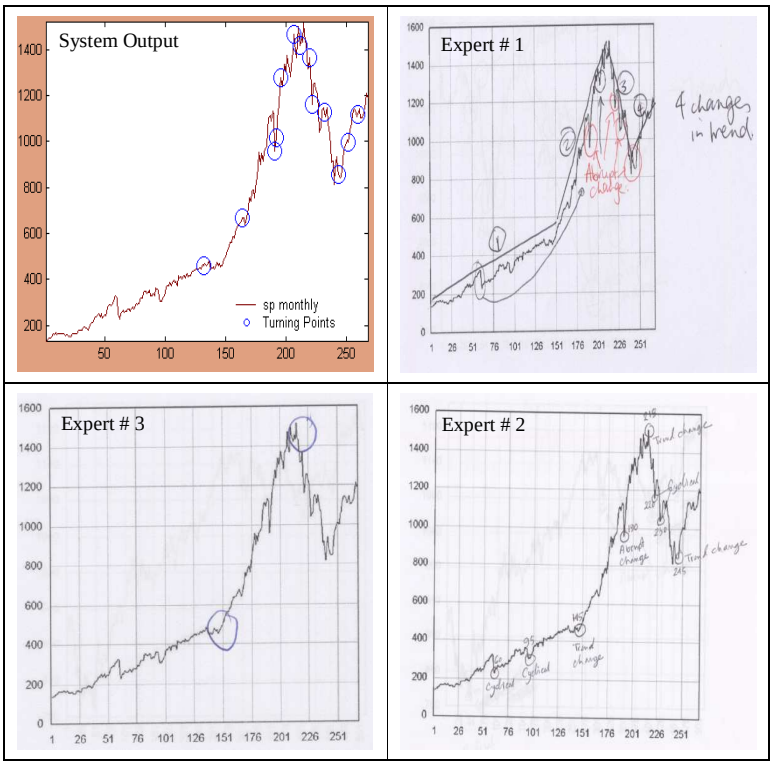


Figure 4.19 System output and expert opinion for turning points in the S&P index.

Table 4.6 presents the results of the experts' opinion on the patterns identified by our system. Here again, we find very subjective and varied opinions. For example, Expert # 8 gives us a five out of five for identifying a variance change (Q9 in Table 4.6), while Expert # 6 gives us a minus five for the same output. The same is true for all other features identified by our system. Then we

have input from Nobel Laureate Professor Clive Granger who thinks that these time series are 'random' or 'stochastic' and hence there is no point studying turning points in them (Table 4.7).

Another important opinion that we were able to gather was from Steve Pincus who has created a method for measuring a signal's "entropy," or disorder and is the author of several papers in the Proceedings of the National Academy of Sciences (PNAS). Steve Pincus did not actually answer our questionnaire, but was satisfied with its design and content. He was of the opinion that we should not restrict ourselves to traders only and should forward our questionnaire to knowledgeable statisticians, and econometricians which we tried to do. Table 2f in *Appendix B* shows Steve Pincus's response.

Table 4.7 Nobel Laureate Clive Granger's views on turning points.

"I do not see the point in discussing turning points for a series that would well be a random walk, particularly for data measured every minute."
– Professor Sir Clive W. J. Granger, Nobel Laureate

We also analyzed some mechanomyographic (MMG) data for Professor Travis Beck at the University of Nebraska. Professor Beck found our analysis useful for "identifying nonstationary versus stationary signals". More specifically, he found the capability of our system to locate a variance change in a signal useful. Table 2f in *Appendix B* shows Professor Beck's response.

We showed the existence of power laws in financial data (Figure 2.15), in data characterizing physical phenomena like ocean shears (Figure 3.13), and in medical data (R-R recordings). With this, we laid the foundation for the use of wavelets as a universal tool across various domains to study the universal principle of scaling in complex systems. Our approach and hypothesis was corroborated by Professor Stephen Burroughs at the University of Tampa, USA, who has co-authored a paper in the Proceedings of the National Academy of Science (PNAS) in which wavelet analysis is used to study shoreline changes [149]. Table 2g in *Appendix B* shows Professor Burroughs's response.

4.4.2 Statistical Evaluation

Identification of Trends

In this section, we will assess the accuracy with which our pattern identification module can detect trends, in artificially generated time series with successive additions of noise.

Let us consider a 2000 samples time series comprising two trends which are essentially two straight lines with different slopes and intercepts joined together at sample number 1024. This two-trended time series is described by Eq (4.11), and is shown in the top panel (blue plot) of Figure 4.20.

$$Trend_1(t) = 0.008\,t - 0.008,\ t < 1024$$
$$Trend_2(t) = -0.006\,t + 8.01,\ 1024 < t < 2000 \qquad 4.11$$

Let us also consider a normally distributed random noise process of 2000 samples which has a sample mean of approximately one and a standard deviation of approximately five. This random noise process (red plot) is shown in the bottom panel of Figure 4.20. We will now test the robustness of our pattern identification module to accurately identify the two slopes (m_1 = 0.008 and m_2 = -0.006) and inflexion (at 1024) in the two-trended synthetic time series (top Panel, Figure 4.20) with the successive addition of increasing levels of random noise (bottom Panel, Figure 4.20). The procedure of adding noise to the synthetic time series is described in the algorithm of Figure 4.21.

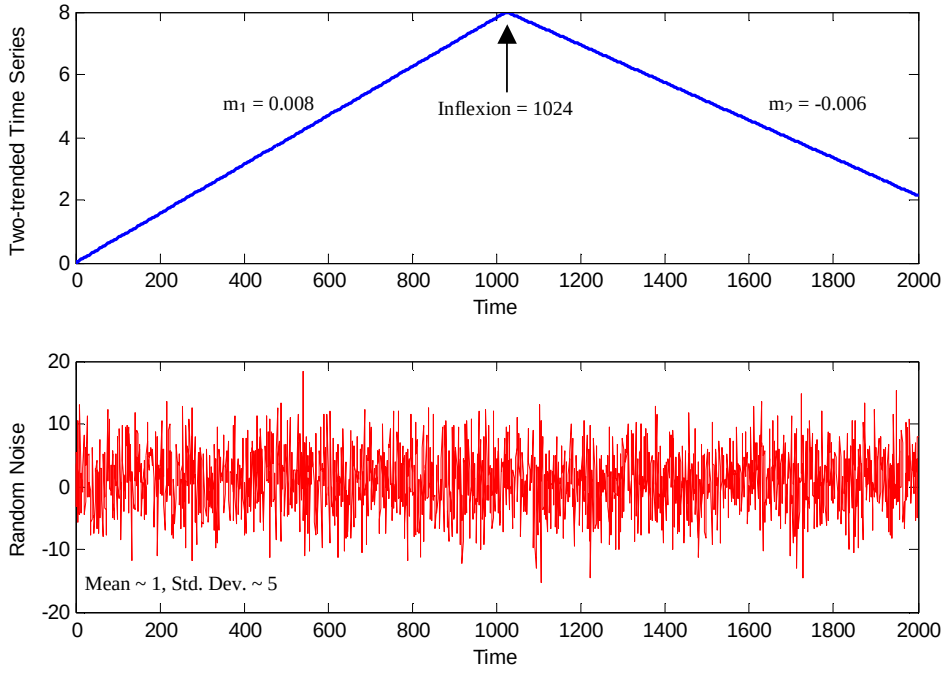


Figure 4.20 Synthetic time series with two trends (top panel) and normally distributed random noise (bottom panel) with N = 2000 samples.

From the algorithm of Figure 4.21 it is clear that we iteratively increase the amplitude of the noise process (ε_t) from 0 to 4.9 in steps of 0.1 which results in a total of 50 such iterations. We also compute a signal to noise ratio (SNR) in decibels (dB) which is defined as "the ratio of a given transmitted signal to the background noise of the transmission medium" [137]. We note that the numerator and denominator of the SNR formula is the *root mean square* (RMS) amplitude of the synthetic time series and the added noise respectively. The SNR statistic offers us a standard measure of the levels of noise in the input signal.

The pattern identification module uses the methodology described in Section 4.1.1, Figure 4.4a, Figure 4.4b, and Algorithm 1 (Figure 4.11) to detect the two slopes m_1 and m_2 and the inflexion

point in the input signal. For a comprehensive evaluation of the performance of the pattern identification module, we will investigate its accuracy in identifying trends with successive addition of increasing levels of noise for different wavelet filters namely Haar, D(4): Daubechies 4-tap, D(6): Daubechies 6-tap, LA(8): Daubechies least asymmetric 8-tap, and C(6): Daubechies Coiflet 6-tap.

Noise Addition and SNR Computation Algorithm

Input : (1) Synthetic time series (x_t) with N samples
(2) Random noise (ε_t) with N samples
Output : Noisy time series (X_t) with N samples

for (i = 0, 0.1, … 4.9)

$X_t = x_t + i\varepsilon_t$

$$SNR = 20\log_{10}\left[\sqrt{\frac{x_1^2 + x_2^2 + ... + x_N^2}{N}} \Big/ \sqrt{\frac{i\varepsilon_1^2 + i\varepsilon_2^2 + ... + i\varepsilon_N^2}{N}}\right]$$

end;

Figure 4.21 Noise addition algorithm.

In addition, we will compare the output from our system with the famous moving average (MA) technique which is often used by *technical analysts* to assess trends in financial time series [26], [109]. Essentially, we will compute a 20-sample (or 20-day if the x-axis of our synthetic time series represented days) MA on the input signal and then determine m_1, m_2 and inflexion point of the MA time series based on methodology described in Section 4.1.1, Figure 4.4a, Figure 4.4b, and Algorithm 1 (Figure 4.11). This will allow us to carry out a one-to-one comparison between the performance of various wavelet filters and the MA technique in identifying trends in the input signal with increasing levels of noise.

Results of the simulations of iteratively adding noise to the input signal and detecting features m_1, m_2 and inflexions by utilizing different wavelet filters and MA technique are presented in Table 4.8; a Microsoft Excel Spreadsheet with the actual numbers for Table 4.8 can also be downloaded from [113]. In this simulation, we stop adding noise to the input signal after 50 iterations (Figure 4.21, Table 4.8) when SNR reaches a value of -13.8566 dB. This is because at this noise level, the accuracy in detecting an inflexion point at 1024 of MA, Haar, and LA (8) filters deteriorates to 6.25% or less (which is significantly low) and remains so for at least 10 iterations. Therefore, a SNR = -13.8566 dB corresponds to a satisfactory level of maximum noise to examine the performance of various wavelet filters and the MA technique in detecting m_1, m_2, and inflexion in the input time series – adding further noise does not change the results in any manner.

By observing the numbers in Table 4.8, we can notice deterioration in performance of wavelet filters and MA as the SNR decreases from infinity up to -13.8566. At SNR = ∞, all wavelet filters

(except Haar) and the MA detect m_1, m_2 and inflexion with an accuracy of 97.5% and above whereas at SNR = -13.8566 there is a significant compromise in performance.

Table 4.8 Simulations for trend detection (m_1, m_2, and inflexion) with increasing levels of noise using various wavelet filters and moving average (MA) technique. *An equivalent Microsoft Excel Spreadsheet can be downloaded from [113].*

SNR	Haar			D(4)			D(6)			LA(8)			C(6)			Moving Average		
	m1	m2	Inflexion	m1	m2	Inflexion	m1	m2	Inflexion	m1	m2	Inflexion	m1	m2	Inflexion	m1	m2	Inflexion
Inf	0.0078	-0.0056	993	0.0078	-0.0060	1026	0.0078	-0.0060	1028	0.0078	-0.0060	1027	0.0079	-0.0060	1027	0.0078	-0.0060	1034
19.9473	0.0076	-0.0058	1025	0.0079	-0.0057	1026	0.0077	-0.0058	1028	0.0077	-0.0059	1029	0.0079	-0.0058	1027	0.0079	-0.0059	1031
13.9267	0.0077	-0.0057	1025	0.0079	-0.0055	1026	0.0077	-0.0056	1029	0.0077	-0.0058	1030	0.0079	-0.0057	1027	0.0080	-0.0060	1031
10.4049	0.0077	-0.0057	1025	0.0079	-0.0052	1026	0.0077	-0.0054	1029	0.0077	-0.0057	1030	0.0079	-0.0055	1027	0.0081	-0.0060	1031
7.9061	0.0078	-0.0056	1025	0.0080	-0.0050	1026	0.0077	-0.0053	1029	0.0077	-0.0056	1030	0.0079	-0.0053	1027	0.0081	-0.0060	1031
5.9679	0.0078	-0.0055	1025	0.0080	-0.0047	1026	0.0076	-0.0051	1029	0.0077	-0.0055	1031	0.0079	-0.0051	1027	0.0082	-0.0060	1031
4.3843	0.0078	-0.0054	1025	0.0079	-0.0047	1058	0.0076	-0.0049	1029	0.0076	-0.0054	1032	0.0079	-0.0049	1027	0.0081	-0.0063	1063
3.0454	0.0079	-0.0054	1025	0.0080	-0.0045	1059	0.0076	-0.0047	1029	0.0076	-0.0053	1032	0.0079	-0.0048	1027	0.0082	-0.0063	1063
1.8855	0.0079	-0.0053	1025	0.0080	-0.0043	1058	0.0076	-0.0046	1029	0.0076	-0.0052	1032	0.0079	-0.0046	1027	0.0083	-0.0064	1063
0.8625	0.0079	-0.0052	1025	0.0081	-0.0041	1058	0.0075	-0.0044	1029	0.0076	-0.0051	1032	0.0080	-0.0044	1027	0.0085	-0.0064	1063
-0.0527	0.0080	-0.0052	1025	0.0082	-0.0038	1058	0.0075	-0.0042	1029	0.0076	-0.0050	1033	0.0080	-0.0042	1027	0.0086	-0.0065	1063
-0.8805	0.0080	-0.0051	1025	0.0083	-0.0036	1058	0.0075	-0.0040	1029	0.0076	-0.0050	1033	0.0080	-0.0040	1027	0.0087	-0.0065	1063
-1.6363	0.0081	-0.0050	1025	0.0083	-0.0034	1058	0.0075	-0.0038	1029	0.0076	-0.0049	1033	0.0080	-0.0038	1027	0.0088	-0.0066	1063
-2.3315	0.0081	-0.0050	1025	0.0084	-0.0032	1058	0.0074	-0.0037	1029	0.0076	-0.0048	1033	0.0080	-0.0037	1027	0.0089	-0.0066	1063
-2.9752	0.0081	-0.0049	1025	0.0085	-0.0030	1058	0.0074	-0.0035	1029	0.0075	-0.0047	1033	0.0080	-0.0035	1027	0.0090	-0.0067	1063
-3.5745	0.0082	-0.0048	1025	0.0001	0.0358	1826	0.0074	-0.0033	1029	0.0075	-0.0046	1033	-0.0007	0.0365	1827	0.0091	-0.0067	1063
-4.1351	0.0082	-0.0048	1025	0.0000	0.0385	1826	-0.0008	0.0352	1828	0.0075	-0.0045	1033	-0.0009	0.0393	1827	0.0092	-0.0068	1063
-4.6617	0.0082	-0.0047	1025	-0.0001	0.0413	1826	-0.0010	0.0377	1828	0.0075	-0.0044	1033	-0.0010	0.0421	1827	0.0094	-0.0068	1063
-5.1581	0.0083	-0.0046	1025	-0.0002	0.0441	1826	-0.0012	0.0403	1828	0.0075	-0.0043	1033	-0.0012	0.0450	1827	0.0095	-0.0069	1063
-5.6277	0.0083	-0.0046	1025	-0.0003	0.0468	1826	-0.0013	0.0428	1828	-0.1486	0.0048	40	-0.0014	0.0478	1827	0.0096	-0.0069	1063
-6.0733	-0.1016	0.0055	64	-0.0004	0.0496	1826	-0.0015	0.0454	1828	-0.1568	0.0050	40	-0.0015	0.0506	1827	0.0097	-0.0070	1063
-6.4971	-0.1069	0.0057	64	-0.0005	0.0524	1826	-0.0016	0.0480	1828	-0.1650	0.0052	40	-0.0017	0.0534	1827	0.0098	-0.0070	1063
-6.9011	-0.1121	0.0059	64	-0.0006	0.0551	1826	-0.0018	0.0505	1828	-0.1732	0.0054	40	-0.0019	0.0562	1827	-0.2380	0.0063	56
-7.2872	-0.1174	0.0061	64	-0.0007	0.0579	1826	-0.0020	0.0531	1828	-0.1814	0.0056	40	-0.0020	0.0591	1827	-0.2491	0.0066	56
-7.6569	-0.1227	0.0064	64	-0.0009	0.0607	1826	-0.0021	0.0557	1828	-0.1896	0.0058	40	-0.0022	0.0619	1827	-0.2603	0.0068	56
-8.0115	-0.1280	0.0066	64	-0.0010	0.0634	1826	-0.0023	0.0582	1828	-0.1978	0.0060	40	-0.0023	0.0647	1827	-0.2715	0.0070	56
-8.3521	-0.1332	0.0068	64	-0.0011	0.0662	1826	-0.0024	0.0608	1828	-0.2061	0.0062	40	-0.0025	0.0675	1827	-0.2826	0.0073	56
-8.6799	-0.1385	0.0070	64	-0.0012	0.0690	1826	-0.0026	0.0633	1828	-0.2143	0.0065	40	-0.0027	0.0703	1827	-0.2938	0.0075	56
-8.9958	-0.1438	0.0073	64	-0.0013	0.0717	1826	-0.0028	0.0659	1828	-0.2225	0.0067	40	-0.0028	0.0732	1827	-0.3050	0.0078	56
-9.3006	-0.1491	0.0075	64	-0.0014	0.0745	1826	-0.0029	0.0685	1828	-0.2307	0.0069	40	-0.0030	0.0760	1827	-0.3161	0.0080	56
-9.5951	-0.1543	0.0077	64	-0.0015	0.0773	1826	-0.0031	0.0710	1828	-0.2389	0.0071	40	-0.0032	0.0788	1827	-0.3273	0.0083	56
-9.8799	-0.1596	0.0079	64	-0.0016	0.0800	1826	-0.0032	0.0736	1828	-0.2471	0.0073	40	-0.0033	0.0816	1827	-0.3385	0.0085	56
-10.1557	-0.1649	0.0082	64	-0.0017	0.0828	1826	-0.0034	0.0761	1828	-0.2553	0.0075	40	-0.0035	0.0844	1827	-0.3496	0.0088	56
-10.4230	-0.1702	0.0084	64	-0.0018	0.0856	1826	-0.0036	0.0787	1828	-0.2635	0.0077	40	-0.0037	0.0872	1827	-0.3608	0.0090	56
-10.6823	-0.1755	0.0086	64	-0.0019	0.0883	1826	-0.0037	0.0813	1828	-0.2717	0.0079	40	-0.0038	0.0901	1827	-0.3720	0.0092	56
-10.9340	-0.1807	0.0088	64	-0.0020	0.0911	1826	-0.0039	0.0838	1828	-0.2799	0.0081	40	-0.0040	0.0929	1827	-0.3831	0.0095	56
-11.1787	-0.1860	0.0091	64	-0.0022	0.0939	1826	-0.0040	0.0864	1828	-0.2882	0.0083	40	-0.0041	0.0957	1827	-0.3943	0.0097	56
-11.4167	-0.1913	0.0093	64	-0.0023	0.0966	1826	-0.0042	0.0890	1828	-0.2964	0.0085	40	-0.0043	0.0985	1827	-0.4055	0.0100	56
-11.6483	-0.1966	0.0095	64	-0.0024	0.0994	1826	-0.0044	0.0915	1828	-0.3046	0.0087	40	-0.0045	0.1013	1827	-0.4166	0.0102	56
-11.8740	-0.2018	0.0097	64	-0.0025	0.1022	1826	-0.0045	0.0941	1828	-0.3128	0.0089	40	-0.0046	0.1042	1827	-0.4278	0.0105	56
-12.0939	-0.2071	0.0100	64	-0.0026	0.1049	1826	-0.0047	0.0966	1828	-0.3210	0.0091	40	-0.0048	0.1070	1827	-0.4390	0.0107	56
-12.3083	-0.2124	0.0102	64	-0.0027	0.1077	1826	-0.0048	0.0992	1828	-0.3292	0.0093	40	-0.0050	0.1098	1827	-0.4501	0.0109	56
-12.5177	-0.2177	0.0104	64	-0.0028	0.1105	1826	-0.0050	0.1018	1828	-0.3374	0.0095	40	-0.0051	0.1126	1827	-0.4613	0.0112	56
-12.7220	-0.2230	0.0107	64	-0.0029	0.1132	1826	-0.0052	0.1043	1828	-0.3456	0.0097	40	-0.0053	0.1154	1827	-0.4725	0.0114	56
-12.9217	-0.2282	0.0109	64	-0.0030	0.1160	1826	-0.0053	0.1069	1828	-0.3538	0.0099	40	-0.0055	0.1183	1827	-0.4836	0.0117	56
-13.1169	-0.2335	0.0111	64	-0.0031	0.1188	1826	-0.0055	0.1094	1828	-0.3620	0.0101	40	-0.0056	0.1211	1827	-0.4948	0.0119	56
-13.3078	-0.2388	0.0113	64	-0.0032	0.1216	1826	-0.0056	0.1120	1828	-0.3703	0.0103	40	-0.0058	0.1239	1827	-0.5060	0.0122	56
-13.4946	-0.2441	0.0116	64	-0.0034	0.1243	1826	-0.0058	0.1146	1828	-0.3785	0.0105	40	-0.0059	0.1267	1827	-0.5171	0.0124	56
-13.6775	-0.2493	0.0118	64	-0.0035	0.1271	1826	-0.0060	0.1171	1828	-0.3867	0.0107	40	-0.0061	0.1295	1827	-0.5283	0.0127	56
-13.8566	-0.2546	0.0120	64	-0.0036	0.1299	1826	-0.0061	0.1197	1828	-0.3949	0.0109	40	-0.0063	0.1323	1827	-0.5395	0.0129	56

As per Table 4.8, in Figure 4.22 we present a visual comparison between the trends detected using DWT LA (8) filter and MA for SNR = ∞ or zero noise. The straight line fits (red plots, Figure 4.22) to the wavelet smooth (middle panel, Figure 4.22) and the MA time series (bottom panel, Figure 4.22) are accurate – the values of m_1, m_2 and inflexion obtained from LA (8) are 0.0078, -0.0060, and 1027 respectively and those obtained from MA are 0.0078, -0.0060, and 1034 respectively, which are close to the actual values (0.008, -0.006, and 1024) of these parameters (Eq (4.11)) in the input time series.

As per Table 4.8, in Figure 4.23 we present another comparison between LA (8) and MA for trend detection but for a SNR = -5.1581 instead of a SNR = ∞. We can observe that both the wavelet smooth (black plot, middle panel, Figure 4.23) and the MA time series (green plot, bottom panel, Figure 4.23) have become quite jagged with the addition of noise in the input signal. Hence the accuracy of the trends detected (red plots, Figure 4.23) has also deteriorated. As a result, the DWT now identifies the inflexion as 1033 instead of 1027 when SNR was equal to infinity. Similarly, the MA now identifies an inflexion at 1063 instead of 1034 when SNR was equal to infinity.

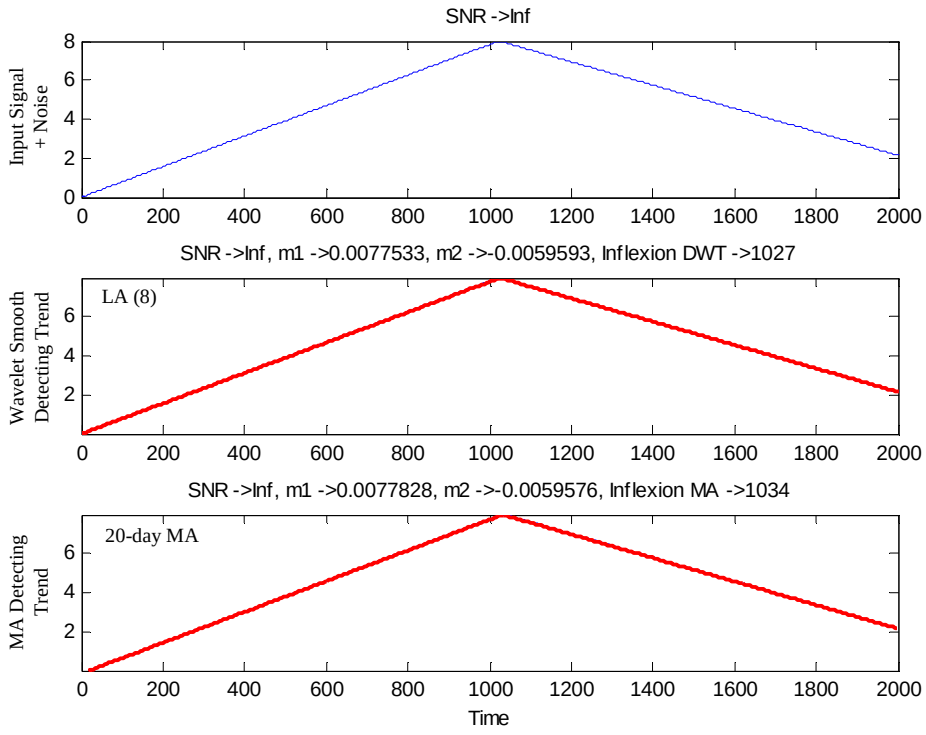


Figure 4.22 Comparison between trends detected using LA (8) DWT filter and MA for SNR = infinity.

In the considered examples (Figures 4.22 and 4.23), the DWT LA (8) filter seems to be performing better than the MA technique in identifying an inflexion point in our synthetic time series for increasing levels of noise. That is, the MA technique seems to be more sensitive to the addition of noise – the MA time series appears more jagged than the wavelet smooth time series for the same level of noise (Figure 4.23). However, for a more comprehensive and complete evaluation, we need to compare the performances of all wavelet filters used and the MA technique in identifying features m_1, m_2, and inflexion in our synthetic time series – the results of this analysis are presented in Figure 4.24.

For comparison purposes, we have set six counters (C_1 to C_6) to zero, where each counter represents one wavelet filter and the MA technique as per Table 4.9.

Table 4.9 Set of counters to assess performance of wavelet filters and MA in identifying m_1, m_2, and inflexion.

Counter Name	Identity	Initial Value
C_1	Haar	0
C_2	D(4)	0
C_3	D(6)	0
C_4	LA(8)	0
C_5	C(6)	0
C_6	MA	0

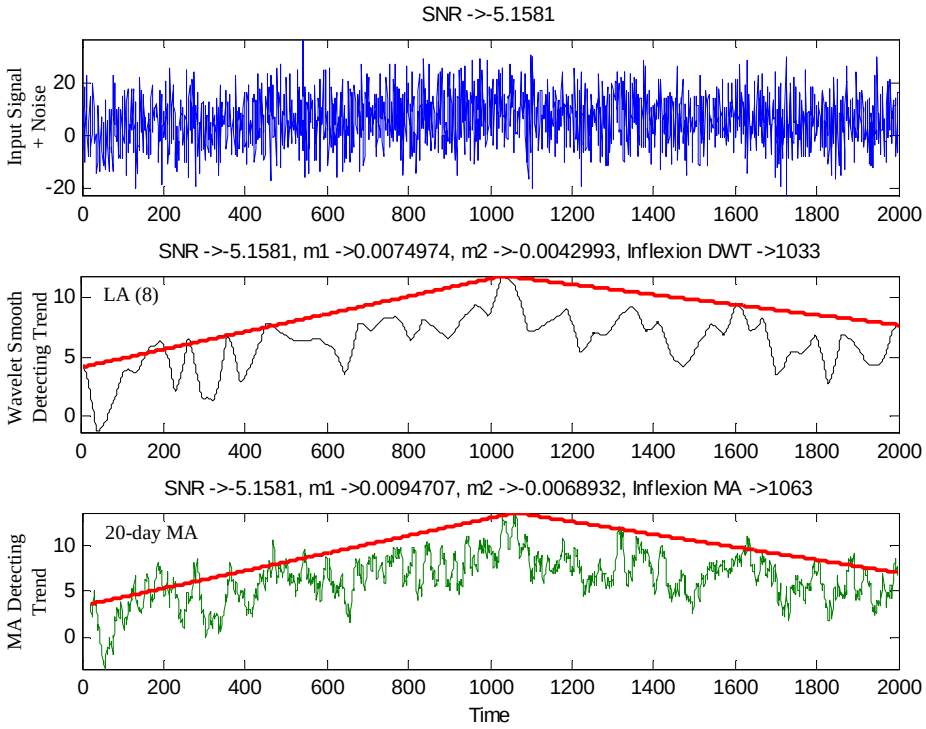


Figure 4.23 Comparison between trends detected using LA (8) DWT filter and MA for SNR = -5.16.

The counters described in Table 4.9 are built three times – once for assessing the accuracy with which m_1 is detected, once for assessing the accuracy with which m_2 is detected, and once for assessing the accuracy with which inflexion is detected. For assessing the accuracy with which m_1 is detected, we set the six counters to zero as per Table 4.9. Now we sequentially go through each row of Table 4.8 and compute the absolute differences between the actual value of m_1 and those detected by the various wavelet filters and MA. Therefore for each row of Table 4.8, we have six values (differences) which correspond to each of the five wavelet filters and the MA. If the first value (difference for Haar filter) is the smallest of the six, then counter C_1 is incremented by one, if the second value (difference for D(4) filter) is the smallest of the six, then counter C_2 is incremented by one, and so on. In other words, for each iteration through Table 4.8, a counter corresponding to a wavelet filter or MA is incremented by one if the difference between the actual value of m_1 and the value of m_1 detected by that filter or MA is the smallest. Also, if more than one technique returns an equal smallest value of the difference between the actual and detected m_1, then counters corresponding to those techniques are also incremented by one. This procedure is repeated from row one to row fifty of Table 4.8 for m_1, m_2, and inflexion separately.

Therefore, for SNR = ∞ to SNR = -13.8566, we have counters described by Table 4.9 set in different states for features m_1, m_2, and inflexion. It is clear that for each feature (m_1, m_2 or inflexion), the counter (C_1 to C_6) with the highest value indicates the best performance for its

corresponding wavelet filter or MA. The graphical and numerical results of this simulation are presented in Figure 4.24.

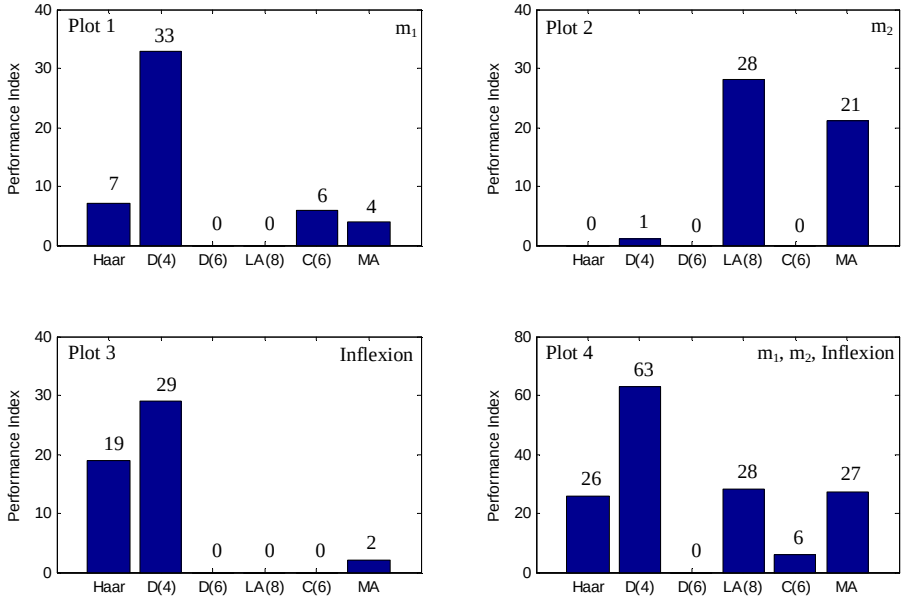


Figure 4.24 Comparison of the trend detection performance of various wavelet filters and MA from a SNR = ∞ to a SNR = -10.9340. Plots 1, 2, and 3 respectively compare the performances for features m_1, m_2 and inflexion considered individually, whereas Plot 4 compares the performance for features m_1, m_2, and inflexion considered together – i.e. the best overall performance.

In Figure 4.24, Plot 1 visualizes the performance of various wavelet filters and MA for detecting m_1, Plot 2 visualizes the performance of various wavelet filters and MA for detecting m_2, and Plot 3 visualizes the performance of various wavelet filters and MA for detecting inflexion. We observe that the wavelet analysis proves superior in identifying all the three features with increasing levels of noise, where one wavelet filter or the other out performs the MA technique (Plots 1, 2, 3, Figure 4.24). In Plot 4 of Figure 4.24 we visualize the performance of various wavelet filters and MA in detecting all three features (m_1, m_2, and inflexion) together. In other words, Plot 4 of Figure 4.24 visualizes the *best overall performance* of all the techniques studied and is the sum of Plots 1, 2, and 3. We can see in Plot 4 of Figure 4.24 that the best overall performance is achieved by wavelet filter D (4) owing to its superior performance in detecting m_1 (Plot 1, Figure 4.24) and inflexion (Plot 3, Figure 4.24). The D (4) wavelet filter is followed by LA (8) and MA for the best overall performance in detecting features m_1, m_2, and inflexion (Plot 4, Figure 4.24).

In this study, we have demonstrated the robustness of our pattern identification module in identifying trends in time series in the presence of noise. Our system shows a fair degree of stability and accuracy in trend identification with addition of increasing levels of noise (Table 4.8). A comparison with the MA technique indicates that our wavelet framework is more robust and superior in detecting trends for increasing noise levels (Table 4.8, Figure 4.23, and Figure

4.24). A clear benefit of the wavelet technique lies in the availability of various filters which may be well-suited for particular tasks. For example, Daubechies D (4) filter performed the best in detecting the slope m_1 (Plot 1, Figure 4.24), the Daubechies LA (8) filter performed the best in detecting the slope m_2 (Plot 2, Figure 4.24), and the Haar filter performed quite well (second only to D (4)) in detecting the inflexion point (Plot 3, Figure 4.24).

Detecting Variance Breaks

We now focus our attention towards the variance break identification capability of our pattern identification module. Let us consider two normally distributed random noise processes, say, ε_1 and ε_2 of 1000 samples each with respective standard deviations of 2 and 5, and a mean of 1 each. These two random noise processes are shown in the left hand and right hand plots of the top panel of Figure 4.25. Let us now append ε_1 and ε_2 to create a new time series ε_3 of 2000 samples as shown in the middle panel of Figure 4.25. Clearly the change in variance or standard deviation of ε_3 lies at sample number 1000 and is depicted by a vertical red line.

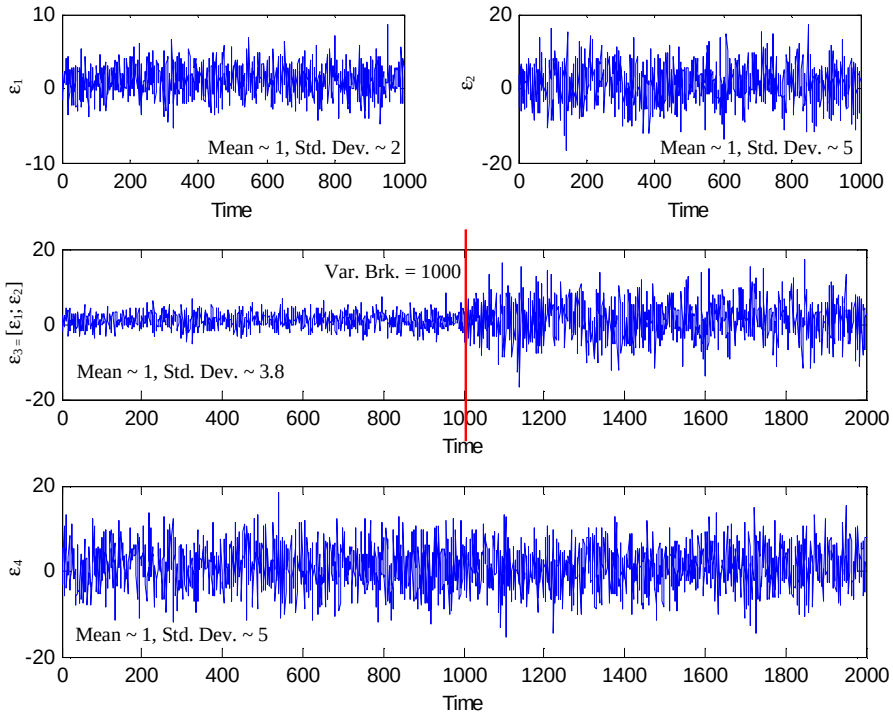


Figure 4.25 Two random noise processes ε_1 (top panel, left plot), and ε_2 (top panel, right plot), of 1000 samples each, with significantly different standard deviations, appended to form a synthetic time series ε_3 (middle panel), resulting in a variance break at t = 1000. Bottom panel is another random noise process ε_4 of 2000 samples which will be iteratively added with increasing amplitudes to ε_3 to obscure its prominent variance change at t = 1000.

Let us also consider a normally distributed random noise process of 2000 samples which has a sample mean of approximately one and a standard deviation of approximately five. This random noise process is shown in the bottom panel of Figure 4.25. We will now test the robustness of our

pattern identification module to accurately identify a change in variance at sample number 1000 in the synthetic time series (middle panel, Figure 4.25) with the successive addition of increasing levels of random noise (bottom panel, Figure 4.25). The procedure of adding noise (ε_4) to the synthetic time series (ε_3) will be similar to the one described in the algorithm of Figure 4.21.

The pattern identification module uses the NCSS statistic described in Section 3.4, Section 4.1.1, and Algorithm 1 (Figure 4.11) to detect a change in variance in the input signal. For a comprehensive evaluation of the performance of the pattern identification module, we will investigate its accuracy in identifying a variance break with successive addition of increasing levels of noise for different wavelet filters namely Haar, D(4): Daubechies 4-tap, D(6): Daubechies 6-tap, LA(8): Daubechies least asymmetric 8-tap, and C(6): Daubechies Coiflet 6-tap.

In addition, we will compare the output for detecting a variance change from our system with the cumulative variance statistic which is a direct outcome of the *null hypothesis* testing technique often used by *statisticians* and *econometricians* to assess stationarity of time serial data [161], [74], [26]. Essentially, the null hypothesis H_0 states that,

$$H_0 : \text{var}\{x_0\} = \text{var}\{x_0, x_1\} = \ldots = \text{var}\{x_0, x_1, \ldots, x_{N-1}\}, \qquad 4.12$$

for a time series $X \equiv [x_0, x_1, \ldots, x_{N-1}]$ with N samples. Based on Eq (4.12), the cumulative variance statistic is defined as,

$$CV \equiv [\text{var}\{x_0\}, \text{var}\{x_0, x_1\}, \ldots, \text{var}\{x_0, x_1, \ldots, x_{N-1}\}]. \qquad 4.13$$

For a strictly stationary process, the cumulative variance (CV) statistic described by Eq (4.13) will comprise a smooth curve almost resembling a straight line. For a nonstationary process, there will be an inflexion in the CV curve where the nonstationarity occurs. An inflexion point in the cumulative variance statistic of Eq (4.13) is determined based on methodology described in Section 4.1.1, Figure 4.4a, Figure 4.4b, and Algorithm 1 (Figure 4.11). This inflexion point will correspond to the location of a variance change in the input signal. This will allow us to carry out a one-to-one comparison between the performance of various wavelet filters inside the NCSS statistic and the cumulative variance technique in identifying a variance break in the input signal with increasing levels of noise.

Results of the simulations of iteratively adding noise (ε_4) to the input signal (ε_3) of Figure 4.25 and detecting a variance break by utilizing different wavelet filters (i.e. computation of the NCSS statistic with different mother wavelets) and cumulative variance technique are presented in Table 4.10. In this simulation, we stop adding noise to the input signal when SNR reaches a value of -12.1696 dB (Table 4.8). This is because at this noise level, the accuracy of the CV statistic in detecting a variance break point at 1000 drops to 30.3% and remains so for at least 10 iterations. On the other hand, the accuracy of all wavelet filters inside the NCSS index in detecting the

variance break point at 1000 remains at 78.4% or higher (resulting in a significantly superior performance of the wavelet filters as compared to the CV statistic) for the last 10 iterations or more. Therefore, a SNR = -12.1696 dB corresponds to a satisfactory level of maximum noise to examine the performance of various wavelet filters inside the NCSS statistic and the CV technique in detecting a variance break in the input time series – adding further noise does not change the results in any manner.

By observing the numbers in Table 4.10, we can notice that at SNR = ∞ or zero noise, all wavelet filters and the cumulative variance detect the change in variance between 998 and 1000 achieving an accuracy of 99.8% or higher. An apparent deterioration in performance of wavelet filters and cumulative variance in detecting the variance change-point at 1000 is observed as the SNR decreases from infinity up to -12.1696 dB.

Table 4.10 Simulations for variance change detection with increasing levels of noise using various wavelet filters (i.e. computation of the NCSS statistic with different mother wavelets) and cumulative variance technique.

SNR	Haar	D(4)	D(6)	LA(8)	C(6)	Cumulative Variance
Inf	1000	1000	1000	998	998	1001
17.9334	1000	1000	1000	998	998	1001
11.9128	1000	1000	1000	998	998	1001
8.3910	1000	1000	1000	998	998	1001
5.8922	1000	1000	1000	998	998	1001
3.9540	1000	1000	1000	998	998	1001
2.3704	1042	1000	1000	998	998	1001
1.0315	1042	1000	1000	998	998	1001
-0.1284	1042	1041	1041	968	1039	1041
-1.1514	1042	1041	969	968	1039	874
-2.0666	1042	895	895	896	896	874
-2.8944	895	895	895	893	893	874
-3.6502	895	895	895	893	893	874
-4.3454	895	895	895	893	893	874
-4.9891	895	895	895	893	893	541
-5.5884	895	895	895	893	893	290
-6.1490	895	895	895	893	893	303
-6.6755	895	895	895	893	893	303
-7.1720	895	895	895	893	893	303
-7.6416	895	895	786	784	893	303
-8.0872	895	786	786	784	784	303
-8.5110	786	786	786	784	784	303
-8.9150	786	786	786	784	784	303
-9.3011	786	786	786	784	784	303
-9.6708	786	786	786	784	784	303
-10.0254	786	786	786	784	784	303
-10.3660	786	786	786	784	784	303
-10.6938	786	786	786	784	784	303
-11.0097	786	786	786	784	784	303
-11.3145	786	786	786	784	784	303
-11.6090	786	786	786	784	784	303
-11.8938	786	786	786	784	784	303
-12.1696	786	786	786	784	784	303

As per table 4.10, in Figure 4.26 we present a visual comparison between the variance break detected using DWT Haar filter inside the NCSS statistic and cumulative variance statistic for SNR = ∞. It is clear from Figure 4.26 (middle panel) that the wavelet NCSS statistic offers a very clean detection of the variance break. On the other hand, the green plot for the cumulative

variance (bottom panel, Figure 4.26) appears distorted at the beginning. Hence we apply our inflexion point detection methodology (Figure 4.4a and Figure 4.4b) to the cumulative variance time series from sample number 200 onwards. This can be observed in the bottom panel of Figure 4.26 where the slanting red line is drawn from sample number 200 till the end. The maximum distance between the slanting red line and the cumulative variance time series determines the location of variance change which is shown by a vertical red line (bottom panel, Figure 4.26).

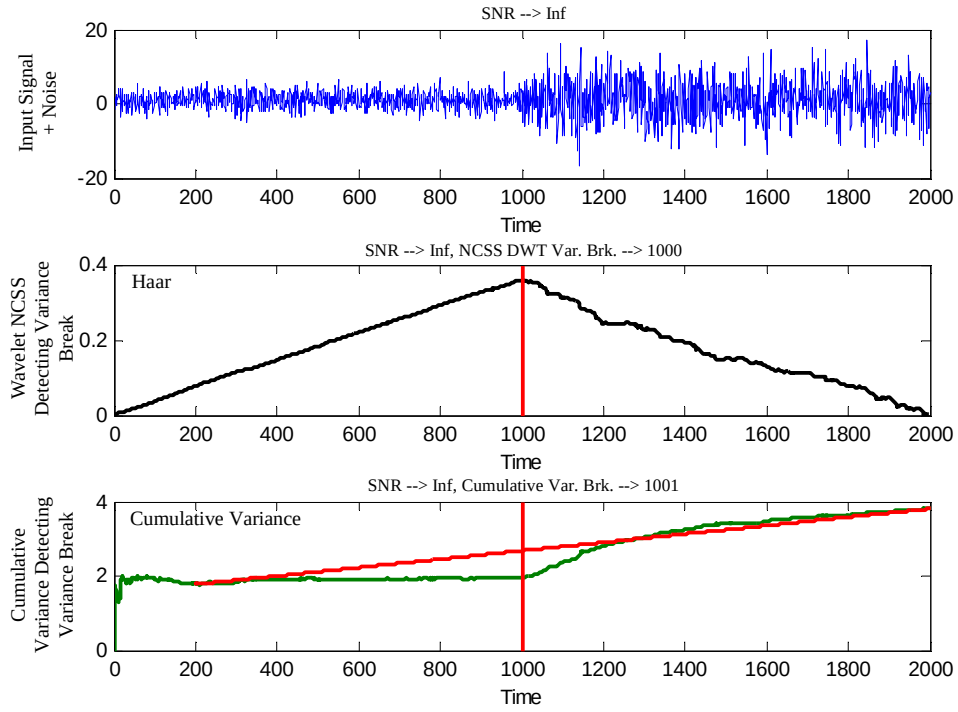


Figure 4.26 Comparison between variance break detected using Haar DWT filter inside the NCSS statistic and the cumulative variance technique for SNR = infinity.

As per Table 4.10, in Figure 4.27 we present another comparison between Haar based DWT NCSS and cumulative variance for detecting the variance break at 1000 but for a SNR = -2.066. We can observe that the wavelet NCSS statistic (black plot, middle panel, Figure 4.27) has become jagged with the addition of noise in the input signal. Hence the accuracy of variance break detected (red vertical line, middle panel, Figure 4.27) has also deteriorated. As a result, our system now identifies a variance break at 1042 instead of 1000. A more pronounced deterioration has been observed in the cumulative variance statistic for detecting a break in variance (bottom panel, Figure 4.27). The cumulative variance statistic now seems relatively flat and the inflexion at the variance change-point does not seem to be significant. As a result, the cumulative variance statistic identifies the variance break point as 874.

In the considered examples (Figures 4.26 and 4.27), the DWT Haar filter inside the NCSS statistic seems to be performing better than the cumulative variance technique in identifying a variance

break-point in our synthetic time series for increasing levels of noise. The cumulative variance technique seems to be more sensitive to the addition of noise – the cumulative variance time series appears flat as compared to the wavelet NCSS time series for the same level of noise (Figure 4.27) thus making the detection of an inflexion point (and hence a variance break-point) in the cumulative variance statistic more difficult. However, for a more comprehensive and complete evaluation, we need to compare the performances of all wavelet filters used inside the NCSS statistic and the cumulative variance technique in identifying variance breaks in our synthetic time series (middle panel, Figure 4.25) with addition of increasing levels of noise (bottom panel, Figure 4.25).

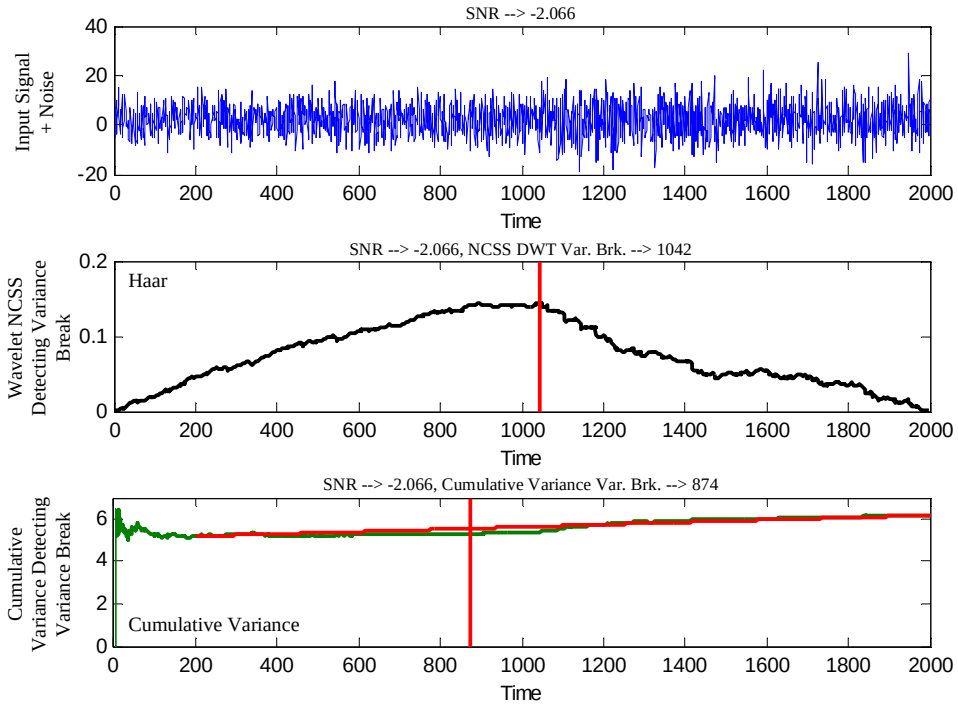


Figure 4.27 Comparison between variance break detected using Haar DWT filter inside the NCSS statistic and the cumulative variance technique for SNR = -2.066.

In Figure 4.28, we present a graphical comparison of the variance break detection performance of various wavelet filters inside the NCSS statistic and the cumulative variance statistic from a SNR = ∞ to a SNR = -12.1696. As per Table 4.10, each curve in Figure 4.28 represents the deviation (or absolute difference) from the correct value of variance change (which is 1000) with increasing noise levels or decreasing SNR ratio. This means that the more a particular curve deviates from the zero y-value for increasing levels of noise (or decreasing SNR), the worst is the performance of the statistic associated with that curve.

We note in Figure 4.28 that the greatest deviation from zero with increasing levels of noise can be observed for the cumulative variance curve. As the SNR drops below -5.5, we notice a deterioration of 57.5% in the performance of the cumulative variance statistic in accurately

detecting the variance break at 1000 – from detecting a variance break point at 874 it now detects the break point at 303. The NCSS statistic which employs different wavelet filters also shows deterioration with increasing noise levels but not as much as that shown by the cumulative variance statistic – for example the variance break detection accuracy of the Haar filter inside the NCSS statistic only drops from 89.5% for SNR = -8.0872 to 78.6% for SNR = -12.1696. Moreover, the accuracy of the various wavelet filters in detecting a variance change seems to be consistent in the sense that they all tend to deviate from zero together (*clustered*) with increasing levels of noise. At the minimum SNR of -12.1696, all wavelet filters deviate by about 215 from the correct value of 1000 whereas the cumulative variance statistic shows a deviation of about 700 for the same noise level.

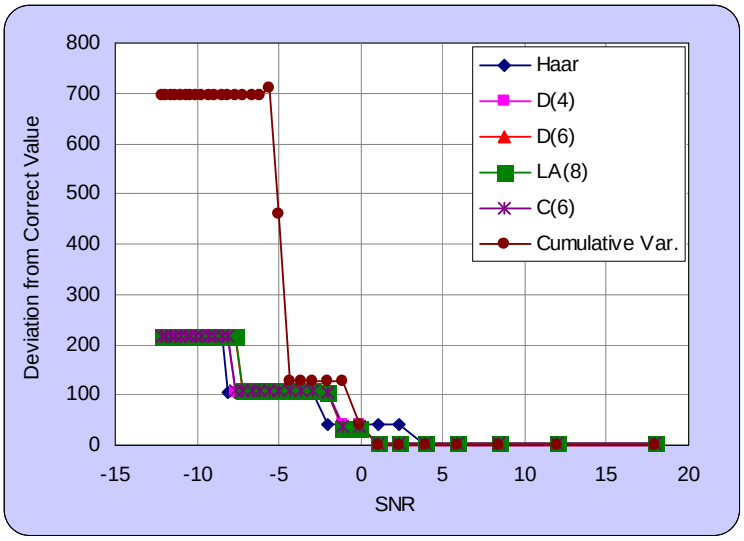


Figure 4.28 Comparison of the variance break detection performance of various wavelet filters inside the NCSS statistic and the cumulative variance statistic from a SNR = ∞ to a SNR = -12.1696. As per Table 4.10 each curve represents the deviation (or absolute difference) from the correct value of variance change (which is 1000) with increasing noise levels or decreasing SNR ratio. Note that we have not shown values for SNR = ∞ on the plot for plotting ease and clarity.

We conclude that our system shows stability and accuracy in detecting variance breaks in time serial data with addition of increasing levels of noise. The performance of all wavelet filters inside the NCSS statistic is similar in accuracy for detecting variance breaks with increasing levels of noise. A comparison with the cumulative variance statistic indicates that our wavelet framework is more robust and superior in detecting variance breaks in a time series for increasing noise levels (Table 4.10, Figure 4.26, Figure 4.27, and Figure 4.28).

We note that we used non-zero mean random noise processes (Figure 4.23 and Figure 4.25) to test the robustness of our pattern identification module and its comparison with other techniques. The choice of non-zero mean noise processes was purely arbitrary. We believe that our results would not have changed if we had used zero mean noise processes instead. This is because the only

parameter on which system performance was mapped was the SNR. The SNR was computed in a similar and consistent fashion for all simulations regardless of the statistical properties of the underlying noise process, thus giving an unbiased evaluation of system robustness and its comparison with other techniques.

Medical Diagnosis Module

In this section, we evaluate the statistical robustness of our medical diagnosis module. In the case study of Section 4.3.3 we introduced the datasets that we wish to study and presented results produced by our medical diagnosis module for these datasets. We compared our methodology and results with those published by Thurner et al. in *Physical Review Letters* [152]. For a dataset of 45 subjects (23 CHF, and 22 normal), the *sensitivity* of our system was 83% and the *specificity* of our system was 100% employing the MODWT-based wavelet variance statistic (or SDF, $\hat{C}_j$), which our system uses as default for separating CHF and normal subjects.

However, for a statistically more rigorous evaluation of our medical diagnosis module, here we tackle the classification of 23 CHF and 22 normal subjects studied in Section 4.3.3 on two fronts:

(a) We employ a *leave-one-out* methodology to train our system to detect an *inflexion point* (or *critical scale r1* of Step II-2 of Algorithm 3 of Figure 4.13) in the wavelet SDF curves for computing ΔS from which we could determine a threshold for ΔS (i.e. *ΔS threshold*) to partition the CHF and normal groups. We then test our system on the *left-out* signal by computing its ΔS value (based on the *inflexion point* determined during training) and classifying it as CHF or normal (based on the *ΔS threshold* determined during training). We repeat this procedure 45 times by *leaving out* each of the 45 signals once.

(b) Using the methodology described in (a), we compare the performance of the MODWT versus the DWT in classifying CHF and normal subjects based on statistics like *sensitivity*, *specificity*, and *accuracy* (Eq (4.10), and Table 4.5).

To the best of our knowledge, no one has ever undertaken a study to classify CHF and normal subjects based on methods described in (a) and (b). Moreover, to the best of our knowledge, nobody has ever analyzed so many ECG signals (i.e. 45) in an effort to classify CHF and normal subjects.

For an initial comparison between MODWT and DWT, let us now produce the same output as produced by our medical diagnosis module in Figure 4.17 but for the DWT instead of the MODWT. In Figure 4.29 we present the output from our medical diagnosis module for classifying 23 CHF and 22 normal subjects (of Table 4.3) using the DWT. If we look at the DWT SDF curves (left-hand plot, Figure 4.29) we do observe a separation between CHF and normal subjects at

scales 4 and 5. However, we also observe that the DWT SDF curves for both CHF and normal subjects appear more curvy or rounded as compared to the MODWT SDF curves (green plots and red plots, left-hand plot, Figure 4.17). A fractal behaviour (SDF curves resembling straight lines), especially like the one observed for normal subjects using the MODWT (green plots, left-hand plot, Figure 4.17), is missing.

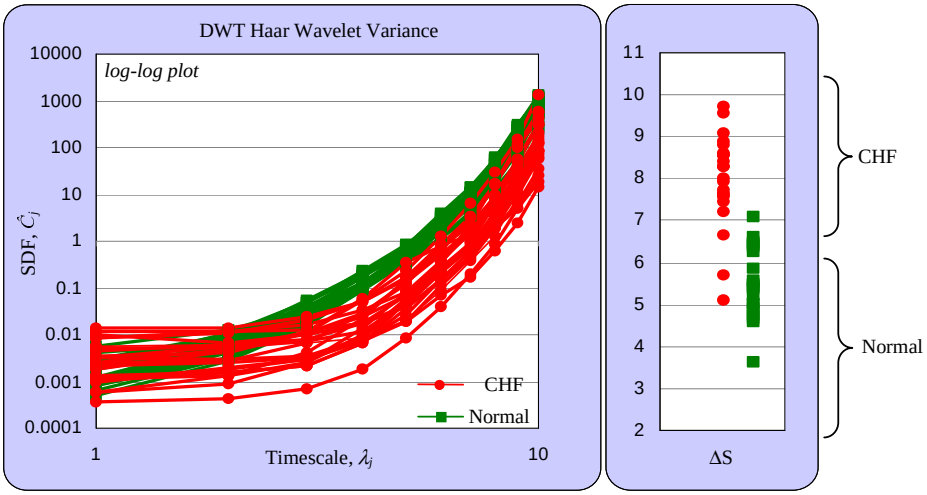


Figure 4.29 Output of the medical diagnosis module: the left-hand plot shows the variation of the DWT SDF ($\hat{C}_j$) against timescale (λ_j) on a log-log graph while the right-hand plot shows the values of the scaling instability index (ΔS) for all the 45 patients examined.

Despite the lack of fractal behaviour in the DWT SDF curves for normal subjects (green plots, left-hand plot, Figure 4.29), and a rather rounded appearance of the DWT SDF curves for all 45 subjects (green plots and red plots, left-hand plot, Figure 4.29), we determine the scaling instability index (ΔS) for each of the 45 subjects in an effort to classify CHF and normal subjects. As per Step II-2 of Algorithm 3, Figure 4.13, the right-hand plot of Figure 4.29 displays the values of ΔS computed based on an *inflexion point* or *critical scale* of 3 for all the 45 DWT SDF curves. We note that the ΔS values obtained using the DWT (right-hand plot, Figure 4.29) are significantly different than those obtained using the MODWT (right-hand plot, Figure 4.17).

In the right-hand plot of Figure 4.29, we observe a separation between CHF and normal subjects based on ΔS values computed for the DWT SDF curves at an *inflexion point* corresponding to scale 3. There are only 3 instances out of 23 where a CHF subject has been wrongly classified as normal resulting in a *sensitivity* of 87% and 0 instances out of 22 where a normal subject has been classified as CHF resulting in a *specificity* of 100%. We recall that using the MODWT for classifying the same dataset (Section 4.3.3) we had obtained a *sensitivity* of 83% and *specificity* of 100%. In Table 4.11, we present the actual numbers for ΔS for all the 45 subjects studied using the DWT SDF curves.

Table 4.11 Values of scaling instability index (ΔS) for all the 45 subjects using the DWT.

DWT				
S.No.	Normal		CHF	
	Signal	ΔS	Signal	ΔS
1	NSR001	6.4765	CHF01	7.9435
2	NSR002	5.8647	CHF02	8.5707
3	NSR003	6.5046	CHF03	8.3937
4	NSR004	5.3756	CHF04	7.6606
5	NSR005	5.5689	CHF05	5.6952
6	NSR006	7.0867	CHF06	8.6103
7	NSR007	6.4262	CHF07	9.5822
8	NSR008	5.5229	CHF08	5.1009
9	NSR009	4.6720	CHF09	8.0119
10	NSR010	6.3740	CHF10	7.6077
11	NSR011	6.2714	CHF11	8.8762
12	NSR012	4.8156	CHF12	8.2906
13	NSR013	5.4189	CHF13	7.5555
14	NSR014	5.0229	CHF14	8.8021
15	NSR015	4.7401	CHF15	6.6621
16	NSR016	5.2120	CHF201	9.0687
17	NSR017	5.4721	CHF202	7.2311
18	NSR018	6.6167	CHF203	7.7268
19	NSR019	4.6029	CHF204	8.2769
20	NSR020	3.6255	CHF205	8.5797
21	NSR021	4.9019	CHF206	8.5509
22	NSR022	5.0142	CHF207	7.4522
23	-	-	CHF208	9.7062

Although the *sensitivity* in classifying CHF and normal subjects is higher using the DWT instead of the MODWT, the differences in the ΔS values of CHF and normal subjects seem statistically less significant for the DWT as compared to the MODWT. If we look at the right-hand plot of Figure 4.29, we observe that the three CHF signals which have been wrongly classified as normal using the DWT tend to go quite deep into the *normal region* in contrast to the classification achieved by the MODWT (right-hand plot, Figure 4.17).

In Figure 4.30, we present bar plots showing average values of the DWT-based scaling instability index (ΔS) for the 23 CHF and 22 normal subjects. We do observe a distinction between the red and the green bar in Figure 4.30, however this distinction is less pronounced than the distinction that was observed for the MODWT (Figure 4.18). This distinction is further confirmed by a p value of 1.20E-10 (for a double tailed, two-sample equal variance student's *t test*) which is 923 times higher than the p value of 1.30E-13 observed for the MODWT – this means that ΔS values for the CHF and normal groups obtained using the MODWT are statistically more different than those obtained using the DWT. Moreover, the error bars for the standard deviation of each group tend to be closer for the DWT ΔS as compared to the MODWT ΔS which confirms the above

observation. From a *data classification* perspective, the more statistically significant the difference in means of ΔS for CHF and normal subjects, the more easier and accurate would be the determination of the ΔS threshold to partition the two groups and vice-versa. Thus, in this case, determination of ΔS threshold for MODWT would be easier as compared to DWT.

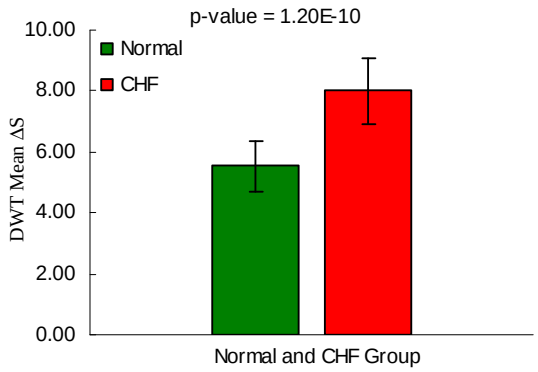


Figure 4.30 Bar plot showing average values of the DWT-based scaling instability index (ΔS) of the 23 CHF and 22 normal subjects along with the *p value* obtained from a student *t-test*. The positive and negative error bars show the standard deviation of each group.

Let us now look at a *leave-one-out* methodology to compare the performance of the MODWT and the DWT in classifying CHF and normal subjects. During *training*, we want our system to automatically determine the *scale ranges* on the SDF curves between which the *scaling instability index* (*ΔS*) could be computed, following which a *ΔS threshold* could be computed so that the *incoming* (or *left-out*) signal could be classified as either CHF or normal based on its *ΔS* during *testing*. To keep the computation of ΔS simple and consistent, we aim to determine only one *critical scale* (or *inflexion point*) for all SDF curves around which we would compute ΔS. Once determined, this *critical scale* would correspond to scale *r1* of Step II-2 of Algorithm 3 of Figure 4.13.

To determine a single *critical scale* (*r1*) around which we could compute ΔS, during training we first determine individual *critical scales* but only on the CHF SDF curves. We determine individual *critical scales* only on the CHF SDF curves because theoretically the normal SDF curves show fractal behaviour and are roughly linear on a log-log graph, hence they do not have an *inflexion point* or a *critical scale* as such. The individual *critical scales* or *inflexion points* in the CHF SDF curves are determined using methodology described in Section 4.1.1, Figure 4.4a, Figure 4.4b, and Algorithm 1 (Figure 4.11). We then take a mean of the *critical scales* or *inflexion points* determined for each of the CHF SDF curves and round it off to zero places of decimal to determine a single *critical scale* (*r1*). We use *r1* as the *critical scale* for computing ΔS for training and testing our data.

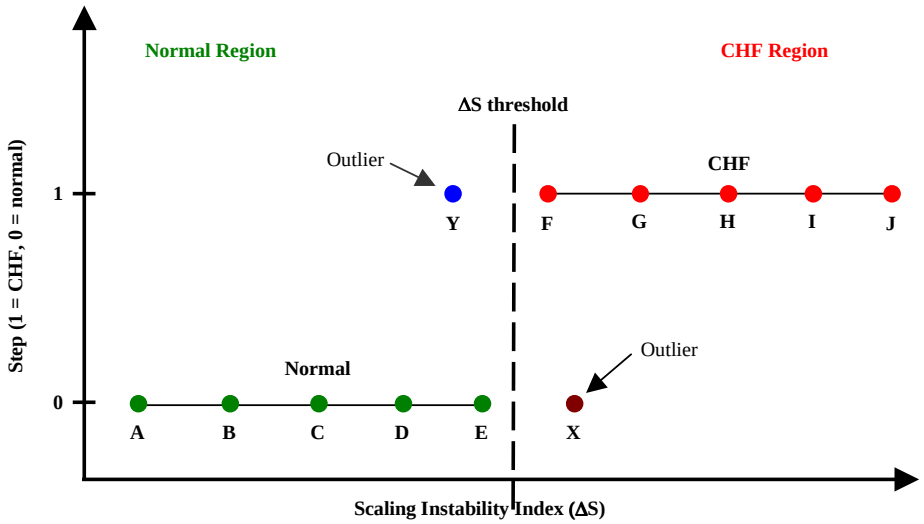


Figure 4.31 Step function learner to classify CHF and normal subjects.

Once a *critical scale* (*r1*) has been determined, our attempt is to use a *step function learner* of the type shown in Figure 4.31 to classify CHF and normal subjects. The *x-axis* of the plot shown in Figure 4.31 represents the scaling instability index (ΔS), whereas the *y-axis* represents the type of signal where 1 implies CHF and 0 implies normal.

Leave-One-Out Algorithm for Classifying CHF and Normal Subjects

Input : 45 R-R signals of CHF and normal patients
Output : Classification of CHF and normal subjects, *sensitivity*, *specificity*, and *accuracy*

counters A = 0; B = 0; C = 0; D = 0;
for signal i = 1 to 45
leave the i^{th} signal out;

I. Compute **SDF curves** for all 44 signals except the i^{th} signal.

I-1. Compute **inflexion points** only for CHF SDF curves of Step I.

I-2. Compute the **average inflexion point** or **critical scale (r1)** by taking mean of results of Step I-1 and rounding it off.

II. Based on the result of Step I-2, **compute ΔS for all 44 SDF curves** of Step I.

II-1. Based on results of Step II, **compute ΔS threshold** by determining the midpoint between the largest normal ΔS value and the smallest CHF ΔS value:

$$\Delta S\ threshold\ = \frac{\left(\Delta S_{L\ \arg\ est}^{Normal} + \Delta S_{Smallest}^{CHF}\right)}{2}.$$

III. Based on the result of Step I-2, **compute ΔS** for the ***i^{th}*** (left-out) **signal**.

IV. Based on results of Steps II-1 and III, **classify the *i^{th}* signal** as either CHF (if $\Delta S^i > \Delta S$ threshold) or normal (if it has a $\Delta S^i \leq \Delta S$ threshold).

V. *if* system classifies ***i*** as CHF and ***i*** is CHF, *then* A = A +1;
else if system classifies ***i*** as CHF and ***i*** is normal, *then* B = B +1;
else if system classifies ***i*** as normal and ***i*** is CHF, *then* C = C +1;
else if system classifies ***i*** as normal and ***i*** is normal, *then* D = D +1;
end;

end;

VI. After 45 iterations (of Step I to Step V), fill Table 4.5 with values for A, B, C, and D. Compute ***sensitivity***, ***specificity***, and ***accuracy*** as per Eq (4.10).

Training

Testing

Figure 4.32 Leave-one-out algorithm for classifying CHF and normal subjects.

Ideally, the *ΔS threshold* partitioning the two datasets would be as shown by the vertical dotted line in Figure 4.31, where *ΔS threshold = (E+F)/2*. However, there may be outliers (blue and dark red dots in Figure 4.31) for both CHF and normal subjects in which case *ΔS threshold = (Y+X)/2*.

Essentially, we have programmed our system to determine *ΔS threshold* as the midpoint between the *largest normal ΔS value* and the *smallest CHF ΔS value* – we believe that this is the most unbiased way of determining the *ΔS threshold*. The *ΔS threshold* line of Figure 4.31 would shift towards left or right depending on how deep the outliers invade a particular (CHF or normal) region.

Once *ΔS threshold* is determined we compute the ΔS value for the incoming (or *left-out*) signal and classify it as either CHF (if $\Delta S^{\text{left-out}} > \Delta S$ *threshold*) or normal (if it has a $\Delta S^{\text{left-out}} \leq \Delta S$ *threshold*). A complete algorithm describing the *leave-one-out* methodology that our system uses for classifying CHF and normal subjects is presented in Figure 4.32.

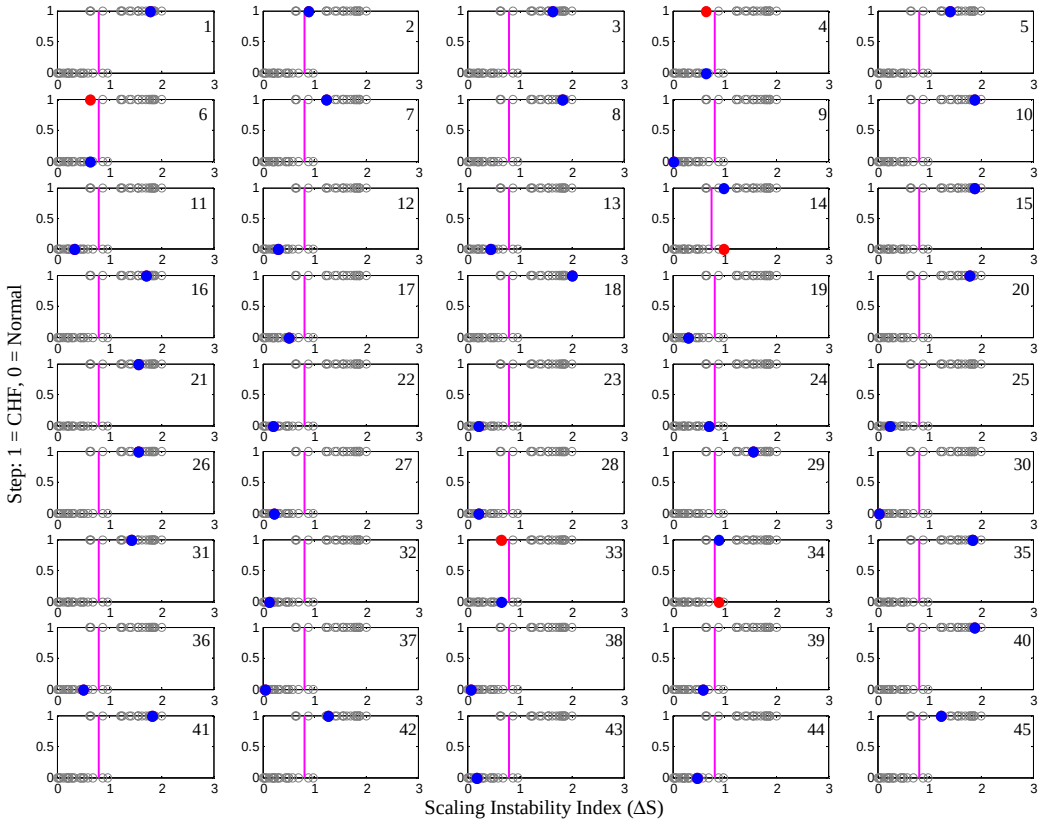


Figure 4.33 Visualization of *leave-one-out* MODWT-based classification of CHF and normal subjects using a step function learner. The blue dots are the subjects identified as CHF or normal by the system: if a blue dot lies on the zero line, it implies normal while if it lies on the one line, it implies CHF. The red dots are where the system went wrong: they show the actual status of a subject when the system makes a mistake. The vertical magenta line represents the *ΔS threshold* value for each case. *A high resolution image of this figure can be downloaded from [157].*

Table 4.12 Numerical values for the visualization of leave-one-out MODWT-based classification of CHF and normal subjects shown in Figure 4.33.

MODWT						
Training				Testing		
Plot #	**Signal Left Out**	**Mean Infl. (r1)**	**ΔS Threshold**	**ΔS for Left Out Signal**	**System (CHF=1, Normal=0)**	**Actual (CHF=1, Normal=0)**
1	CHF14	3.2273	0.8025	1.7845	1	1
2	CHF05	3.2273	0.8025	0.8752	1	1
3	CHF12	3.3182	0.8025	1.6257	1	1
4	CHF202	3.3182	0.8083	0.6281	0	1
5	CHF09	3.2273	0.8025	1.3962	1	1
6	CHF08	3.2727	0.8025	0.6397	0	1
7	CHF205	3.3182	0.8025	1.2200	1	1
8	CHF206	3.2727	0.8025	1.8113	1	1
9	NSR012	3.2609	0.8025	0.0069	0	0
10	CHF208	3.3182	0.8025	1.8652	1	1
11	NSR008	3.2609	0.8025	0.3369	0	0
12	NSR003	3.2609	0.8025	0.3000	0	0
13	NSR014	3.2609	0.8025	0.4441	0	0
14	NSR020	3.2609	0.7572	0.9769	1	0
15	CHF06	3.2727	0.8025	1.8708	1	1
16	CHF204	3.2273	0.8025	1.7008	1	1
17	NSR019	3.2609	0.8025	0.5122	0	0
18	CHF02	3.2727	0.8025	2.0018	1	1
19	NSR015	3.2609	0.8025	0.2962	0	0
20	CHF11	3.2727	0.8025	1.7697	1	1
21	CHF10	3.2727	0.8025	1.5605	1	1
22	NSR013	3.2609	0.8025	0.1918	0	0
23	NSR017	3.2609	0.8025	0.2168	0	0
24	NSR007	3.2609	0.8025	0.6855	0	0
25	NSR004	3.2609	0.8025	0.2398	0	0
26	CHF207	3.1818	0.8025	1.5509	1	1
27	NSR022	3.2609	0.8025	0.2251	0	0
28	NSR001	3.2609	0.8025	0.2223	0	0
29	CHF201	3.3182	0.8025	1.5529	1	1
30	NSR010	3.2609	0.8025	0.0198	0	0
31	CHF203	3.2273	0.8025	1.4290	1	1
32	NSR021	3.2609	0.8025	0.1208	0	0
33	CHF15	3.1364	0.8025	0.6585	0	1
34	NSR006	3.2609	0.8025	0.8862	1	0
35	CHF07	3.2727	0.8025	1.8316	1	1
36	NSR009	3.2609	0.8025	0.5048	0	0
37	NSR002	3.2609	0.8025	0.0446	0	0
38	NSR005	3.2609	0.8025	0.0749	0	0
39	NSR018	3.2609	0.8025	0.5874	0	0
40	CHF13	3.2273	0.8025	1.8764	1	1
41	CHF03	3.2273	0.8025	1.8126	1	1
42	CHF04	3.3182	0.8025	1.2577	1	1
43	NSR011	3.2609	0.8025	0.1848	0	0
44	NSR016	3.2609	0.8025	0.4705	0	0
45	CHF01	3.2727	0.8025	1.2285	1	1

Based on the *leave-one-out algorithm* of Figure 4.32, in Figure 4.33 we present the MODWT-based classification of CHF and normal subjects of Table 4.3 produced by our medical diagnosis module. A high resolution image of Figure 4.32 can be downloaded from [157]. Since one signal out of a total of 45 is *left out* each time during training and then tested, Figure 4.33 has a total of 45 graphs. The vertical magenta lines represent the *ΔS threshold* in each case. The grey dots represent the ΔS values for the *training* set. The blue dots represent the *tested* (or *left out*) signal which the system classifies as either CHF or normal. Where the system calls it correct, there is only the blue dot in that particular plot (case) at step = 1 for CHF or at step = 0 for normal.

However, when the system makes a mistake, we also see a red dot (along with a blue dot) where the red dot tells us the actual status (CHF or normal) of the signal being tested. For example, in Plot #1, first row, Figure 4.33, the system classifies the tested signal as CHF (blue dot) which is correct and hence no red dot. Similarly, in Plot # 4, first row, Figure 4.33, the system classifies the tested signal as normal (blue dot) whereas in actuality this signal is CHF which is shown by the red dot.

The numerical values corresponding to Figure 4.33 are presented in Table 4.12. Numerical values for the individual inflexion points detected on the CHF SDF curves during training (before the computation of the *mean critical scale r1*, or *inflexion point*) for each of the 45 iterations can be downloaded from [108].

Table 4.13 Contingency table for measuring *specificity*, *sensitivity*, and *accuracy* of leave-one-out MODWT-based classification of CHF and normal subjects.

MODWT	Patients with CHF	Patients without CHF
Test is positive	**20**	**2**
Test is negative	**3**	**20**

If we focus our attention on Figure 4.33 and Table 4.12 we can observe that the system makes a mistake on 5 occasions in a total of 45 iterations. The contingency table for the *leave-one-out* MODWT-based classification of CHF and normal subjects is presented in Table 4.13. As per Table 4.13, and Eq (4.10), the *leave-one-out* MODWT-based classification of CHF and normal subjects produces a *sensitivity* of 87%, a *specificity* of 91%, and an *accuracy* of 89%.

For comparison purposes, we present the outputs produced by our system for the *leave-one-out* DWT-based classification of CHF and normal subjects of Table 4.3 in Figure 4.34, and Table 4.14. Numerical values for the individual inflexion points detected on the DWT CHF SDF curves during training (before the computation of the *mean critical scale r1* or *inflexion point*) for each of the 45 iterations can be downloaded from [37].

The contingency table for the *leave-one-out* DWT-based classification of CHF and normal subjects is presented in Table 4.15. As per Table 4.15, and Eq (4.10), the *leave-one-out* DWT-based classification of CHF and normal subjects produces a *sensitivity* of 91%, a *specificity* of 73%, and an *accuracy* of 82%.

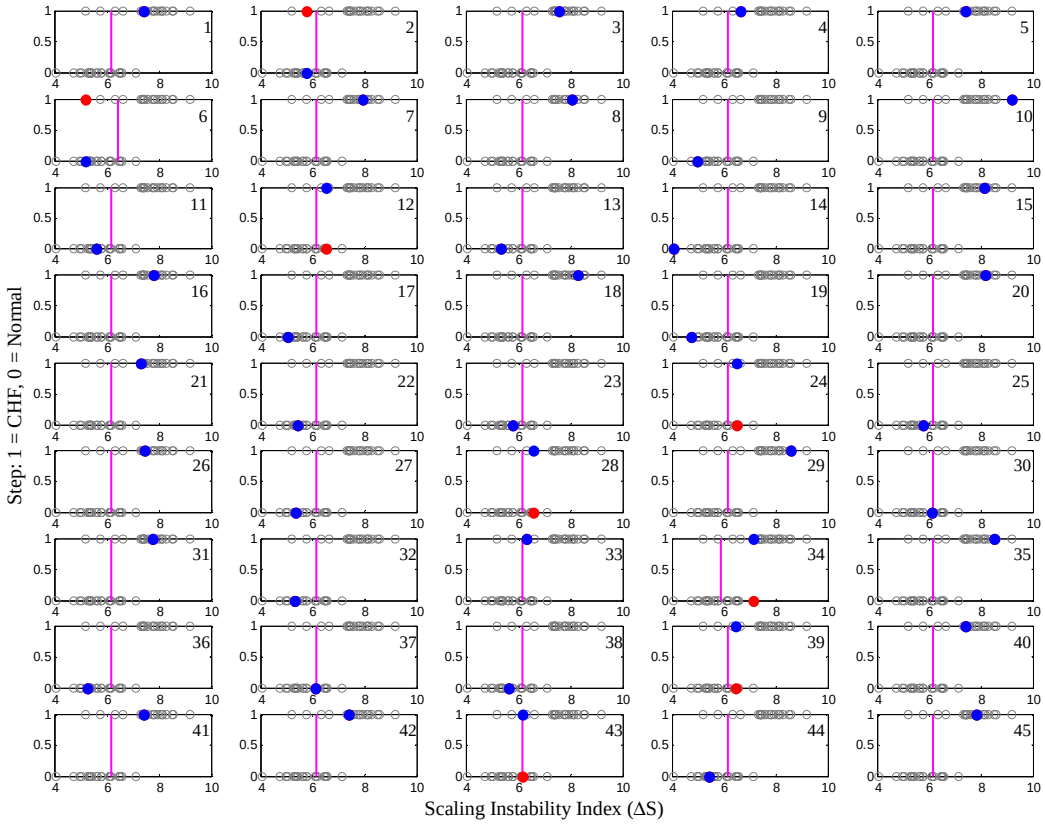


Figure 4.34 Visualization of *leave-one-out* DWT-based classification of CHF and normal subjects using a step function learner. The blue dots are the subjects identified as CHF or normal by the system: if a blue dot lies on the zero line, it implies normal while if it lies on the one line, it implies CHF. The red dots are where the system went wrong: they show the actual status of a subject when the system makes a mistake. The vertical magenta line represents the *ΔS threshold* value for each case. *A high resolution image of this figure can be downloaded from [156].*

In Table 4.16, we summarize the performance of the MODWT and DWT for classifying CHF and normal subjects of Table 4.3, based on the *sensitivity*, *specificity*, and *accuracy* statistics. The DWT shows a higher *sensitivity* of 91% as compared to a lower *sensitivity* of 87% obtained using the MODWT for classifying CHF and normal subjects. This implies that the DWT is more efficient than the MODWT in detecting *true positives* or in other words the DWT proves to be better than the MODWT in correctly classifying ill subjects (CHF subjects).

On the other hand, the MODWT shows a significantly higher *specificity* of 91% as compared to a low *specificity* of 73% obtained using the DWT for classifying CHF and normal subjects. This implies that the MODWT performs significantly better than the DWT in terms of correctly classifying normal subjects or in other words the DWT tends to generate more *false positives* than the MODWT, i.e. the DWT has a greater tendency than the MODWT to wrongly classify normal subjects as CHF. On an overall basis, the MODWT proves to be superior than the DWT in classifying CHF and normal subjects with an *accuracy* of 89% as compared to an *accuracy* of 82% achieved by the DWT.

Table 4.14 Numerical values for the visualization of leave-one-out DWT-based classification of CHF and normal subjects shown in Figure 4.34.

DWT						
Training				Testing		
Plot #	**Signal Left Out**	**Mean Infl. (r1)**	**ΔS Threshold**	**ΔS for Left Out Signal**	**System (CHF=1, Normal=0)**	**Actual (CHF=1, Normal=0)**
1	CHF14	2.9545	6.1602	7.4229	1	1
2	CHF05	2.9091	6.1602	5.7560	0	1
3	CHF12	2.9545	6.1602	7.5798	1	1
4	CHF202	3.0000	6.1602	6.6213	1	1
5	CHF09	2.9091	6.1602	7.3806	1	1
6	CHF08	2.9545	6.4361	5.2042	0	1
7	CHF205	3.0000	6.1602	7.9489	1	1
8	CHF206	2.9545	6.1602	8.0664	1	1
9	NSR012	2.9565	6.1602	4.9541	0	0
10	CHF208	2.9545	6.1602	9.1713	1	1
11	NSR008	2.9565	6.1602	5.6154	0	0
12	NSR003	2.9565	6.1602	6.5237	1	0
13	NSR014	2.9565	6.1602	5.3514	0	0
14	NSR020	2.9565	6.1602	4.0468	0	0
15	CHF06	2.9545	6.1602	8.1323	1	1
16	CHF204	2.9545	6.1602	7.8107	1	1
17	NSR019	2.9565	6.1602	5.0381	0	0
18	CHF02	2.9545	6.1602	8.2793	1	1
19	NSR015	2.9565	6.1602	4.7536	0	0
20	CHF11	2.9545	6.1602	8.1474	1	1
21	CHF10	2.9545	6.1602	7.3112	1	1
22	NSR013	2.9565	6.1602	5.4063	0	0
23	NSR017	2.9565	6.1602	5.7840	0	0
24	NSR007	2.9565	6.1602	6.492	1	0
25	NSR004	2.9565	6.1602	5.7792	0	0
26	CHF207	2.9091	6.1602	7.4565	1	1
27	NSR022	2.9565	6.1602	5.3629	0	0
28	NSR001	2.9565	6.1602	6.5759	1	0
29	CHF201	3.0000	6.1602	8.5601	1	1
30	NSR010	2.9565	6.1602	6.0911	0	0
31	CHF203	2.9545	6.1602	7.7393	1	1
32	NSR021	2.9565	6.1602	5.3225	0	0
33	CHF15	3.0000	6.1602	6.3394	1	1
34	NSR006	2.9565	5.8901	7.1162	1	0
35	CHF07	2.9545	6.1602	8.5078	1	1
36	NSR009	2.9565	6.1602	5.2552	0	0
37	NSR002	2.9565	6.1602	6.1157	0	0
38	NSR005	2.9565	6.1602	5.6588	0	0
39	NSR018	2.9565	6.1602	6.4604	1	0
40	CHF13	2.9091	6.1602	7.3881	1	1
41	CHF03	2.9545	6.1602	7.4197	1	1
42	CHF04	3.0000	6.1602	7.3891	1	1
43	NSR011	2.9565	6.1602	6.1872	1	0
44	NSR016	2.9565	6.1602	5.4127	0	0
45	CHF01	2.9545	6.1602	7.8146	1	1

Table 4.15 Contingency table for measuring *specificity*, *sensitivity*, and *accuracy* of leave-one-out DWT-based classification of CHF and normal subjects.

DWT	**Patients with CHF**	**Patients without CHF**
Test is positive	**21**	**6**
Test is negative	**2**	**16**

Table 4.16 Comparison between MODWT and DWT based on *sensitivity*, *specificity*, and *accuracy* for classifying CHF and normal subjects using the *leave-one-out* methodology.

Measure	MODWT	DWT
Sensitivity	87%	91%
Specificity	91%	73%
Accuracy	89%	82%

In this study we have demonstrated that our system can robustly and accurately classify CHF and normal subjects. With an overall higher *accuracy*, the MODWT proves to be more efficient than the DWT in classifying CHF and normal subjects. Moreover, the MODWT SDF curves show a more fractal behaviour (Figure 4.17) as compared to the DWT SDF curves (Figure 4.29) supporting the hypothesis that the heartbeat intervals of the complex human organism exhibit scaling behaviour. We note here that our medical diagnosis module uses the MODWT as default for separating CHF and normal subjects.

4.5 Conclusion

In Chapter 2, we developed the concept of universal scaling in complex systems and the motivation for a formal method that can help us study this phenomenon systematically. In Chapter 3, we saw the systemacity with which wavelet analysis can decompose complex time series to elucidate complex dynamics at various time horizons.

In this chapter, we have presented a wavelet framework for the analysis of time series produced by complex systems. We have shown the extensibility of our framework for analyzing time series from different domains. We have presented architectures and algorithms for three modules namely, pattern identification, risk management, and medical diagnosis. The first two modules operate in the financial domain whereas the third module analyzes signals in the medical domain (ECG signals). Three comprehensive case studies have been performed to explain the kinds of outputs each of the three modules generate.

In Section 4.4, we provided details of the evaluation undertaken to assess the performance of our system. The evaluation of our system was performed on two fronts, namely *human evaluation* and *statistical evaluation.* The *human evaluation* assessed the performance of the pattern identification module by comparing its outputs for trends, cycles, turning points, and variance changes with those identified by experts. The results of the *human evaluation* were satisfactory, though some discrepancies were found not only between the system and the experts but also amongst the experts themselves. From the human evaluation of our system we learnt that such an

evaluation is very subjective where every expert has his or her own distinctive opinion on the properties and behaviour of the time series being studied.

The *statistical evaluation* of our system assessed the performance of the pattern analysis module and the medical diagnosis module. For the pattern identification module, excellent results were obtained for identifying trends and variance breaks in artificial time series of known properties with successive addition of noise. A comparison with other techniques like the moving average and cumulative variance showed the superiority in performance of our wavelet framework for identifying patterns in time serial data in the presence of noise.

In the *statistical evaluation* of the medical diagnosis module, we adopted a *leave-one-out* methodology to examine a total of 45 ECG signals of CHF and normal subjects. A comparison was made between the performance of the MODWT and DWT for the classification of CHF and normal subjects. Excellent results were obtained for both the MODWT which showed an *accuracy* of 89% and the DWT which showed an *accuracy* of 82% in classifying CHF and normal subjects.

Based on the above, we can suggest that the method presented is suitable for analyzing complex time series across various disciplines. The system presented is a first step towards the development of a robust autonomous system for the analysis of high-frequency complex temporal data produced by various social, biological, and physical systems. The key advantages that the wavelet analysis presents vis-à-vis the development of a robust interdisciplinary time series analysis system are: (1) wavelet analysis is a nonparametric method for time series analysis, and hence no parameters need to be set beforehand (no *a priori* knowledge is required about the dynamics of the data being analyzed) for undertaking an analysis; (2) wavelet analysis has formalized old notions of decomposing complex time series into its various components by making use of a simple pyramidal algorithm [97]; (3) wavelet analysis is not restricted by the assumption of stationarity and is effective in studying *locally volatile* phenomena naturally; and finally, (4) wavelets are effective for studying, in a unified framework, the phenomenon of universal scaling or self organized criticality that exists across various complex systems.

For a system designer, universal tools and principles are important in the development of robust systems that render themselves useful across different domains. In this research, we have laid the foundation for the development of a robust system that can study the universal principle of scaling in complex systems by making use of a universal tool, namely wavelet analysis.

5 Conclusion

In this chapter we provide an overview of the main findings of this research. We discuss and compare the main findings with the aims declared in the Introduction (Chapter 1). We also make suggestions for future research.

Let us first briefly review the work undertaken as part of this research. In Chapter 1, we outlined the major aim of this investigation, namely the development of a novel automatic interdisciplinary time series analysis system. Initial thinking and information about the subject was presented, and expectations related to these questions were outlined.

In Chapter 2, we introduced the concept of complex systems. We discussed that though complex systems exhibit great variety and complexity, universal principles and tools are important in understanding their behaviour. We went on to discuss the universal property of 'scaling' in complex systems, which served as the main motivation for using multiscale wavelet analysis for analyzing time series produced by these systems. The scale invariant power law behaviour was discussed and demonstrated for a range of physical, social and biological systems. We also provided a literature review on various univariate and multivariate time series analysis methods and procedures. We concluded Chapter 2 with a conceptual outline of our interdisciplinary time series analysis system.

In Chapter 3, we motivated a discussion for the search of a universal tool to study the universal phenomenon of scaling in complex systems. We discussed in detail various wavelet filtering methods and techniques. Important statistics based on the wavelet transform (e.g. wavelet variance and covariance) were also discussed. As an application of the wavelet variance, we discussed how wavelets can be used to study power law behaviour in complex systems. Based on the wavelet covariance statistic we demonstrated how we can undertake a multiscale multivariate analysis – we discussed the implementation of the multiscale CAPM with specific examples. In Chapter 3, we also discussed how the wavelet analysis can be used to locate nonstationarities in time series data. Based on wavelet filtering methods and statistics studied, we concluded Chapter 3 with the problem formulation for analyzing complex time series.

In Chapter 4, we provided the architecture and details of our wavelet framework that is capable of accomplishing automatic time series analysis in economics and medicine. We presented three modules, namely, pattern identification, risk management, and medical diagnosis. We also presented three algorithms that summarized the performance of each of these modules. This was followed by three comprehensive case studies to show the types of outputs our system generates. In Chapter 4 we also provided details of the *human* and *statistical* evaluation we undertook to

assess the performance of our system. We concluded that the performance of our system was excellent, and that it is possible to build an autonomous system based on universal principles (scaling) and tools (wavelets).

5.1 Discussion and Directions for Future Work

Our research supports the hypothesis that universal scaling does exist across various social, physical, and biological systems. The universal scaling exhibits itself in two forms, namely, as *scale variance* and as *scale invariance*. We observed that several time series exhibit different dynamics at different timescales, for example, the famous IBM return series showed different temporal patterns and characteristics at different scales. Such behaviour can be termed as *scale variant*, that is, the behaviour and dynamics change with timescale. The other case is that of time series that showed similar dynamics at different timescales. This was observed for R-R heart rate time series where scale invariant power law behaviour was observed for the wavelet variance across timescales, at least for normal subjects. Such behaviour can be termed as fractal behaviour, since the system exhibits similar dynamics at all scales. We conclude that whether scaling is variant, or invariant, it does exist across complex systems, and hence is a universal property of these systems.

Given the existence of universal scaling in complex systems, the discovery of a universal tool would certainly make the analysis of these systems simple and systematic. A range of experiments with various wavelet filtering methods suggest that wavelets are indeed very useful in elucidating complex dynamics across social, physical and biological systems. We found that the formalism provided by the wavelet transform in decomposing complex time series on a timescale basis is useful in the development of robust automatic procedures for time series analysis.

For example, consider the decomposition of a univariate time series into J_0 wavelet levels. We can decompose any complex time series into J_0 wavelet levels based on the same filtering procedure, that is Mallat's pyramidal algorithm [97]. The implementation of Mallat's algorithm is easy and results in a systematic characterization of timescales at each level. For example, scale λ_j at *level j* will correspond to changes in a signal on a scale of 2^{j-1}. This universality will hold true each time we decompose any given signal with wavelets.

Another important possibility provided by the wavelet covariance statistic is to undertake a multiscale multivariate analysis of temporal data. That is, we can study scaling behaviour of bivariate and univariate temporal processes with wavelets. The analysis of risk-return tradeoffs in the DJIA within a CAPM framework, showed the existence of scaling in multivariate processes as well. A multivariate wavelet analysis showed that the *risk* or *beta* of a stock changes as a function

of timescale and that this information could be useful to traders who operate or invest at different timescales.

We thus propose that wavelet analysis is a universal tool for the analysis of the universal principle of scaling in complex systems, and that it can be implemented automatically. Based on our findings, we have developed a wavelet framework for analyzing time series produced by complex systems. We refer to this framework as an *interdisciplinary time series analysis and summarization system*. Our framework has both univariate and multivariate time series analysis capability. We have evaluated our system based on expert opinion, empirical results, and statistical tests. The results of our evaluation are excellent, and suggest that the system developed is not only useful but also robust for studying the *state and behaviour* of various complex systems that surround us.

We believe that automatic time series analysis and summarization using wavelets is a rich area for further study, where several improvements, expansions, reworkings and alternatives are possible. We presented a wavelet framework that has three modules, namely pattern identification, risk management, and medical diagnosis. Several improvements are possible that would enhance the functionality and capability of each of these modules. This would in turn be helpful in the development of a more robust and versatile time series analysis and summarization framework.

Further experiments need to be carried out to improve the accuracy of pattern identification in temporal data using wavelets. For example, in this study, to locate turning points, we just picked up the extremas of each DWT MRA detail d_j. Better results could have been obtained by looking for corresponding extremas across the details d_j. For example, if an extrema occurs at time point t_1 in d_1 and another extrema again appears at time point t_1 in d_2, then the chances of t_1 being a turning point would be higher. Moreover, there is a possibility that the patterns identified using the MODWT instead of the DWT would be more accurate. This is because the *detail* and *smooth* coefficients of a MODWT MRA are associated with zero phase filters. This means that temporal events and patterns in the original signal are meaningfully aligned with the features in the MRA. Experimentation and subsequent *human* and *statistical* evaluation for pattern identification using the MODWT MRA instead of the DWT MRA could be a useful exercise towards improving the accuracy of pattern identification in temporal data using wavelets.

For testing the practical usefulness of the risk management module, a statistical evaluation is important. For example, we could have hypothetical traders 'invest' in stocks as determined by the multiscale risk-return tradeoffs for some training data. This could be followed by investigating the value of the investment made by each trader over different time horizons in the testing data. Such a study would be able to throw light on the quality of the portfolio investment suggested by the risk management module.

A study to examine broader applicability of the wavelet transform to medical and critical care data needs to be undertaken. In this research we showed good results for separating CHF patients from normal patients by examining R-R heart rate time series data. However, further experimentation and evaluation with bigger R-R time series databases would certainly be a useful exercise. Further, it would be interesting to study whether other medical signals like EEGs could be analyzed in a similar fashion to separate *seizure* and *seizure-free* patients (for example).

Prediction of time series data using wavelets is another important area for future research. In previous studies we have shown how wavelet pre-processing can improve the prediction capability of fuzzy systems [123], [122]. However, other statistical time series analysis strategies may also be combined with wavelet analysis to check whether improvements in forecast can be obtained. Since the wavelet transform has the capability of decomposing a time series into the trend, seasonal and irregular components, appropriate prediction techniques can be applied to each component to gain an overall efficiency in forecast. For example, fitting an AR model to the "smooth" or "trend" component, a seasonal autoregressive moving average (SARIMA) model to the seasonal component, and a bootstrapping model to the noise component could result in better forecasts.

5.2 Afterword

As stated in the introduction (Section 1.1, p.2), the major aim of the study reported in this thesis was to develop an automatic procedure for analyzing complex time serial data with three fundamental characteristics: 1) the procedure should be based on a superior analytical method that can provide time series analysis without human assistance; 2) it should borrow insights from research in complex systems, time series analysis, statistics and artificial intelligence, so that it could make a contribution to these fields; and 3) it should have the robustness and capability to analyze signals across different domains. A major characteristic of the procedure developed was that it was not only tested on temporal datasets from two domains, namely, economics and medicine, but also on artificially generated data with pre-defined properties. Because of this, several conclusions may be drawn:

- Ad-hoc techniques for time series analysis and decomposition like first differencing or moving average might prove useful but mainly for data exhibiting simple dynamics (e.g. stationarity, periodicity, linearity etc.) and low noise levels. This has been well proven by the study we undertook in Section 4.4.2 for comparing the performance of our system with the moving average technique for trend detection in a synthetic time series with successive addition of noise. For high signal to noise ratios the moving average technique performed at par with wavelet filtering to detect trends in the artificially generated time

series. However, with successive increase in noise levels, the performance of the moving average technique deteriorated more drastically as compared to wavelet filtering.

- Statistical techniques like the GARCH model might be useful for studying the variance and standard deviation of a process. The GARCH model, developed in econometrics can also be used in general for processes exhibiting volatility clustering. However, there is no provision to study the dynamics of a process at various time horizons in GARCH models. Since temporal scaling is an important universal property of real-world systems, the insights gained by applying a GARCH model to these systems might be somewhat limited.

- The Fourier analysis is a useful alternative to time domain analysis (for example the GARCH analysis) since it determines how much of each frequency the signal contains. However, Fourier analysis alone cannot possibly deal with non-periodic and unobvious temporal patterns that could be contaminated with noise or hidden in the data. The basic premise of Fourier analysis is that the analyzed signal is periodic and stationary. For complex, real-world time series, the periodicity and stationarity conditions are not met. The Fourier analysis decomposes a finite signal into several sine waves (each oscillating at one characteristic frequency) of infinite duration (from $-\infty$ to $+\infty$) to characterize the frequency content of the input signal – no time information is available. On the other hand, the wavelet analysis decomposes a finite signal into several wavelet sub-series (oscillating at different frequencies or timescales) of exactly the same duration as the input signal to characterize not only the frequency content of the input signal but also the local temporal features which are characterized by the extrema of the various wavelet sub-series. Hence wavelets are a better suited tool for analyzing nonstationary, non-periodic, nonlinear, and noisy signals.

The key achievements of the present study are:

- Several wavelet filtering techniques and statistics were used to analyze time series data from different domains. It was demonstrated that wavelets can provide a formal and unique decomposition of processes on a timescale basis to throw light on the complex dynamics of these processes.

- A wavelet-based time series analysis framework was implemented that has multiscale univariate and multivariate analysis capabilities.

- The multiscale univariate capabilities of our framework include automatic pattern identification and summarization in financial time series data, and the analysis of medical and critical care data like ECG signals.

- The multiscale multivariate capabilities of our framework include automatic risk management in financial markets to study risk-return tradeoffs at different time horizons.
- A series of experiments with humans were conducted using our pattern identification module, and positive conclusions were drawn about the performance of our system.
- Statistical evaluation of our pattern identification module was carried out by detecting trends and variance breaks in artificially generated time series with increasing levels of noise. A comparison was made with existing techniques like the moving average and the cumulative variance statistics. Our wavelet framework showed considerable robustness in pattern identification with increasing levels of noise and performed much better than the moving average and cumulative variance statistics.
- We evaluated the performance of our medical diagnosis module based on a *leave one out* methodology and the computation of statistics such as *specificity*, *sensitivity* and *accuracy* to undertake a comparison between the MODWT and the DWT. Although the MODWT proved slightly superior to the DWT in overall performance but excellent values were obtained for *specificity*, *sensitivity*, and *accuracy* for both techniques in classifying CHF and normal subjects.

In summary, the study reported in this thesis has managed to develop a computer-implemented robust automatic procedure for analyzing time series produced by complex systems, following insights from complex systems, time series analysis, statistics and artificial intelligence, and therefore the main aims of the thesis have been achieved.

References

[1] Ahmad, S., Taskaya, T., and Ahmad, K., "Summarizing Time Series: Learning Patterns in 'Volatile' Series," In Yang Z.R., Everson R., Yin H. (Eds.), *Proceedings of the 5th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL)*, Exeter, UK, *Lecture Notes in Computer Science (LNCS)* 3177: 523-532, Heidelberg: © Springer-Verlag, Germany, August 2004.

[2] Allan, D. W., "Statistics of atomic frequency standards," *Proceedings of the IEEE* 31: 221-230, 1966.

[3] Amaral, L. A. N., Goldberger, A. L., Ivanov, P. C., and Stanley, H. E., "Scale-Independent Measures and Pathologic Cardiac Dynamics," *Physical Review Letters* 81 (11): 2388-2391, 1998.

[4] Aussem, A., and Murtagh, F., "Combining neural networks forecasts on wavelet-transformed time series," *Connection Science* 9: 113-121, 1997.

[5] Aussem, A., Campbell, J., and Murtagh, F., "Wavelet-based feature extraction and decomposition strategies for financial forecasting," *J. Comput. Intell. Finance*: 5-12, 1998.

[6] Bak, P., "How Nature Works," *Copernicus-Springer-Verlag*, New York, 1996.

[7] Bar-Yam, Y., "Dynamics of Complex Systems," *Westview Press*, 1997.

[8] Bassingthwaighte, J. B., Liebovitch, L. S., and West, B. J., "Fractal Physiology," *Oxford University Press*, New York, 1994.

[9] Bekaert, G., and Harvey, C. R., "Emerging equity market volatility," *Journal of Financial Economics* 43: 29-78, 1997.

[10] Bekaert, G., and Harvey, C. R., "Time-varying world market integration," *Journal of Finance* 50: 403-444, 1995.

[11] Berardi, V. L., and Zhang, G. P., "An Empirical Investigation of Bias and Variance in Time Series Forecasting: Modelling Considerations and Error Evaluation," *IEEE Transactions on Neural Networks* 14 (3): 668-679, 2003.

[12] Bigger, J. T. (Jr.), Steinman, R. C., Rolnitzky, L. M., Fleiss, J. L., Albrecht, P., and Cohen, R. J., "Power law behaviour of RR-interval variability in healthy middle-aged persons, patients with recent acute myocardial infarction, and patients with heart transplants," *Circulation* 93: 2142-2151, 1996.

[13] Black, F., "Capital market equilibrium with restricted borrowing," *Journal of Business* 45: 444-454, 1972.

[14] Black, F., Jensen, M., and Scholes, M., "Capital asset pricing model: Some empirical tests," in Jensen, M. (Ed.), *Studies in the Theory of Capital Markets*, Praeger, New York, 1972.

[15] Bloomfield, B., "Fourier Analysis of Time Series: An Introduction," *John Wiley & Sons (2nd Edition)*, New York, 2000.

[16] Bollerslev, T., "Generalised Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics* 31: 307-327, 1986.

[17] Bollinger, J., "Bollinger On Bollinger Bands," *McGraw-Hill*, New York, 2002.

[18] Box, G. E. P, and Jenkins, G. M., "Time Series Analysis: Forecasting and Control," *Time Series and Digital Processing* (2nd Edition), Holden Day, San Francisco, 1976.

[19] Boyd, S., "TREND: A System for Generating Intelligent Description of Time-Series Data," *Proceedings of the International Conference on Intelligent Processing Systems (ICIPS'98)*, Australia, 1998.

[20] Bracewell, R. N., "Fourier Transform and Its Applications," *McGraw Hill*, New York, 2000.

[21] Brailsford, T. J., and Faff, R. W., "Testing the conditional CAPM and the effect of intervaling: A note," *Pacific-Basin Finance Journal* 5: 527-537, 1997.

[22] Brailsford, T. J., and Josev, T., "The impact of return interval on the estimation of systematic risk," *Pacific-Basin Finance Journal* 5: 353-372, 1997.

[23] Brealey, R., and Myers, S., "Principles of Corporate Finance," *McGraw Hill (7th Edition)*, New York, 2001.

[24] Campbell, J. Y., "Asset pricing at the millennium," *Journal of Finance* 55: 1515-1567, 2000.

[25] Casillas, J., Cordón, O., and Herrera, F., "COR: A Methodology to Improve *Ad hoc* Data-Driven Linguistic Rule Learning Methods by Inducing Cooperation Among Rules," *IEEE Transactions on Systems, Man, and Cybernetics* 32 (4): 526-537, 2002.

[26] Chatfield, C., "The Analysis of Time Series: An Introduction," *Chapman and Hall / CRC (5th Edition)*, 1996.

[27] Cohen, K., Hawawini, G., Mayer, S., Schwartz, R., and Whitcomb, D., "The Microstructure of Securities Markets," *Prentice-Hall*, Sydney, 1986.

[28] Congestive Heart Failure RR Interval Database, "http://www.physionet.org/physiobank/database/chf2db/,' Accessed on December 1, 2006.

[29] Crosby, J. V., "Cycles, Trends and Turning Points: Practical Marketing and Sales Forecasting Techniques," *McGraw-Hill*, 2000.

[30] Daubechies, I., "Ten Lectures on Wavelets," *SIAM*, Philadelphia, 1992.

[31] Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, S., "Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in United Kingdom," *Economic Journal* 88: 661-692, 1978.

[32] De Alba, E., and Boue, M., "A Bayesian algorithm for detecting multiple changes of variance in a time series," *ITAM and the University of Waterloo*, 2000.

[33] Dempster, M. A. H., and Jones, C. M., "Can channel pattern trading be profitably automated?" *European Journal of Finance* 8 (3): 275-301, 2002.

[34] Donoho, D. L., and Johnstone, I. M., "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association* 90: 1200-1224, 1995.

[35] Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D., "Wavelet shrinkage: asymptopia?" *Journal of the Royal Statistical Society* 57: 301-369, 1995.

[36] Dubrulle, B., Graner, F., and Sornette, D. (Eds.), "Scale Invariance and Beyond," *Les Houches Workshop*, Springer, 1997.

[37] DWT Inflexion Point Detection During Training, "http://www.saifahmad.com/dwt_infl.xls," Accessed on January 24, 2007.

[38] Earthquake Data Source, "http://www.data.scec.org/Module/s2act08.html," Accessed on October 14, 2005.

[39] Elhadad, M., and Robin, J., "An overview of SURGE: A Reusable Comprehensive Syntactic Realization Component," *Technical Report*, Department of Mathematics and Computer Science, Ben Gurion University, Israel, 1996.

[40] Engle, R. F., "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica* 50: 987-1007, 1982.

[41] Engle, R. F., and Ng, V. K., "Measuring and Testing the Impact of News on Volatility," *The Journal of Finance* 48 (5): 1749-1778, 1993.

[42] Fabry, B., Maksym, G. N., Butler, J. P., Glogauer, M., Navajas, D., and Fredberg, J. J., "Scaling the microrheology of living cells," *Physical Review Letters* 87, 2001.

[43] Fama, E. F., and MacBeth, J., "Risk, return and equilibrium: Empirical tests," *Journal of Political Economy* 71: 607-636, 1973.

[44] Faraway, J., and Chatfield, C., "Time series forecasting with neural networks: a comparative study using the airline data," *Applied Statistics* 47: 231-250, 1998.

[45] Farmer, D. J., and Lo, A. W., "Frontiers of finance: Evolution and efficient markets," *Proceedings of the National Academy of Sciences* 96: 9991-9992, 1999.

[46] Fernando, D. A. K., and Jayawardena, A. W., "Runoff forecasting using RBF networks with OLS algorithm," *Journal of Hydrologic Engineering* 3 (3): 203-209, 1998.

[47] Financial Times Online, "http://news.ft.com/home/uk," Accessed on October 8, 2005.

[48] Frankfurter, G., Leung, W., and Brockman, W., "Compounding period length and the market model," *Journal of Economics and Business* 46: 179-193, 1994.

[49] Freeman, J. A., "Simulating Neural Networks with Mathematica," *Addison-Wesley*, Reading, MA, 1994.

[50] Fuller, W. A., "Introduction to Statistical Time Series," *John Wiley & Sons*, New Jersey, 1996.

[51] Gabor, D., "Theory of Communication," *Journal of IEE* 93: 429-457, 1946.

[52] Gallagher, R., and Appenzeller, T., "Beyond reductionism," *Science* 284: 79, 1999.

[53] Garcia, R., and Ghysels, E., "Structural change and asset pricing in emerging markets," *Journal of International Money and Finance* 17: 455-473, 1998.

[54] Gençay, R., Selçuk, F., and Whitcher, B., "An Introduction to Wavelets and Other Filtering Methods in Finance and Economics," *Academic Press*, 2002.

[55] Gencay, R., Selcuk, F., and Whitcher, B., "Systematic risk and timescales," *Quantitative Finance* 3: 108-116, 2003.

[56] Gendron, M., and Ioup, J., "Wavelet Multi-Scale Edge Detection for Extraction of Geographic Features to Improve Vector Map Databases," *Journal of Navigation*, 2000.

[57] Gisiger, T., "Scale invariance in biology: coincidence or footprint of a universal mechanism?" *Biol Rev Camb Philos Soc* 76:161-209 2001.

[58] Goldberger, A. L., "Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside," *Lancet* 347: 1312-1314, 1996.

[59] Gourieroux, C., and Monfort, A., "Time Series and Dynamic Models," *Cambridge University Press*, 1997.

[60] Granger, C. W. J., "Some properties of time series data and their use in econometric model specification," *Journal of Econometrics* 16: 121-130, 1981.

[61] Granger, C. W. J., "The Typical Shape of an Econometric Variable," *Econometrica* 34: 150-161, 1966.

[62] Granger, C. W. J., and Engle, R., "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica* 55: 251-76, 1987.

[63] Granger, C. W. J., and Newbold, P., "Spurious regressions in econometrics," *Journal of Econometrics* 2: 111-120, 1974.

[64] Granger, C. W. J., and Weiss, A. A., "Time series analysis of error-correction models," in S. Karlin, T. Amemiya and L. A. Goodman (eds), *Studies in Econometrics, Time Series and Multivariate Statistics, in Honor of T. W. Anderson*: 255-278, Academic Press, San Diego, 1983.

[65] Gutenberg, B., and Richter, C. F., "Seismicity of the Earth and Associated Phenomenon," *Princeton University Press (2nd Ed.)*, Princeton, 1954.

[66] Handa, P., Kothari, S. P., and Wasley, C., "The relation between the return interval and *betas*: Implications for the size effect," *Journal of Financial Economics* 23: 79-100, 1989.

[67] Handa, P., Kothari. S. P., and Wasley, C., "Sensitivity of multivariate tests of the capital asset pricing to the return interval measurement," *Journal of Finance* 48: 15-43, 1993.

[68] Harvey, C. R., "The world price of covariance risk," *Journal of Finance* 46: 111-157, 1991.

[69] Harvey, C. R., "Time-varying conditional covariances in tests of asset pricing models," *Journal of Financial Economics* 24: 289-317, 1989.

[70] Hawawini, G., "Why *beta* shifts as the return interval changes," *Financial Analysts* 39: 73-77, 1983.

[71] Heisenberg, W., "The Physical Content of Quantum Kinematics and Mechanics," *Magazine for Physics* 43: 172-98, 1927.

[72] Hill, T., O'Conner, M., and Remus, W., "Neural network models for time series forecasts," *Management Science* 42 (7): 1082-1092, 1996.

[73] Ho, D. W. C., Zhang, P. A., and Xu, J., "Fuzzy Wavelet Networks for Function Learning," *IEEE Transactions on Fuzzy Systems* 9 (1): 200-211, 2001.

[74] Hsu, D. A., "Tests for Variance Shift at an Unknown Time Point," *Applied Statistics* 26 (3): 279-284, 1977.

[75] Hull, J., "Options, Futures and other Derivatives," *Prentice Hall (Fourth Edition)*, Upper Saddle River, 2000.

[76] Hunter, J., and McIntosh, N., "Knowledge-Based Event Detection in Complex Time Series Data," In W. Horn et al. (ed), *AIMDM'99: Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making* 271-280, Springer Verlag, 1999.

[77] Hutchinson, J. M., 1994, "A radial basis function approach to financial time series analysis," *PhD dissertation*, Massachusetts Institute of Technology, Cambridge, MA, 1994.

[78] Inclan, C., "Detection of multiple changes of variance using posterior odds," *Journal of Business and Economic Statistics* 11: 289-300, 1993.

[79] J. P. Morgan / Reuters, "RiskMetrics™," *Technical Document (Fourth Edition)*, New York, 1996.

[80] Jang, J. S. R., Sum, C. T., and Mizutani, E., "Neuro-fuzzy and Soft Computing," *Prentice-Hall*, Upper Saddle River, NJ, 1997.

[81] Jayawardena, A. W., and Fernando, D. A. K., "Artificial neural networks in hydro-meteorological modelling," *In Developments in Neural Networks and Evolutionary Computing for Civil and Structural Engineering*, Topping, B. H. V. (ed.), Civil-Comp, Edinburgh 115-120, 1995.

[82] Jayawardena, A. W., and Fernando, D. A. K., "Use of radial basis function type artificial neural networks for runoff simulation," *Computer-Aided Civil and Infrastructure Engineering* 13 (2): 91-99, 1998.

[83] Jayawardena, A. W., Fernando, D. A. K., and Zhou, M. C., "Comparison of multilayer perceptron and radial basis function networks as tools for flood forecasting," *In Destructive Water: Water-Caused Natural Disaster, their Abatement and Control*, International Association of Hydrological Sciences Press: Oxfordshire 173-182, 1996.

[84] Jokinen, V., Syvanne, M., Makikallio, T. H., Airaksinen, K. E., and Huikuri, H. V., "Temporal age-related changes in spectral, fractal and complexity characteristics of heart rate variability," *Clin Physiol* 21: 273-281, 2001.

[85] Juang, C. F., and Lin C. T., "An on-line self-constructing neural fuzzy inference network and its applications," *IEEE Transactions on Fuzzy Systems* 6: 12-32, 1998.

[86] Kaynak, O., Erbatur, K., and Ertugrul, M., "The Fusion of Computationally Intelligent Methodologies and Sliding-Mode Control – A Survey," *IEEE Transactions on Industrial Electronics* 48 (1): 4-17, 2001.

[87] Kazuko, Y., Muchnik, L., Havlin, S., Bunde, A., and Stanley, H. E., "Scaling and memory in volatility return intervals in financial markets," *Proceedings of the National Academy of Sciences* 102 (26): 9424-9428, 2005.

[88] Keogh, E., "A fast and robust method for pattern matching in time-series data," *In Proceedings of WUSS-97*, USA, 1997.

[89] Keogh, E., Chu, S., Hart, D. and Pazzani, M., "An Online Algorithm for Segmenting Time Series," *In Proceedings of the First IEEE International Conference on Data Mining* 289-296, USA, 2001.

[90] Kobayashi, M., and Musha T, "1/f fluctuation of heartbeat period," *IEEE Transactions on Biomedical Engineering* 29: 456-457, 1982.

[91] Lachtermacher, G., and Fuller, J., D., "Backpropagation in time-series forecasting," *Journal of Forecasting* 14: 381-393, 1995.

[92] Lerner, B., Guterman, H., and Dinstein, I., "A Classification-Driven Partially Occluded Object Segmentation (CPOOS) Method with Application to Chromosome Analysis," *IEEE Transactions on Signal Processing* 46 (10): 2841-2847, 1998.

[93] Levhari, D., and Levy, H., "The capital asset pricing model and the investment horizon," *Review of Economics and Statistics* 59: 92-104, 1977.

[94] Lintner, J., "The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets," *Review of Economics and Statistics* 47: 13-37, 1965.

[95] Lisi, F., and Schiavo, R. A., "A comparison between neural networks and chaotic models for exchange rate prediction," *Computational Statistics and Data Analysis* 30: 87-102, 1999.

[96] Löwenborg, P., Johansson, H., and Wanhammar, L., "Two-Channel Digital and Hybrid Analog / Digital Multirate Filter Banks With Very Low-Complexity Analysis or Synthesis Filters," *IEEE transactions on Circuits and Systems II (Analog and Digital Signal Processing)* 50 (7): 355-367, 2003.

[97] Mallat, S., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11: 674-93, 1989.

[98] Mallat, S., and Zhong, S., "Characterization of Signals from Multiscale Edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (7): 710-732, 1992.

[99] Mandelbrot, B. B., "The Fractal Geometry of Nature," (French edition published 1975), *Freeman*, New York, 1983.

[100] Mandelbrot, B. B., and Hudson, R. L., "The (mis) Behaviour of Markets: A Fractal View of Risk, Ruin, and Reward," *Profile Books Ltd.*, London, 2004.

[101] MarketVolume™, "http://www.marketvolume.com/," Accessed on October 9, 2005.

[102] Marr, D., "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information," *W. H. Freeman*, San Francisco, 1982.

[103] Matlab Wavelet Toolbox, "http://www.mathworks.com/products/wavelet/," Accessed on November 22, 2005.

[104] McCoy, E. J., and Walden, A. T., "Wavelet analysis and synthesis of stationary long-memory processes," *Journal of Computational and Graphical Statistics* 5: 26-56, 1996.

[105] McKeown, K., Kukich, K., and Shaw, J., "Practical issues in automatic documentation generation," *In Proceedings of the Fourth ACL Conference on Applied Natural Language Processing (Association for Computational Linguistics)*, Stuttgart, 1994.

[106] Medeiros, M. C., Veiga, A., and Pedreira, C. E., "Modelling Exchange Rates: Smooth Transitions, Neural Networks, and Linear Models," *Transactions on Neural Networks* 12 (4): 755-764, 2001.

[107] Meyer, Y., "Wavelets: Algorithms and Applications," *Society for Industrial and Applied Mathematics*, Philadelphia, 1993.

[108] MODWT Inflexion Point Detection During Training, "http://www.saifahmad.com/modwt_infl.xls," Accessed on January 24, 2007.

[109] Moving Average on Wikipedia, "http://en.wikipedia.org/wiki/Weighted_moving_average," Accessed on January 6, 2007.

[110] New England Complex Systems Institute (NECSI), "http://www.necsi.org," Accessed on September 22, 2005.

[111] Normal Sinus Rhythm RR Interval Database, "http://www.physionet.org/physiobank/database/nsr2db/," Accessed on December 1, 2006.

[112] Nozaki, K., Ishibuchi, H., Tanaka, H., "A simple but powerful heuristic method for generating fuzzy rules from numerical data," *Fuzzy Sets and Systems* 86 (3): 251-270, 1997.

[113] Numbers for Trend Detection, "http://www.saifahmad.com/trend.xls," Accessed on January 21, 2007.

[114] Oppenheim, A. V., and Schafer, R. W., "Discrete Signal Processing," *Prentice-Hall*, New Jersey, 1989.

[115] Palit, A. K., and Popovic, D., "Fuzzy Logic Automatic Rule Generation and Forecasting of Time Series," *In Proceedings of IEEE International Fuzzy Systems Conference* 360-365, Seoul, South Korea, 1999.

[116] Percival, D. B., "On estimation of the wavelet variance," *Biometrika* 82: 619-631, 1995.

[117] Percival, D. B., and Guttorp, P., "Long-memory processes, the Allan variance and wavelets," In E. Foufoula-Georgiou and P. Kumar (Eds.), *Wavelets in Geophysics, Wavelet Analysis and its Applications*: 325-344, Academic Press, San Diego, 1994.

[118] Percival, D. B., and Walden, A. T., "Wavelet Methods for Time Series Analysis," *Cambridge University Press*, 2000.

[119] Phillips, A. W., "Stabilization policy and the time forms of lagged responses," *Economic Journal* 67: 265-277, 1957.

[120] Physionet, "http://www.physionet.org/," Accessed on January 14, 2007.

[121] Pikkujamsa, S. M., Makikallio, T. H., Sourander, L. B., Raiha, I. J., Puukka, P., Skytta, J., Peng, C. K., Goldberger, A. L., and Huikuri, H. V., "Cardiac interbeat interval dynamics from childhood to senescence: comparison of conventional and new measures based on fractals and chaos theory," *Circulation* 100: 393-399, 1999.

[122] Popoola, A., Ahmad, S., and Ahmad, K., "A Fuzzy-Wavelet Method for Analyzing Non-Stationary Time Series," *Proceedings of the 5th International Conference on Recent Advances in Soft Computing (RASC)*, Nottingham, UK, December 2004.

[123] Popoola, A., Ahmad, S., and Ahmad, K., "Multiscale Wavelet Preprocessing for Fuzzy Systems," *Proceedings of ICSC Congress on Computational Intelligence Methods and Applications (CIMA 05)*, Istanbul, Turkey, December 2005.

[124] Precision and Recall on Wikipedia, "http://en.wikipedia.org/wiki/Information_retrieval#Precision," Accessed on December 26, 2006.

[125] Priestley, M. B., "Spectral Analysis and Time Series," *Academic Press*, London, 1981.

[126] Radcliffe, R. C., "Investment," *Harper Collins*, New York, 1994.

[127] Ramsey, J. B., "The contribution of wavelets to the analysis of economic and financial data," *Phil. Trans. R.. Soc. Lond. A* 357: 2593-2606, 1999.

[128] Reiter, E., and Dale, R., "Building Natural Language Generation Systems," *Cambridge University Press*, 2000.

[129] Reuters Kobra, "http://www.about.reuters.com/productinfo/kobra/index.aspx," Accessed on October 7, 2005.

[130] Reynolds, S. B., Mellichamp, J. M., and Smith, R. E., "Box-Jenkins Forecast Model Identification," *AI Expert*: 15-28, 1995.

[131] Ross, S. A., Westerfield, R. W., and Jaffe, J., "Corporate Finance," *Richard Irwin*, Chicago, 1996.

[132] Sargan, J. D., "Wages and prices in the United Kingdom: A study in econometric methodology," in P. E. Hart, G. Mills and J. N. Whittaker (eds), *Econometric Analysis for National Economic Planning*, Butterworths, London, 1964.

[133] Scholes, M., and Williams, J., "Estimating *betas* from non-synchronous data," *Journal of Financial Economics* 5: 309-327, 1977.

[134] Seely, A. J. E., and Macklem, P. T., "Complex systems and the technology of variability analysis," *Critical Care* 8 (6): 367-384, 2004.

[135] Sensitivity and Specificity, "http://www.poems.msu.edu/EBM/Diagnosis/SensSpec.htm," Accessed on December 26, 2006.

[136] Sharpe, W., "Capital asset prices: A theory of market equilibrium under conditions of risk," *Journal of Finance* 19: 425-442, 1964.

[137] SNR on Wikipedia, "http://en.wikipedia.org/wiki/Signal-to-noise_ratio," Accessed on January 5, 2007.

[138] Sripada, S. G., Reiter, E., Hunter, J. and Yu, J., "Modelling the Task of Summarising Time Series Data Using KA Techniques," In A. Macintosh, M. Moulton, and A Preece (eds), *Applications and Innovations in Intelligent Systems IX* 183-196, Springer-Verlag, 2001.

[139] Sripada, S. G., Reiter, E., Hunter, J., and Yu, J., "Generating English Summaries of Time Series Data Using the Gricean Maxims," *In Proceedings of the Ninth ACM SIGMOD International Conference on Knowledge Discovery and Data Mining (KDD-2003)* 187-196, USA, 2003.

[140] Stanley, H. E., Amaral, L. A. N., Buldyrev, S. V. , Gopikrishnan, P., Plerou, V., and Salinger, M. A., "Self-organized complexity in economics and finance," *Proceedings of the National Academy of Sciences* 99 (1): 2561-2565, 2002.

[141] Stanley, H. E., Amaral, L. A., Goldberger, A. L., Havlin, S., Ivanov, P., and Peng, C. K., "Statistical physics and physiology: monofractal and multifractal approaches," *Physica A* 270: 309-324, 1999.

[142] Statistically Significant Consulting, "http://www.statisticallysignificantconsulting.com/Ttest.htm?gclid=CJe_9Nnm6IkCFSFWPgodiS6OJw," Accessed on December 25, 2006.

[143] Stern, H. S., 1996, "Neural networks in applied statistics," *Technometrics* 38: 205-220, 1996.

[144] Student's t-test on Wikipedia, "http://en.wikipedia.org/wiki/Student's_t-test," Accessed on December 25, 2006.

[145] Surrey Market Report, "http://www.saifahmad.com/May6/SurreyMarketReport2.htm," Accessed on December 1, 2005.

[146] Surrey Wavelet Framework Evaluation Document, "www.saifahmad.com/evaluation.pdf," Accessed on December 6, 2005.

[147] Sweeney, L., "Performance Measures for Non-Statistical Computer Scientists," *Lecture 3a*, Data Privacy Laboratory, Carnegie Mellon University, USA. URL: http://privacy.cs.cmu.edu/courses/dp1/refs/general/analysis1.pdf, Accessed on January 2, 2006.

[148] Szeto, H. H., Cheng, P. Y., Decena, J. A., Cheng, Y., Wu, D. L., and Dwyer, G., "Fractal properties in foetal breathing dynamics," *American Journal of Physiology* 263: 141-147, 1992.

[149] Tebbens, S. F., Burroughs, S. M., and Nelson, E. E., "Wavelet analysis of shoreline change on the Outer Banks of North Carolina: An example of complexity in the marine sciences," *Proceedings of the National Academy of Sciences* 99 (1): 2554-2560, 2002.

[150] The BIDMC Congestive Heart Failure Database, "http://www.physionet.org/physiobank/database/chfdb/," Accessed on December 1, 2006.

[151] Thuillard, M., "Wavelets in Soft Computing," *World Scientific Series in Robotics and Intelligent Systems* 25, World Scientific Publishing Co., Singapore, 2001.

[152] Thurner, S., Feurstein, M. C., and Teich, M. C., "Multiresolution Wavelet Analysis of Heartbeat Intervals Discriminates Healthy Patients from Those with Cardiac Pathology," *Physical Review Letters* 80: 1544-1547, 1998.

[153] Turcotte, D. L., and Rundle, J. B., "Self-organized complexity in the physical, biological, and social sciences," *Proceedings of the National Academy of Sciences* 99 (1): 2463-2465, 2002.

[154] Two-sample F-test, "http://www.vias.org/tmdatanaleng/cc_test_2sample_ftest.html," Accessed on January 25, 2007.

[155] Van den Berg, J., Kaymak, U., and Van den Bergh, W. M., "Financial markets analysis by using a probabilistic fuzzy modelling approach," *International Journal of Approximate Reasoning* 35: 291-305, 2004.

[156] Visualization of DWT-based CHF and Normal Classification, "http://www.saifahmad.com/dwt_medical.bmp," Accessed on January 25, 2007.

[157] Visualization of MODWT-based CHF and Normal Classification, "http://www.saifahmad.com/modwt_medical.bmp," Accessed on January 25, 2007.

[158] Wang, L. X., and Mendel, J., "Generating fuzzy rules by learning from examples," *IEEE Transactions on Systems, Man, and Cybernetics* 22: 1414-1427, 1992.

[159] Wang, X., and Tugnait, J. K., "A Bit-Map-Assisted Dynamic Queue Protocol for Multiaccess Wireless Networks with Multiple Packet Reception," *IEEE Transactions on Signal Processing* 51 (8): 2068-2081, 2003.

[160] Whitcher, B., Percival, B., and Guttorp, P., "Multiscale Detection and Location of Multiple Variance Changes in the Presence of Long Memory," *Journal of Statistical Computation and Simulation* 68: 65-88, 2000.

[161] Wichern, D. W., Miller, R. B., and Hsu, D. A., "Changes of Variance in First-Order Autoregressive Time Series Models-With an Application," *Applied Statistics* 25 (3): 248-256, 1976.

[162] WMTSA Wavelet Toolkit for MATLAB, "http://www.atmos.washington.edu/~wmtsa/," Accessed on November 22, 2005.

[163] WOODSON WAVE REPORT, "http://www.woodsonwave.com/," Accessed on October 9, 2005.

[164] Yahoo® Finance, "http://finance.yahoo.com," Accessed on October 7, 2005.

[165] Zhang, B., Coggins, R., Jabri, M. A., Dersch, D., and Flower, B., "Multiresolution Forecasting for Futures Trading Using Wavelet Decompositions," *Transactions on Neural Networks* 12 (4): 765-775, 2001.

[166] Zhang, B., Wang, L., and Wang, J., "The Research of Fuzzy Modelling Using Multiresolution Analysis," *In Proceedings of IEEE International Conference on Fuzzy Systems* 378-383, 2003.

Abbreviations

ACVS	autocovariance sequence
AI	artificial intelligence
ANOVA	analysis of variance
AR	autoregressive / autoregression
ARCH	autoregressive conditional heteroskedasticity
ARIMA	autoregressive integrated moving average
ARMA	autoregressive moving average
CAPM	capital asset pricing model
CHF	congestive heart failure
CV	cumulative variance
CWT	continuous wavelet transform
DFT	discrete Fourier transform
DJIA	Dow Jones industrial average
DWT	discrete wavelet transform
ECG	electrocardiogram
EEG	electroencephalogram
FFT	fast Fourier transform
FTSE	financial times stock exchange
GARCH	generalized autoregressive conditional heteroskedasticity
GE	general electric
GM	general motors
IBM	international business machines
KA	knowledge acquisition
LNCS	lecture notes in computer science
MA	moving average
MLP	multilayer perceptron
MMG	mechanomyographic
MODWT	maximal overlap discrete wavelet transform
MRA	multiresolution analysis
NASDAQ	national association of securities dealers automated quotation
NCSS	normalized cumulative sum of squares index
NLG	natural language generation
NYSE	New York stock exchange
OLS	ordinary least square
PPL	pure power law
RBF	radial basis function
S&P	standard & poor
SARIMA	seasonal autoregressive moving average
SDF	spectral density function
SML	security market line
STFT	short time Fourier transform
VaR	value at risk
IR	information retrieval
SNR	signal to noise ratio
RMS	root mean square
dB	decibels

Appendix A – DJIA Risk-Return Tradeoffs

Table 1a: Risk-return tradeoffs in the DJIA for raw data.

Portfolio	Company	Raw			
		Beta	**Return**	**Portfolio *Beta***	**Portfolio Return**
P_1	PFIZER INC	0.77	10.62	0.80	9.17
	MCDONALDS CP	0.82	9.18		
	EXXON MOBIL CP	0.82	7.72		
P_2	COCA COLA CO THE	0.83	7.38	0.85	7.42
	JOHNSON AND JOHNS DC	0.85	6.89		
	VERIZON COMMUN	0.88	7.99		
P_3	ALTRIA GROUP INC	0.88	8.87	0.88	10.17
	MERCK CO INC	0.89	11.07		
	WALT DISNEY-DISNEY C	0.89	10.58		
P_4	PROCTER GAMBLE CO	0.90	7.06	0.94	8.69
	SBC COMMUNICATIONS	0.92	8.69		
	BOEING CO	0.99	10.32		
P_5	3M COMPANY	0.99	8.98	1.00	9.43
	AMER INTL GROUP INC	1.00	9.13		
	GEN MOTORS	1.00	10.18		
P_6	INTL BUSINESS MACH	1.01	7.11	1.02	9.36
	DU PONT E I DE NEM	1.01	7.62		
	ALCOA INC	1.04	13.36		
P_7	CATERPILLAR INC	1.04	11.41	1.05	9.22
	MICROSOFT CP	1.05	8.00		
	WAL MART STORES	1.05	8.25		
P_8	UNITED TECH	1.07	8.51	1.10	8.23
	GEN ELECTRIC CO	1.11	8.07		
	CITIGROUP INC	1.14	8.10		
P_9	AMER EXPRESS INC	1.15	6.66	1.17	9.62
	HONEYWELL INTL INC	1.16	10.39		
	HEWLETT PACKARD CO	1.19	11.82		
P_{10}	JP MORGAN CHASE CO	1.19	8.22	1.20	10.72
	INTEL CP	1.20	14.67		
	HOME DEPOT INC	1.20	9.27		

Table 1b: Risk-return tradeoffs in the DJIA for scale 1 (2 – 4 days).

Portfolio	Company	Scale 1			
		Beta	**Return**	**Portfolio *Beta***	**Portfolio Return**
P_1	PFIZER INC	0.69	10.62	0.76	8.41
	JOHNSON AND JOHNS DC	0.78	6.89		
	EXXON MOBIL CP	0.80	7.72		
P_2	WALT DISNEY-DISNEY C	0.81	10.58	0.82	9.05
	MCDONALDS CP	0.82	9.18		
	COCA COLA CO THE	0.84	7.38		
P_3	ALTRIA GROUP INC	0.86	8.87	0.87	9.31
	MERCK CO INC	0.87	11.07		
	VERIZON COMMUN	0.88	7.99		
P_4	SBC COMMUNICATIONS	0.88	8.69	0.92	8.69
	PROCTER GAMBLE CO	0.90	7.06		
	BOEING CO	0.97	10.32		
P_5	AMER INTL GROUP INC	1.00	9.13	1.00	8.79
	WAL MART STORES	1.00	8.25		
	3M COMPANY	1.01	8.98		
P_6	DU PONT E I DE NEM	1.01	7.62	1.03	10.79
	ALCOA INC	1.03	13.36		
	CATERPILLAR INC	1.04	11.41		
P_7	GEN MOTORS	1.05	10.18	1.05	8.60
	UNITED TECH	1.05	8.51		
	INTL BUSINESS MACH	1.06	7.11		
P_8	AMER EXPRESS INC	1.10	6.66	1.11	7.58
	GEN ELECTRIC CO	1.10	8.07		
	MICROSOFT CP	1.12	8.00		
P_9	HONEYWELL INTL INC	1.14	10.39	1.16	9.25
	CITIGROUP INC	1.15	8.10		
	HOME DEPOT INC	1.18	9.27		
P_{10}	HEWLETT PACKARD CO	1.22	11.82	1.25	11.57
	JP MORGAN CHASE CO	1.23	8.22		
	INTEL CP	1.31	14.67		

Table 1c: Risk-return tradeoffs in the DJIA for scale 2 (4 – 8 days).

Portfolio	Company	Scale 2			
		Beta	**Return**	**Portfolio *Beta***	**Portfolio Return**
P_1	PFIZER INC	0.78	10.62	0.82	9.80
	EXXON MOBIL CP	0.83	7.72		
	MERCK CO INC	0.84	11.07		
P_2	MCDONALDS CP	0.85	9.18	0.86	8.48
	ALTRIA GROUP INC	0.86	8.87		
	COCA COLA CO THE	0.89	7.38		
P_3	VERIZON COMMUN	0.89	7.99	0.90	7.31
	JOHNSON AND JOHNS DC	0.89	6.89		
	PROCTER GAMBLE CO	0.91	7.06		
P_4	GEN MOTORS	0.92	10.18	0.93	9.43
	AMER INTL GROUP INC	0.92	9.13		
	3M COMPANY	0.97	8.98		
P_5	BOEING CO	0.98	10.32	1.00	8.87
	SBC COMMUNICATIONS	1.00	8.69		
	DU PONT E I DE NEM	1.01	7.62		
P_6	WALT DISNEY-DISNEY C	1.02	10.58	1.02	8.73
	UNITED TECH	1.02	8.51		
	INTL BUSINESS MACH	1.03	7.11		
P_7	MICROSOFT CP	1.07	8.00	1.08	9.16
	CATERPILLAR INC	1.07	11.41		
	GEN ELECTRIC CO	1.09	8.07		
P_8	ALCOA INC	1.09	13.36	1.12	11.14
	HEWLETT PACKARD CO	1.12	11.82		
	WAL MART STORES	1.15	8.25		
P_9	HONEYWELL INTL INC	1.15	10.39	1.16	9.25
	CITIGROUP INC	1.15	8.10		
	HOME DEPOT INC	1.18	9.27		
P_{10}	JP MORGAN CHASE CO	1.18	8.22	1.19	9.85
	INTEL CP	1.19	14.67		
	AMER EXPRESS INC	1.21	6.66		

Table 1d: Risk-return tradeoffs in the DJIA for scale 3 (8 – 16 days).

Portfolio	Company	Scale 3			
		Beta	**Return**	**Portfolio *Beta***	**Portfolio Return**
P_1	MCDONALDS CP	0.76	9.18	0.79	8.48
	COCA COLA CO THE	0.78	7.38		
	ALTRIA GROUP INC	0.82	8.87		
P_2	PFIZER INC	0.85	10.62	0.85	8.78
	EXXON MOBIL CP	0.85	7.72		
	MICROSOFT CP	0.86	8.00		
P_3	VERIZON COMMUN	0.87	7.99	0.89	7.39
	PROCTER GAMBLE CO	0.90	7.06		
	INTL BUSINESS MACH	0.90	7.11		
P_4	BOEING CO	0.94	10.32	0.96	10.93
	MERCK CO INC	0.96	11.07		
	CATERPILLAR INC	0.97	11.41		
P_5	ALCOA INC	0.97	13.36	0.98	9.98
	3M COMPANY	0.98	8.98		
	DU PONT E I DE NEM	0.99	7.62		
P_6	JOHNSON AND JOHNS DC	0.99	6.89	1.02	7.93
	JP MORGAN CHASE CO	1.04	8.22		
	SBC COMMUNICATIONS	1.04	8.69		
P_7	GEN MOTORS	1.05	10.18	1.06	9.96
	WALT DISNEY-DISNEY C	1.05	10.58		
	AMER INTL GROUP INC	1.08	9.13		
P_8	HONEYWELL INTL INC	1.10	10.39	1.10	11.10
	INTEL CP	1.10	14.67		
	WAL MART STORES	1.11	8.25		
P_9	CITIGROUP INC	1.13	8.10	1.15	8.23
	UNITED TECH	1.13	8.51		
	GEN ELECTRIC CO	1.21	8.07		
P_{10}	HEWLETT PACKARD CO	1.23	11.82	1.25	9.25
	AMER EXPRESS INC	1.26	6.66		
	HOME DEPOT INC	1.26	9.27		

Table 1e: Risk-return tradeoffs in the DJIA for scale 4 (16 – 32 days).

Portfolio	Company	Scale 4			
		Beta	**Return**	**Portfolio *Beta***	**Portfolio Return**
P_1	MCDONALDS CP	0.76	9.18	0.79	8.48
	COCA COLA CO THE	0.78	7.38		
	ALTRIA GROUP INC	0.82	8.87		
P_2	PFIZER INC	0.85	10.62	0.85	8.78
	EXXON MOBIL CP	0.85	7.72		
	MICROSOFT CP	0.86	8.00		
P_3	VERIZON COMMUN	0.87	7.99	0.89	7.39
	PROCTER GAMBLE CO	0.90	7.06		
	INTL BUSINESS MACH	0.90	7.11		
P_4	BOEING CO	0.94	10.32	0.96	10.93
	MERCK CO INC	0.96	11.07		
	CATERPILLAR INC	0.97	11.41		
P_5	ALCOA INC	0.97	13.36	0.98	9.98
	3M COMPANY	0.98	8.98		
	DU PONT E I DE NEM	0.99	7.62		
P_6	JOHNSON AND JOHNS DC	0.99	6.89	1.02	7.93
	JP MORGAN CHASE CO	1.04	8.22		
	SBC COMMUNICATIONS	1.04	8.69		
P_7	GEN MOTORS	1.05	10.18	1.06	9.96
	WALT DISNEY-DISNEY C	1.05	10.58		
	AMER INTL GROUP INC	1.08	9.13		
P_8	HONEYWELL INTL INC	1.10	10.39	1.10	11.10
	INTEL CP	1.10	14.67		
	WAL MART STORES	1.11	8.25		
P_9	CITIGROUP INC	1.13	8.10	1.15	8.23
	UNITED TECH	1.13	8.51		
	GEN ELECTRIC CO	1.21	8.07		
P_{10}	HEWLETT PACKARD CO	1.23	11.82	1.25	9.25
	AMER EXPRESS INC	1.26	6.66		
	HOME DEPOT INC	1.26	9.27		

Table 1f: Risk-return tradeoffs in the DJIA for scale 5 (32 – 64 days).

Portfolio	Company	Scale 5			
		Beta	**Return**	**Portfolio *Beta***	**Portfolio Return**
P_1	SBC COMMUNICATIONS	0.50	8.69	0.62	8.31
	WAL MART STORES	0.67	8.25		
	VERIZON COMMUN	0.69	7.99		
P_2	INTL BUSINESS MACH	0.79	7.11	0.81	7.19
	PROCTER GAMBLE CO	0.81	7.06		
	COCA COLA CO THE	0.82	7.38		
P_3	MCDONALDS CP	0.86	9.18	0.88	8.82
	HOME DEPOT INC	0.87	9.27		
	MICROSOFT CP	0.92	8.00		
P_4	WALT DISNEY-DISNEY C	0.94	10.58	0.97	10.26
	MERCK CO INC	0.97	11.07		
	AMER INTL GROUP INC	1.01	9.13		
P_5	EXXON MOBIL CP	1.03	7.72	1.05	8.41
	JOHNSON AND JOHNS DC	1.05	6.89		
	PFIZER INC	1.07	10.62		
P_6	CITIGROUP INC	1.08	8.10	1.12	9.90
	GEN MOTORS	1.13	10.18		
	CATERPILLAR INC	1.17	11.41		
P_7	DU PONT E I DE NEM	1.20	7.62	1.20	8.06
	UNITED TECH	1.20	8.51		
	GEN ELECTRIC CO	1.21	8.07		
P_8	3M COMPANY	1.21	8.98	1.26	11.32
	INTEL CP	1.28	14.67		
	BOEING CO	1.30	10.32		
P_9	ALCOA INC	1.35	13.36	1.36	9.41
	JP MORGAN CHASE CO	1.35	8.22		
	AMER EXPRESS INC	1.38	6.66		
P_{10}	ALTRIA GROUP INC	1.42	8.87	1.54	10.36
	HEWLETT PACKARD CO	1.51	11.82		
	HONEYWELL INTL INC	1.68	10.39		

Table 1g: Risk-return tradeoffs in the DJIA for scale 6 (64 – 128 days).

Portfolio	Company	Scale 6			
		Beta	**Return**	**Portfolio *Beta***	**Portfolio Return**
P_1	SBC COMMUNICATIONS	0.07	8.69	0.11	8.13
	VERIZON COMMUN	0.09	7.99		
	EXXON MOBIL CP	0.16	7.72		
P_2	MICROSOFT CP	0.23	8.00	0.25	8.17
	AMER INTL GROUP INC	0.25	9.13		
	COCA COLA CO THE	0.28	7.38		
P_3	PROCTER GAMBLE CO	0.31	7.06	0.32	7.71
	JOHNSON AND JOHNS DC	0.31	6.89		
	MCDONALDS CP	0.33	9.18		
P_4	WAL MART STORES	0.35	8.25	0.36	9.27
	3M COMPANY	0.35	8.98		
	WALT DISNEY-DISNEY C	0.37	10.58		
P_5	UNITED TECH	0.38	8.51	0.38	8.91
	DU PONT E I DE NEM	0.38	7.62		
	PFIZER INC	0.39	10.62		
P_6	BOEING CO	0.40	10.32	0.40	10.20
	ALTRIA GROUP INC	0.40	8.87		
	CATERPILLAR INC	0.41	11.41		
P_7	MERCK CO INC	0.42	11.07	0.43	10.84
	CITIGROUP INC	0.43	8.10		
	ALCOA INC	0.45	13.36		
P_8	INTL BUSINESS MACH	0.47	7.11	0.49	8.52
	GEN ELECTRIC CO	0.47	8.07		
	HONEYWELL INTL INC	0.53	10.39		
P_9	AMER EXPRESS INC	0.55	6.66	0.57	10.20
	INTEL CP	0.58	14.67		
	HOME DEPOT INC	0.58	9.27		
P_{10}	HEWLETT PACKARD CO	0.61	11.82	0.63	10.07
	GEN MOTORS	0.62	10.18		
	JP MORGAN CHASE CO	0.67	8.22		

Appendix B – System Evaluation

Here we show some excerpts of the evaluation undertaken to assess our system. A complete version of the evaluation document can be downloaded from:

www.saifahmad.com/evaluation.pdf

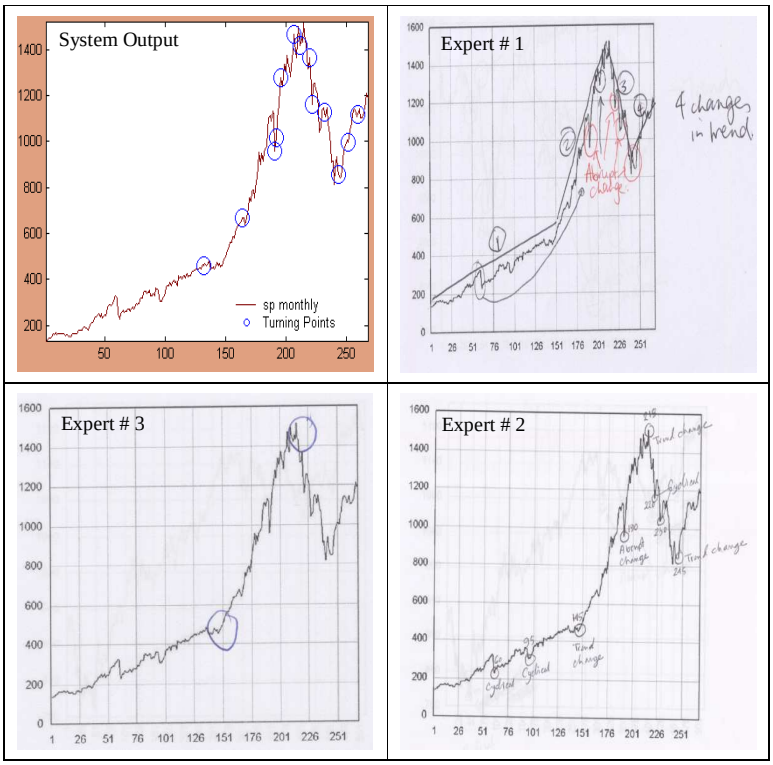


Figure 2a System output and expert opinion for turning points in the S&P index.

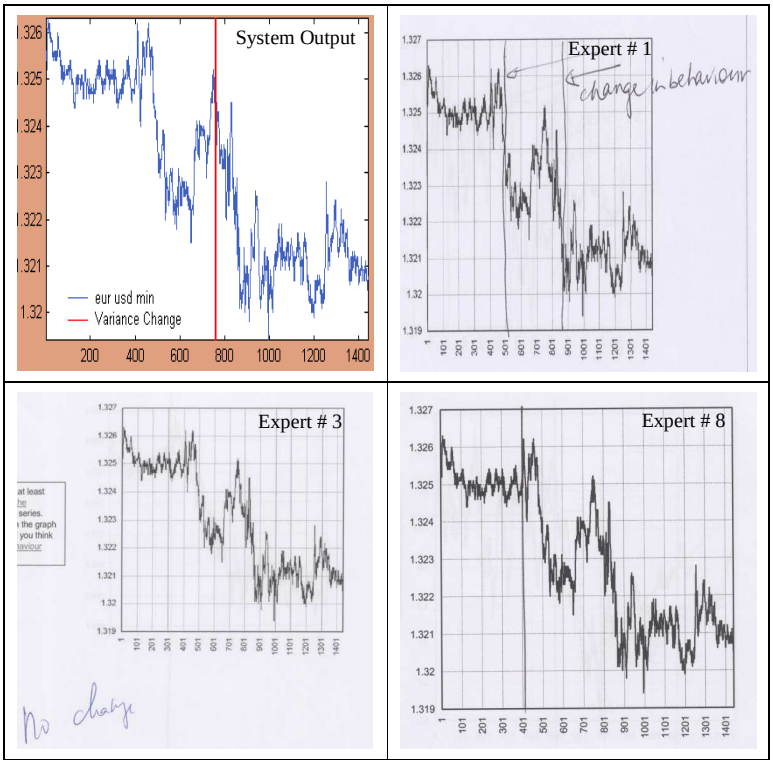


Figure 2b System output and expert opinion for variance change in the EUR-USD exchange rate series.

Table 2a Summary of answers for Q1 to Q3.

Expert #	Name and Details	Turning Points		
		Q1	**Q2**	**Q3**
1	**Dr. Jo Evans**, Lecturer, Economics, University of Surrey	60, 148, 191, 204, 222, 229, 244	47, 54, 150, 187, 225, 308, 349, 406, 443, 456	37, 46, 56, 65, 179, 207
2	**Dr. Shah Jalaal Sarker**, Department of Public Health Sciences, King's College London	60, 95, 145, 190, 218, 220, 230, 245	50, 305, 345, 405, 455	61, 160
3	**Prof. Philip Hans B. F. Franses**, Econometric Institute, & Department of Marketing & Organization, Erasmus University Rotterdam	144, 213	None	None
4	**Prof. Martin Crowder**, Department of Maths / Stats, Imperial College	148, 213, 243	47, 115, 223, 284, 407, 456	57, 65, 98, 165, 179
5	**Prof. John Nankervis**, Department of Economics, University of Essex	60, 63, 98, 148, 190, 191, 217, 244	47, 54, 97, 115, 151, 187, 206, 225, 284, 308, 317, 349, 368, 405, 456	46, 56, 65, 97, 109, 165, 179, 204, 207, 220, 232
6	**Mr. Florian Weiners**, Trader, JRC, Berlin	64, 98, 148, 191, 204, 217, 225, 234, 244	47, 60, 96, 126, 150, 187, 206 225, 284, 321, 375, 407, 456, 509	9, 36, 46, 61, 65, 78, 91, 107, 119, 138, 148, 165, 179, 199, 207, 229, 238
7	**Dr. A. Heen**, Researcher, London Hedge Fund	149, 220, 235	50, 260	180
8	**Dr. Steve Pollock**, Reader, Economics, Queen Mary University of London	142, 213, 245	47, 61, 96, 115, 151, 187, 225, 268, 275, 307, 345, 407, 443, 456	165, 179
9	**System Output**	132, 164, 191, 192, 196, 207, 207, 212, 220, 222, 232, 244, 252, 260	8, 15, 36, 44, 46, 55, 68, 96, 132, 260, 372, 389, 404, 484	36, 44, 46, 68, 96, 100, 132, 174, 175, 176, 178, 180, 196, 197

Table 2b Summary of answers for Q4 to Q7.

Expert #	Name and Details	Trend	Structural Change		
		Q4	Q5	Q6	Q7
1	**Dr. Jo Evans**, Lecturer, Economics, UniS	1 to 272 uptrend; 272 to 519 uptrend	501, 862	710	556
2	**Dr. Shah Jalaal Sarker**, Department of Public Health Sciences, King's College London	1 to 255 uptrend; 255 to 519 uptrend	501, 868	569, 801	570
3	**Prof. Philip Hans B. F. Franses**, Econometric Institute, & Department of Marketing & Organization, Erasmus University Rotterdam	1 to 242 uptrend; 242 to 519 uptrend	No change	412	No change
4	**Prof. Martin Crowder**, Department of Maths / Stats, Imperial College	1 to 272 uptrend 272 to 519 constant	No change	409, 941	557
5	**Prof. John Nankervis**, Department of Economics, University of Essex	1 to 272 uptrend; 272 to 519 uptrend	467, 836	459, 817	557
6	**Mr. Florian Weiners**, Trader, JRC, Berlin	None	501, 836	401, 560	578
7	**Dr. A. Heen**, Researcher, London Hedge Fund	1 to 251 uptrend; 251 to 519 constant	850	500	550
8	**Dr. Steve Pollock**, Reader, Economics, Queen Mary University of London	1 to 50 downtrend; 50 to 272 uptrend; 272 to 519 uptrend	404	438, 849	498, 689, 1004
9	**System Output**	1 to 519 uptrend	761	531	691

Table 2c Rating of system output for Q8 to Q12.

Expert #	Name & Details	Trend	Var. Chng.	Turning Points		All Three	Avg. Score
		Q8	Q9	Q10	Q11	Q12	
1	**Dr. Jo Evans** Lecturer, Economics UniS	4.0	3.0	-2.0	-2.0	3.0	**1.2**
2	**Dr. Shah Jalaal Sarker** Department of Public Health Sciences King's College London	4.0	4.0	1.0	3.0	3.0	**3.0**
3	**Prof. Philip Hans B.F. Franses** Econometric Institute, & Department of Marketing & Organization Erasmus University Rotterdam	-5.0	-2.0	-5.0	-1.0	1.0	**-2.4**
4	**Prof. Martin Crowder** Department of Maths / Stats Imperial College	-2.0	-3.0	-5.0	-5.0	-5.0	**-4.0**
5	**Prof. John Nankervis** Department of Economics Univ. of Essex	-3.0	-3.0	1.0	2.0	-2.0	**-1.0**
6	**Mr. Florian Wieners** Trader JRC Berlin	-4.0	-5.0	3.0	4.0	3.0	**0.2**
7	**Dr. A Heen** Researcher London Hedge Fund	2.0	-2.0	0.0	2.0	1.0	**0.6**
8	**Dr. Steve Pollock** Reader, Economics Queen Mary University of London	0.0	5.0	-3.0	-5.0	-3.0	**-1.2**

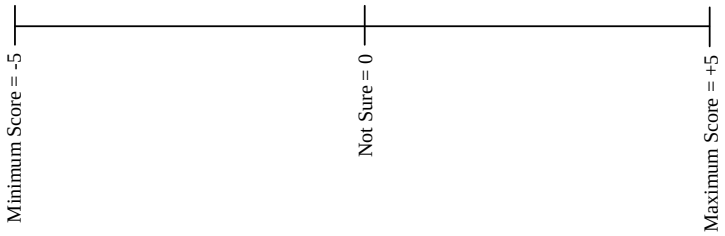


Table 2d Reply from Professor Granger.

From: Michael Bacci [mailto:mbacci@ucsd.edu]
Sent: Wed 02/03/2005 22:08
To: Ahmad S Mr (PG/R - Computing)
Subject: REPLY FROM PROF. GRANGER

The following is from Prof. Sir Clive W.J. Granger:

Dear Saif,

I am just preparing to go on a long trip and do not have time to answer your questions.

I do not see the point in discussing turning points for a series that would well be a random walk, particularly for data measured every minute.

Yours sincerely,

Sir Clive W.J. Granger
Professor Emeritus
Nobel Laureate

Professor Clive Granger was awarded the 2003 Nobel Prize in Economics "for methods of analyzing economic time series with common trends (cointegration)".

References

Granger, C. W. J., and Engle, R., "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica* 55: 251-76, 1987.

Table 2e Reply from Professor Beck.

From: Travis Beck [mailto:tbeck@unlserve.unl.edu]
Sent: Mon 10/10/2005 16:43
To: Ahmad S Mr (PG/R - Computing)
Subject: Re: URGENT: Beck's Reply

Hi Saif. I have been thinking about the analyses that you performed on our MMG data, and we appreciate all of the work that you have done. Generally speaking, MMG signals are considered to be random, stationary signals. However, your analyses indicated that for the signals I sent you, there were several different changes in variance (i.e. indicating that variance was not constant, and, therefore, the signals were nonstationary). Thus, your analysis may be potentially useful for identifying nonstationary versus stationary signals. However, because MMG signals are random, I don't think that there would be much use in trying to predict how the signal will "behave" (i.e. a forecast). Nevertheless, we appreciate the work you have done, and hope that this information is useful to you.

Travis

Professor Travis Beck is a faculty in the Department of Health & Human Performance, Center for Youth Fitness and Sports Research, University of Nebraska, Lincoln, USA.

References

Beck T. W., Housh T. J., Johnson G. O., Cramer J. T., Weir J. P., Coburn J. W., Malek M. H., "Comparison of the fast Fourier transform and continuous wavelet transform for examining mechanomyographic (MMG) frequency versus eccentric torque relationships," *Journal of Neuroscience Methods*, 2005.

Table 2f Reply from Steve Pincus

From: Steve Pincus [mailto:stevepincus@alum.mit.edu]
Sent: Thu 03/02/2005 22:06
To: Ahmad S Mr (PG/R - Computing)
Subject: Re: Turning Points

Dear Mr. Ahmad,

Thanks for keeping me in the loop. I took a quick glance at your data sets (time-series), and agree that these are worth evaluating. However, my view is that in many instances, there may be both subtle changes in dynamics, in addition to one or several key visually obvious features. Anyhow, I do intend to take a look at these test series, but to be candid, a serious look (here) is a lot more than a 5-10 min evaluation, and my demands are very extensive presently. So it may be a while before you hear back from me.

The questionnaire design looks to be quite reasonable, and your queries are clear.

My one primary concern is an insidious study bias if you are consistently asking readers for 5-10 min total time for this study. You will de facto be excluding more in-depth statistically knowledgeable readers, a subgroup whose input you definitely want to encourage, be they classical ARMA modelers or otherwise. I strongly suggest that you find a few cooperative econometrics types, a like-minded yet distinct group of applications-oriented statisticians, give them the time-series and tasks, and ask for general comments (do not `lead the witness` so much); and most especially, do not impose or suggest any time limit.

Otherwise, good luck, and do keep me apprised of further progress.

Best regards,

Steve Pincus

Steve Pincus, is a free-lance mathematician based in Guilford, Connecticut, USA. He has created a method for measuring a signal's "entropy," or disorder. He has co-authored a paper in the Proceedings of the National Academy of Sciences (PNAS) with Rudolf E. Kalman who is famous for developing, the so-called "Kalman filter".

References

Pincus, S., and Kalman, R. E., "Irregularity, volatility, risk, and financial market time series," *Proceedings of the National Academy of Sciences* 101 (38): 13709-13714, 2004.

Table 2g Reply from Professor Burroughs.

From: Stephen Burroughs [mailto:sburroughs@ut.edu]
Sent: Thu 03/11/2005 14:39
To: Ahmad S Mr (PG/R - Computing)
Cc: Sarah Tebbens
Subject: Re: NC Shoreline Changes Data

Dear Saif,

I have forwarded your request to Dr. Sarah Tebbens, the lead author of the paper you reference in your email. She is now at Wright State University. Hopefully, she will be able to direct you to data sets you can use for your analysis. We find wavelet analysis to be very useful for understanding complex data sets and we encourage you in your efforts to demonstrate its wide applicability.

Best regards,

Steve

Stephen Burroughs is Associate Professor of Physics at the University of Tampa, USA. He is c-author of a paper in the Proceedings of the National Academy of Sciences (PNAS).

References

Tebbens, S. F., **Burroughs, S**. M., Nelson, E. E., "Wavelet analysis of shoreline change on the Outer Banks of North Carolina: An example of complexity in the marine sciences," *Proceedings of the National Academy of Sciences* 99 (1): 2554-2560, 2002.

Appendix C – Research Activities

PUBLICATIONS

A. Popoola, **S. Ahmad**, and K. Ahmad**, "Multiscale Wavelet Preprocessing for Fuzzy Systems"**, *Proceedings of ICSC Congress on Computational Intelligence Methods and Applications (CIMA 05)*, Istanbul, Turkey, December 2005 (to appear).

A. Popoola, **S. Ahmad**, and K. Ahmad, **"A Fuzzy-Wavelet Method for Analyzing Non-Stationary Time Series"**, *Proceedings of the 5th International Conference on Recent Advances in Soft Computing (RASC)*, December 16-18, 2004, Nottingham, UK.

S. Ahmad, T. Taskaya, and K. Ahmad, **"Summarizing Time Series: Learning Patterns in 'Volatile' Series"**, In Yang Z.R., Everson R., Yin H. (Eds.), *Proceedings of the 5th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL)*, August 25-27, 2004, Exeter, UK, volume 3177 of Lecture Notes in Computer Science (**LNCS**), Heidelberg: © Springer-Verlag, Germany.

S. Ahmad, P. Oliviera, and K. Ahmad, **"Summarization of Multimodal Information"**, *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, May 26-28, 2004, Lisbon, Portugal.

K. Ahmad, T. Taskaya, D. Cheng, L. Gillam, **S. Ahmad**, H. Traboulsi, and J. Nankervis, **"FinGrid: Financial Information Grid – an ESRC e-Social Science Pilot Project for the Financial Markets"**, *Proceedings of UK e-Science All Hands Meeting*, August 31-September 3, 2004, Nottingham, UK.

K. Ahmad. T. Taskaya, D. Cheng, P. Manomaisupat, **S. Ahmad**, L. Gillam, H. Trablousi and M. Casey, **"Fundamental Data to SATISFI the Chartist"**, *The Technical Analyst Magazine*, April 2004, UK.

K. Ahmad, D. Cheng, T. Taskaya, **S. Ahmad**, L. Gillam, P. Manomaisupat, H. Traboulsi and A. Hippisley, **"The mood of the (financial) markets: In a corpus of words and of pictures"**, *Proceedings of the Corpus Linguistics Conference*, March 28-31, 2003, Lancaster University, UK.

L. Gillam, K. Ahmad, **S. Ahmad**, M. Casey, D. Cheng, T. Taskaya, P. Oliveira and P. Manomaisupat, **"Economic News and Stock Market Correlation: A Study of the UK Market"**, *Proceedings of the Workshop on Making Money in the Financial Services Industry, at the 6th International Conference on Terminology and Knowledge Engineering (TKE)*, August 28-30, 2002, Nancy, France.

OTHER REPORTS AND PRESENTATIONS

S. Ahmad, and K. Ahmad, **"An Automatic Analysis of Volatile Financial Time Series"**, *Poster Presentation at the House of Commons*, London, under the auspices of Presentations by Britain's Younger Scientists, Engineers and Technologists at the House of Commons, March 15, 2004.

K. Ahmad, and **S. Ahmad**, **"Brand Protection: Comparing the Provenance of two Time Series"**, *A study performed for Reuters PLC (UniS Technical Report)*, London, November 21, 2003.

RESEARCH PROJECTS

GIDA - Generic Information-based Decision Assistant: GIDA was a EU-sponsored R&D project between the University of Surrey (UK), JRC (Germany), Finsoft (UK) and Ibermatica (Spain). The main aim of the project was to develop a generic execution platform by integrating predictive models for financial analysis with advanced text processing techniques for information retrieval, information extraction and language processing. Performed an evaluation to compare the automatic detection of financial turning points using wavelets with those identified by expert traders. The turning point detection accuracy achieved by our system was about sixty percent as compared to humans. This laid the foundation for using the discrete wavelet transform (DWT) to detect turning points in financial time series data.

FINGRID - Financial Information Grid: FINGRID was a 12-month project that began on 1 October 2003 and was sponsored by the ESRC under their Pilot Projects in e-Social Science call (e-Science). It aims to develop a Grid-based demonstrator for sharing and analysing real-time financial data streams, time series of financial returns and financial news. Proposed a technique for detecting and locating cycles on a volatile and nonstationary time series using a combination of the DWT and the fast Fourier transform (FFT). The results of this technique were encouraging.

PROFESSIONAL ACTIVITIES

Presentations at Professional Meetings

2004 International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 04), Exeter, UK (presenter).

2004 PhD Conference, Department of Computing, University of Surrey, UK (presenter).

2003 PhD Conference, Department of Computing, University of Surrey, UK (presenter).

2002 International Conference on Terminology and Knowledge Engineering (TKE 02), Nancy, France (presenter).

Other Presentations

House of Commons (British Parliament), London, UK (2004).
Reuters PLC, London, UK (2003).

HONOURS, SCHOLARSHIPS AND AWARDS

UniS Department of Computing Merit Scholarship, January 2002 – December 2005.
Winner (2nd prize), PhD Conference, Department of Computing, UniS, 2004.
Winner (2nd prize), PhD Conference, Department of Computing, UniS, 2003.