SELF-SIMILAR BASED TIME SERIES ANALYSIS AND PREDICTION

by

Ruoqiu Wang

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Mechanical and Industrial Engineering
University of Toronto

© Copyright 2014 by Ruoqiu Wang
Abstract
Self-Similar Based Time Series Analysis and Prediction

Ruoqiu Wang
Master of Applied Science
Graduate Department of Mechanical and Industrial Engineering
University of Toronto
2014

Fractals have been observed in many natural phenomena, and self-similarity is the most important statistical property of fractals. In the literature, the concepts of self-similarity and long-range-dependence (LRD), the increment process of a self-similar signal when the Hurst parameter is between 0.5 and 1, are often confused. On the other hand, the forecasting of many real-life signals exhibiting these properties is useful yet remains challenging. The objective of this thesis is to provide the readers with a clear and detailed explanation of the relevant concepts, and then to compare forecasting models (including FARIMA, HAR-RV, wavelet-based and average-VAR) for self-similar and LRD signals via real-life data (arterial blood pressure signals and volatility indexes). Numerical studies show that the four models perform similarly, while the FARIMA model is not recommended due to its time-consuming computation; and the performance of detecting large decrease is more accurate than that of detecting large increase.
Acknowledgements

First and foremost, I would like to express my deepest gratitude to Professor Chi-Guhn Lee for his great supervision during my two-year master’s study. I also want to thank him for his kind offer so that I can come to this beautiful city and study in this honored school that I have dreamed of for years. When I felt stuck in the research process, he would always encourage me and provide me with useful advice. I also want to thank him for his time and patience in reviewing this thesis, and providing useful writing suggestions. His superior guidance and knowledge, along with his motivation and patience, have become the most important impetus during my master’s study, without which this thesis would not have been possible.

In addition, I want to express my appreciation to my committee members, Prof. Timothy C. Y. Chan and Prof. Roy H. Kwon, for their time and patience to kindly review my thesis and provide me with insightful feedback.

I also want to thank my lab mates, Luohao Tang, Zhengqin Zeng and Jue Wang, from whom I can get suggestions and comments that are beneficial to me.

Last but not least, I would like to thank my parents Fagang and Yanli for their love and lifelong support, and my boyfriend Yuanyang for his help and encouragement, who is always accompanying me and help me out when I felt down and discouraged.
Table of Contents

Chapter 1 Introduction

1.1 Fractal and self-similarity ................................................................. 1
1.2 History of fractals and self-similar processes ........................................ 3
1.3 Research objectives and contributions .................................................. 3
1.4 Thesis organization structure ................................................................ 4

Chapter 2 Overview on Self-Similarity and Long-Range-Dependence (LRD)

2.1 Fractal dimension ...................................................................................... 6
2.2 Hurst parameter and long-range-dependence (LRD) .................................. 8
  2.2.1 Hurst parameter and R/S analysis ....................................................... 8
  2.2.2 Long-range-dependence (LRD) .......................................................... 9
2.3 Self-similar processes ............................................................................. 10
  2.3.1 Introduction to self-similar processes and classification ....................... 10
  2.3.2 Theory and properties ...................................................................... 11
  2.3.3 Self-similarity and long-range-dependence (LRD) ............................... 12
2.4 Fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) .... 13
2.5 Summary ................................................................................................. 15

Chapter 3 Literature Review

3.1 Time series analysis .................................................................................. 17
  3.1.1 From ARIMA to FARIMA .................................................................. 17
  3.1.2 Modeling realized volatility ................................................................. 18
3.2 Self-similarity related research ................................................................ 19
  3.2.1 Estimation of the self-similar parameter $H$ ........................................ 19
  3.2.2 Fractional Brownian motion .............................................................. 21
  3.2.3 Fractals in human physiology ........................................................... 21

Chapter 4 Algorithms and Models

4.1 FARIMA model ....................................................................................... 23
  4.1.1 Definition .......................................................................................... 23
  4.1.2 Parameter estimation ....................................................................... 24
List of Tables

Table 2.1 Differences between self-similar and LRD processes
Table 4.1 Summary of the models and their time series satisfying conditions
Table 5.1 Volatility indexes
Table 5.2 The Hurst parameters for the analyzed signals
Table 5.3 MSEs for ABP5 for different combinations
Table 5.4 MSEs for ABP9 for different combinations
Table 5.5 MSEs for ABP13 for different combinations
Table 5.6 An illustrative diagram of how TPR and FPR are calculated
Table 5.7 Error metrics and computation time for ABP5 (H=0.9437)
Table 5.8 Error metrics and computation time for ABP13 (H=0.9069)
Table 5.9 Error metrics and computation time for VIX (H=0.9619)
Table 5.10 Percentage of increases followed by a large increase and decreases followed by a large decrease
List of Figures

Fig. 1.1 Fern
Fig. 1.2 Snowflake
Fig. 1.3 The Cantor set
Fig. 1.4 The Koch curve
Fig. 1.5 The arterial blood pressure signal
Fig. 2.1 British coastline
Fig. 2.2 Comparison of the ACFs for short-memory and long-memory signals
Fig. 2.3 Original time series and its aggregated series
Fig. 2.4 Relationship between self-similar processes and processes with LRD
Fig. 2.5 fBm with different Hurst parameters
Fig. 2.6 Comparison between self-similar (fBm) and LRD (fGn) processes
Fig. 4.1 An illustrative diagram of the wavelet spaces
Fig. 4.2 An illustrative diagram of the three-scale MRA
Fig. 4.3 An example of the three-scale MRA
Fig. 4.4 Approximation of some Daubechies wavelet functions
Fig. 4.5 Comparison of the ACFs before and after wavelet transform
Fig. 4.6 Comparison of the ACFs using different wavelets
Fig. 4.7 Across-scale prediction algorithm (down-sampling approach)
Fig. 4.8 An illustration of the three-scale average-VAR model
Fig. 5.1 An illustrative diagram of the analyzing procedure
Fig. 5.2 ACFs for ABP5 and VIX
Fig. 5.3 Across-scale prediction algorithm (redundant wavelet transform approach)
Fig. 5.4 Role of the threshold
Fig. 5.5 A typical ROC curve
Fig. 5.6 FARIMA results for ABP5
Fig. 5.7 HAR-RV results for ABP5
Fig. 5.8 Wavelet-based results for ABP5
Fig. 5.9 Average-VAR results for ABP5
Fig. 5.10 The ROC curve for ABP5 – increase detection
Fig. 5.11 The ROC curve for VIX – increase detection
Fig. 5.12 The ROC curve for ABP5 – decrease detection
Fig. 5.13 The ROC curve for VIX – decrease detection
List of Appendices

Appendix A-1: Numerical studies for other signals
Chapter 1 Introduction

1.1 Fractal and self-similarity

A fractal is referred to an object consisting of small parts that are similar to the whole, and each small part of the object is replicating the whole structure. When zoomed in at finer and finer scales, the same pattern will be found reappearing such that it is almost impossible to know exactly what scale you are staring at [1]. This property is also called scale-invariance, the most important feature of fractals, which is also termed as self-similarity [7]. This feature allows fractals to be analyzed in a mathematical way, making it more practical when applied to a variety of disciplines.

Scientists and researchers have observed fractals everywhere on the earth. Nature is made of fractals, such as ferns, snowflakes, coastlines, etc. [2] In human bodies, many tissues and organs also exhibit the fractal property, including lungs, atrium, etc. [3] In addition, there are fractals generated in a mathematical manner by recursion, such as the Cantor set, the Koch Curve, the Sierpinski Gasket, etc. [1] Fractals are not only limited to geometry, but can also be used to describe the property of a time series. Some typical examples of fractal time series are physiological signals, internet traffic, and financial time series. Figures 1.1-1.5 show some examples of common fractals.

Fig. 1.1 Fern [4]  Fig. 1.2 Snowflake [5]

Fig. 1.3 The Cantor set [6]
For some of the fractals listed above, not exactly the same pattern reappears; instead, they remain the same statistical properties within various degrees of magnification [1], such as coastlines and time series.
This is also a case of scale-invariance. A more rigorous definition of fractals will be given in the next chapter.

1.2 History of fractals and self-similar processes

The fractal phenomena have been observed long before the term fractal is put forward. When scientists attempted to measure complex patterns such as the areas of Aegean islands, the length of British coastline, etc., they often found irregular power laws appearing [8]. These power laws as well as the dimension of these irregular shapes are different from the geometry they usually dealt with.

Benoit Mandelbrot first coined the term fractal in his French book Les Objets Fractals: Forme, Hasard et Dimension in 1975, and later published the famous English edition (revised): The Fractal Geometry of Nature. In the book, he stated that the nature is not simply made up of regular geometry, but some kind of geometry that is more complex and erratic [2]. He termed this new geometry fractal. Common objects in nature, including clouds, mountains, coastlines, the paths that lightning travels, etc., all belong to this new concept. These objects can be described using recursive equations, where similar elements repeat themselves over and over again to form an irregular shape [2].

Self-similarity and non-integer-valued dimension are two most important properties of fractals [8]. The fractal dimension is a non-integer, and it is closely related to the self-similar parameter called Hurst parameter, which is important in characterizing a self-similar process. Besides, the Hurst parameter is used to determine whether a signal has long memory or not. The long-memory signal is also described as long-range-dependent (LRD, to be defined later), which is usually the increment of a self-similar process. As more and more signals in nature are found to possess the self-similar or LRD feature, their statistical properties are further explored [8, 9] and research related to these topics gained more attention ever since. Readers are referred to Chapter 3 for a detailed literature review.

1.3 Research objectives and contributions

Many real-life signals are self-similar or LRD, such as physiological signals, financial signals, etc. It is useful if we can predict these signals accurately. However, traditional time series analysis and forecasting algorithms can only deal with short-memory signals, and the forecasting of self-similar and LRD signals
Chapter 1. Introduction

is challenging due to the existing strong correlation. In this thesis, we generalize some existing forecasting models and compare their performance. The main objective of this thesis is two-fold:

i. To illustrate the theory and properties of self-similar processes;

ii. To compare several forecasting methods via numerical studies.

Since the literature is not consistent in the relevant concepts, we believe it of great importance to first untangle all these concepts and give the readers a clear view. Afterwards, numerical studies will be presented and several error metrics will be used to compare different methods in predicting self-similar and LRD signals. The signals of interest are arterial blood pressure signals and volatility index time series, which have been observed as long-range-dependent.

The main contributions of this thesis are as follows:

First, our study sheds light on the definitions and concepts relevant to self-similarity, which are quite confusing and poorly understood in the current literature, especially self-similarity and long-range-dependence (LRD). Self-similar processes are by definition non-stationary, while LRD time series are stationary. Clearly distinguishing these two concepts and understanding their properties will facilitate the further application of these two processes.

Second, we study several forecasting algorithms for self-similar and LRD signals, and use real life data to compare them, which is also a highlight of this thesis. The forecasting algorithms studied in the thesis will not be limited to any specific signals, such as financial signals, physiological signals, internet traffic signals, etc., so long as the signal is self-similar or LRD. Although the performances of these algorithms are similar in terms of the error metrics we have applied, this numerical study provides some insight on forecasting self-similar and LRD signals and future research can learn from this.

1.4 Thesis organization structure

The rest of the thesis is organized as follows:

Chapter 2 is an overview on self-similar processes and long-range-dependence, which is the theoretical basis of this thesis. In this chapter, a rigorous definition of self-similar processes is given that makes use of the fractal dimension, and then the Hurst parameter is explained in detail, including its relationship to long-range-dependence (LRD). Afterwards, we illustrate the statistical properties of self-similar processes,
which is the most important part of this chapter. Self-similar processes and LRD processes are then compared: the differences and relationship between them are listed clearly for future application. At the end, we take a closer look at a well-known example, fractional Brownian motion, to help better understand the properties.

Chapter 3 is a literature review. In this chapter, research on time series analysis is first reviewed, together with its generalization to long-memory time series. Then a specific time series, realized volatility, is studied, including its definition, properties, as well as some commonly used models. At the end of this chapter, a literature review on the research areas related to fractal and self-similarity is attached to give an idea on its application and research status.

Chapter 4 presents forecasting algorithms and models to be applied. Several prevailing methods for modelling and predicting self-similar and LRD processes are described in detail, including the FARIMA, HAR-RV and wavelet-based model. To illustrate the wavelet-based model more clearly, basic theories of wavelet transform are first presented, especially its unique scaling property that makes it an ideal tool for analyzing self-similar processes. Afterwards, a new model is proposed, the average-VAR model, which utilizes the scaling property of LRD time series. This new model is both straightforward in theory and simple to use. At the end of this chapter, the conditions required for proper use of the models are summarized for future application.

Chapter 5 is the numerical studies and results. First we describe the data sets, i.e. arterial blood pressure signals and volatility indexes, and how they are obtained. Then we study the properties of these data sets, including their stationarity conditions and Hurst parameters, in order to classify them as self-similar processes or LRD, or neither. The models discussed in Chapter 4 are then refined according to the conditions summarized at the end of the previous chapter to be customized to our data sets. To compare the performances of the models, several error metrics are applied including MSE, MAPE, MAE and PRD (to be defined later), as well as the ROC curve, which has been widely used in the field of threshold selection and performance evaluation. At last, we present the comparison results and discussion on the observations.

Chapter 6 is the conclusions and future work. We first summarize the main results of this thesis, and list the problems yet to be unsolved. Afterwards, several potential research topics are proposed based on this thesis, which can make use of and learn from it.
Chapter 2 Overview on Self-Similarity and Long-Range-Dependence (LRD)

In this chapter, the definitions and main properties of self-similar processes are discussed and some confusing concepts are clarified, especially self-similarity and long-range-dependence. Clearly distinguishing these concepts will help the understanding of the models to be discussed in the next chapter. This chapter provides theoretical basis for the following chapters.

2.1 Fractal dimension

The topological dimension of an object is the most commonly used measure of dimension, for example, a line has a topological dimension of 1 and a surface has a topological dimension of 2. However, fractals have their own dimensional measure. Unlike the topological dimension, always an integer, fractal dimension is usually a non-integer.

Fractal dimension is closely related to the measuring of an object. A line segment is not a fractal. When one is trying to measure its length, it remains the same regardless of the scale used to measure it [10]; no matter it is a 1 m stick or a 10 cm one. This is not the case when you are trying to measure the length of a fractal. Take the British coastline as an example – it shows more and more details as we look closer and...
closer [10]. The length depends greatly on how long the measuring stick is [10, 11]. When a shorter measuring unit is used, a longer measuring result will be obtained. Fig. 2.1 shows a comparison of the lengths of the British coastline and the scale at which they are measured. This is the same case as that of the length of a Koch curve (refer to Fig. 1.4) – the length between any two points on the curve is not fixed, depending on the measuring scale. In fact, the length is infinite, as it is made up of infinite scales [11]. This inconsistency in measuring also occurs in other fractal measuring cases such as the determination of the area of snowflakes, the length of a fractal time series plot, etc.

Suppose $r$ is the measuring unit we use to measure the hypervolume of an object (length for line segments, area for surfaces, etc.), and the corresponding measured hypervolume is $L(r)$, then the fractal dimension $D_F$ can be obtained by Equation (2.1):

$$D_F = \frac{\log(L(r))}{\log(r)}$$  \hspace{1cm} (2.1)

To derive the fractal dimension of a certain object, first several measuring scales $r$ are selected and the corresponding hypervolumes $L(r)$ are measured, then a log-log plot is drawn. The slope of the plot is the fractal dimension of the object.

Above is only one way of defining and finding the fractal dimension of an object. For a more detailed view on fractal dimension, please refer to Hastings and Sugihara [8]. With the help of fractal dimension, a more rigorous way of defining a fractal can be introduced.

**Definition 1**: An object with a fractal dimension that is greater than its topological dimension is called a **fractal** [8].

According to this definition, lines with a fractal dimension of 1 are not fractals since its fractal dimension is equal to its topological dimension. This is also the case for triangles that have a fractal dimension of 2, which is equal to its topological dimension. Mandelbrot found out that the fractal dimension for western British coastline is $D_F = 1.25$ [10], which strictly exceeds its topological dimension of 1, thus it is a fractal. A fractional dimension indicates that the property of a fractal falls in between objects with a higher dimension and with a lower dimension [11]. A fractal time series, for example, has a fractal dimension between 1 and 2, which lies between lines (with a fractal dimension of 1) and surfaces (with a fractal dimension of 2). As $D_F$ approaches 1, the time series is closer to a line on the plane; as $D_F$ approaches 2, the time series is more and more winding and is closer to a surface [8].
2.2 Hurst parameter and long-range-dependence (LRD)

2.2.1 Hurst parameter and R/S analysis

Hurst parameter, or Hurst exponent, was originally introduced by a British hydrologist H. E. Hurst in 1951, when he was studying the storage capacity of reservoirs [12]. Hurst performed a rescaled range analysis (R/S analysis for short) on the yearly water level of Nile River. Let \( \{X_i, i = 1, \ldots, N\} \) denote the water level series, and range \( R \) for the first \( n \) data points can be computed as follows [12]:

\[
R(n) = \max_{1 \leq i \leq n} (X_i + X_{i+1} + \cdots + X_{i+n-1} - i\bar{X}) - \min_{1 \leq i \leq n} (X_i + X_{i+1} + \cdots + X_{i+n-1} - i\bar{X})
\]

Standard deviation \( S \):

\[
S(n) = \left[ \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right]^{1/2}.
\]

Then the Hurst parameter can be derived from the averaged rescaled range R/S over the time span [12]:

\[
E\left[ \frac{R(n)}{S(n)} \right] = c \cdot n^H \quad \text{for some constant } c.
\]  

(2.2)

Hurst found the parameter \( H \) obtained from the Nile river data gives a result different from random processes. For the water level of Nile River, \( R/S \sim n^H \) with \( H \approx 0.77 \) [12], while for random processes \( H \approx 0.50 \). Mandelbrot named this parameter \( H \) after H. E. Hurst, and referred to it as the self-similar parameter [13]. It is a measure of the memory of a time series, and can only take numbers between 0 and 1. The Hurst parameter is also related to the fractal dimension \( D_F \) by the following equation:

\[
H = 2 - D_F
\]

Note that stationary condition must be fulfilled in order to perform R/S analysis. If we perform R/S analysis on a non-stationary time series, we get abnormal results. For instance, if the Hurst parameter we get from R/S analysis turns out to be greater than 1, then this means the time series we are analyzing is non-stationary. In such a case, in order to get the true value of \( H \), we must first difference the non-stationary time series to get a stationary one, and then perform the R/S analysis again.
2.2.2 Long-range-dependence (LRD)

For stationary processes, the corresponding time series become persistent if $\frac{1}{2} < H < 1$ [13], resulting in the so-called long-range-dependence (LRD). The successive movements of the long-memory processes are inclined to remain consistent. Conversely, if $0 < H < \frac{1}{2}$, the time series become anti-persistent [13], and the movements tend to be inversely related to the previous period. Lastly, the time series become completely random in a case that $H = \frac{1}{2}$ [13]. Most time series investigated have a Hurst parameter between $0.5$ and $1$, thus show LRD.

Definition 2: Let $X_t$ be a stationary process. If there exists a real number $\alpha \in (0, 1)$ and a constant $c_\alpha > 0$, such that the autocorrelation function $\rho(k)$ satisfies:

$$\rho(k) = c_\alpha k^{-\alpha}, \quad \text{as} \ k \to \infty \quad (2.3)$$

then $X_t$ is called a stationary process with long memory, or long-range-dependent (LRD) [14]. The Hurst parameter $H$ is then obtained by $H = 1 - \frac{\alpha}{2}$. Long memory occurs for $\frac{1}{2} < H < 1$ [14].

Equation (2.3) shows that the autocorrelation decays slowly in a hyperbolic way, much more slowly than the exponential decay for a short-memory process, resulting in a non-summable ACF [16]. Based on this property, LRD can also be defined as a stationary time series with covariance (denoted as $\gamma(k)$) decaying so slowly that their sum diverges [16]:

$$\sum_{k=-\infty}^{\infty} \gamma(k) = \infty \quad \text{or} \quad \sum_{k=0}^{\infty} |\gamma(k)| = \infty \quad (2.4)$$

Two most important properties of LRD are stationarity and slowly decaying autocorrelation. If we compare the ACF of a LRD signal with that of a short-memory one as shown in Fig. 2.2, we can easily see the difference. For short-memory time series (such as white noise), the ACF decays to zero very fast, while for LRD signals (such as fractional Gaussian noise with $H=0.9$), the ACF still exists even after large lags. This is the reason why for short-memory time series, observations in the past have little effect on the future behavior, and for long-memory time series, observations far apart are still strongly
correlated. As a result, we cannot use the same technique as for short-memory time series when dealing with long-memory processes.

For time series with LRD, one can often observe data points clustering on one side of the mean for some time, and then the other side, unlike random noises for which data points are evenly distributed on both sides of the mean [15, 16]. This is referred to Joseph effect. These properties suggest that LRD signals cannot be analyzed in a traditional manner.

![ACF for white noise](image1.png) ![ACF for fGn with H=0.9](image2.png)

(a) (b)

Fig. 2.2 Comparison of the ACFs for short-memory and long-memory signals. (a) the ACF for white noise; (b) the ACF for fractional Gaussian noise with H=0.9

### 2.3 Self-similar processes

#### 2.3.1 Introduction to self-similar processes and classification

Self-similarity is the most important feature of a fractal. It can be described as invariance under suitable scaling of time or space [7]. In general, there are four kinds of self-similarity.

*Exact self-similarity* means each small part of the object is an exact copy of the whole, and this property applies for all scales. The Koch Curve is a good example of exact self-similarity. However, this is only an ideal case. For most of the fractals in nature, the pattern remains similar across different scales, rather than exactly the same, such as ferns, snowflakes, etc. This is called *quasi self-similarity.* *Multifractal* is
another kind of self-similarity that has multiple fractal dimensions, or the scaling rules are different across scales. Statistical self-similarity (sss for short) is the last but most important one, which applies to objects that have the same statistical properties [1]. This is the case of random fractals as mentioned in Chapter 1. Fractal time series are statistically self-similar in the sense that the signals have the same statistical property regardless of the time scale at which they are measured. This is the case we will discuss in this thesis.

2.3.2 Theory and properties

Definition 3: A process \( \{X(t), t \in \mathbb{R}\} \) is self-similar with parameter \( H \) if and only if

\[
X(at) \overset{D}{=} a^H X(t)
\]

(2.5)

where \( \overset{D}{=} \) denotes equivalency in finite joint distribution. \( H \) is the Hurst parameter discussed in the previous section, which is very important in characterizing a self-similar process [9].

In an ideal case, Equation (2.5) should hold for all \( a, t \in \mathbb{R} \). However, in reality, this property only holds for limited scales. The covariance function for a self-similar process \( X(t) \) is as follows [14]:

\[
\gamma_X(t, s) = \frac{1}{2} \sigma^2 [t^{2H} - (t-s)^{2H} + s^{2H}]
\]

(2.6)

From Equations (2.5) and (2.6), it is easy to see that self-similar processes must be non-stationary [17]. This is different from LRD, which requires stationarity.

In practice, signals are recorded in a discrete manner rather than continuously. If a discrete time sample of the process \( X(t) \) is taken, we can get a time series \( X_n \). Among self-similar processes, we are interested in a group of self-similar processes that have stationary increments (H-sssi), because H-sssi processes can produce stationary sequences that are useful in real application [17]. Let \( X_n \) be an H-sssi process and \( Y_n \) be its increment process, or

\[
Y_n = X_{n+1} - X_n
\]

(2.7)

then the time series \( Y_n \) is stationary. Define \( \{Y_i^{(m)}, i = 1, \ldots, n\} \) as the aggregate process of \( Y_n \),

\[
Y_i^{(m)} = \frac{1}{m} (Y_{im-m+1} + \cdots + Y_{im})
\]

(2.8)
then scale-invariance for this discrete stationary time series can be described as:

\[ m^{1-H} Y^{(m)} = Y \]  \hspace{1cm} (2.9)

where “\(D\)” also denotes equivalency in finite joint distribution. \(H\) is the Hurst parameter for the process \(Y_n\), which is the same as the Hurst parameter for the original self-similar process \(X(t)\). The scaling property implied by Equation (2.9) suggests that the averaged sequence has the same distribution as the original time series after proper scaling according to aggregate levels [17] (see Fig. 2.3 for an example).

Fig. 2.3 Original time series and its aggregated sequences. The time series aggregated on different scales (middle: \(m=4\); bottom: \(m=16\)) are similar to the original time series (top).

### 2.3.3 Self-similarity and long-range-dependence (LRD)

Self-similar and LRD are closely related, but they are different and should not be confused. The main difference between self-similar processes and processes with LRD is that self-similar processes are non-stationary, while LRD processes are stationary by definition. The differences between self-similar and LRD processes are listed in Table 2.1.

However, these two kinds of processes are related by a parameter – the Hurst parameter, and one process can be derived from the other. For a self-similar process \(X_n\) with stationary increments (\(H\)-ssi), its increment process \(Y_n\) is stationary. Let \(\gamma_y(k)\) denote the covariance of \(Y_n\), then \(\gamma_y(k)\) can be described as:
\[ \gamma_Y(k) = \frac{\sigma^2}{2} [(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \]  

(2.10)

If the Hurst parameter for \( X_n \) and \( Y_n \) satisfies \( \frac{1}{2} < H < 1 \), then the sum of \( \gamma_Y(k) \) would diverge, or

\[ \sum_{k=1}^{\infty} \gamma_Y(k) = \infty \text{, thus } Y_n \text{ is LRD.} \]

Table 2.1 Differences between self-similar and LRD processes

<table>
<thead>
<tr>
<th>Differences</th>
<th>Self-similar processes</th>
<th>LRD processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stationarity</td>
<td>Non-stationary</td>
<td>Stationary</td>
</tr>
<tr>
<td>2. Scaling law</td>
<td>( X(at) = a^H X(t) )</td>
<td>( m^{1-H} Y(m) = Y )</td>
</tr>
</tbody>
</table>

To sum up, if a non-stationary time series is self-similar with stationary increments, and its Hurst parameter is between 0.5 and 1, then its corresponding increment process is LRD. Conversely, if a stationary time series is LRD, then its cumulative process is non-stationary and self-similar. This relationship is illustrated in Fig. 2.4. A typical example will be given in the next section.

**2.4 Fractional Brownian motion (fBm) and fractional Gaussian noise (fGn)**

Fractional Brownian motion (fBm) is a typical example of the \( H \)-sssi processes. It is a generalization of the standard Brownian motion and can be defined by stochastic calculus [18]:

\[ B_H(t) = \frac{1}{\Gamma(H + 1/2)} \left( \int_{-\infty}^{0} (t - s)^{(H-1/2)} \left[ dB(s) + \int_{0}^{t} (t - s)^{(H-1/2)} dB(s) \right] \right) \]

(2.11)
where $B(t)$ is standard Brownian motion and can be recovered by taking $H = \frac{1}{2}$.

fBm is the only self-similar Gaussian process with stationary increments and has a Hurst parameter between 0 and 1 [17]. Standard Brownian motion is a special case of fBm, with a Hurst parameter of 0.5. It is non-stationary with stationary independent increments, indicating that it is completely random. This suggests that even if the time series is observed increasing or decreasing at this moment, we still have no idea in which direction it will go in the next period. fBm is similar to the standard Brownian motion in the sense that it is also non-stationary and it has stationary increments. However, for fBm with a Hurst parameter not equal to 0.5, the increments are not independent. If $H > 0.5$, then an increasing time series will more likely increase during the next period and in the future [13]. Thus $H$ describes the raggedness of a fractional Brownian motion. There are fewer fluctuations as $H$ increases, and the curve becomes smoother. Figure 2.5 shows fBm signals with different Hurst parameters. When we zoom in Fig. 2.5 (c) and get its partial plot, the same statistical pattern reappears, as shown in Fig. 2.5 (d).

Fig. 2.5 fBm with different Hurst parameters. (a) fBm with H=0.1; (b) standard Brownian motion (fBm with H=0.5); (c) fBm with H=0.9; (d) fBm with H=0.9 (Fig. 2.4 (c) zoomed in)
Define the incremental process \( Y = \{Y_k, k = 1, 2, \cdots \} \) by:

\[
Y_k = B_H (k+1) - B_H (k)
\]  

(2.12)

\( Y_k \) is called fractional Gaussian noise (fGn), and it is the discrete step incremental process of a fractional Brownian motion [16]. The standard Brownian motion is a cumulative sum of the white noise. In a similar way, fractional Brownian motion is also a cumulative sum of fGn [18].

fGn is stationary, and its covariance function is the same as Equation (2.10). If \( H = \frac{1}{2} \), covariances are 0 for all \( k \neq 0 \). This is the case of the white noise, where all observations are independent. When \( H > \frac{1}{2} \), \( Y_k \) is LRD. The autocorrelation function for fGn with \( H > 0.5 \) decays much more slowly than that of the white noise. Readers are referred to Fig. 2.2 for a comparison. Fig 2.6 is a comparison between the self-similar process fBm and its corresponding LRD process fGn with \( H=0.9 \).

![Fig. 2.6 Comparison between self-similar (fBm) and LRD (fGn) processes. (a) fBm with H=0.9; (b) fGn with H=0.9.](image)

2.5 Summary

In this chapter, a detailed theoretical overview on self-similar and LRD processes is presented, and some of the key concepts and statistical properties are illustrated. The Hurst parameter is one of the most important parameters that can characterize a self-similar or LRD signal. It takes a value between 0 and 1, and mostly we deal with processes with a Hurst parameter larger than 0.5, which are known to possess long memory. Both self-similar and LRD processes have the scale-invariance feature, but as discussed in this chapter scale-invariance has different forms of representation for self-similar and LRD processes. The
main reason for this is that self-similar and LRD processes are in theory not identical yet transferable. At the end of this chapter, a typical example of self-similar processes is given, which is fractional Brownian motion (fBm), to help the understanding of the concepts discussed.
Chapter 3 Literature Review

Fractal and self-similarity have been hot topics ever since they were first introduced. They provide a new way to explore the world. This chapter provides a literature review on time series analysis, especially long-memory time series analysis, as well as other research related to fractal and self-similarity.

3.1 Time series analysis

Time series analysis generally consists of the following steps: first analyze the autocorrelation function and determine whether the time series is stationary or not; then choose a proper model and conduct parameter estimation; at last use the model to make forecasts [19]. In real-time application, on-line model update is also essential. In this section, we review general time series analysis technique and focus on how it is extended to long-memory processes. Afterwards, one type of time series that has long-memory is discussed, i.e. realized volatility.

3.1.1 From ARIMA to FARIMA

With time series analysis, we cannot ignore the famous Box & Jenkins ARIMA model [19, 20]. ARIMA is short for autoregressive integrated moving average, and it is a class of linear models that can describe both stationary and non-stationary time series. The shortage is that it can only model short-memory time series. If a time series is stationary and has short memory, then its autocorrelation function would decrease to zero so quickly that their sum converges [19]. If a time series is non-stationary, a stationary series can be derived by differencing the original series, and the degree of differencing depends on the property of the original time series. If the original time series shows a linear trend, differencing once is enough; if the resulting series is still non-stationary, then further differencing should be applied. After proper differencing, the autocorrelation function should decays exponentially fast to zero [19]. However, the degree of differencing can only take integers in the traditional ARIMA model.

In the previous chapter, we discussed a group of stationary processes whose autocorrelation function decays slowly and their sum diverges. The ARIMA model cannot be applied to these long-memory time series directly due to its lack of ability to describe the persistency. Under this situation, a generalization of
the ARIMA model – the FARIMA model – was proposed by J. R. M. Hosking in 1983, which is short for fractional autoregressive integrated moving average [21]. The FARIMA model successfully describes the long-memory property of a time series, and has been applied to model the LRD in economics and hydrology [17], as well as internet traffic data [22]. In his model, fractional differencing is introduced, which allows the degree of differencing to take fractional numbers [19]. More details on fractional differencing can be found in Chapter 4.

3.1.2 Modeling realized volatility

In finance, volatility is a measure of risk and thus important in asset pricing. There are two kinds of volatilities that are widely used, i.e. realized volatility and implied volatility. We focus on the realized volatility in this section, and leave the implied volatility to Chapter 5.

The most basic definition of volatility is the standard deviation of an asset return series. As high frequency trading data is available these days, daily realized volatility can be computed as the sum of squared intraday returns [23]. The most popular daily realized volatility is obtained by summing all the squared 5-min returns in a single trading day [23]. According to the efficient-market hypothesis (EMH), returns are almost uncorrelated, with a Hurst parameter near 0.5. However, research shows that both absolute and squared returns are positively correlated [24]. Realized volatility is the sum of squared returns, and it is LRD with a Hurst parameter between 0.5 and 1. Realized volatility is also found to possess a clustering behavior, which means large volatilities tend to appear consecutively. Based on this, the ARCH (Autoregressive Conditional Heteroscedasticity) model [25] and GARCH (Generalized ARCH) model [26] were proposed to model volatility.

In ARCH model, the basic idea is that the variance of the residuals is not a constant [24], but varies in time. Define $\mu$ and $\sigma_i$ as the mean and standard deviation (volatility) of the asset return respectively, and $\epsilon_i$ as the error term. If the asset return is denoted as $r_i$:

$$r_i = \mu + \sigma_i \epsilon_i,$$

then the forecasted variance for the next period can be described by the following equation [24]:

$$\sigma_{r_{t+1}}^2 = c_0 + c_1 (\sigma_i \epsilon_i)^2$$  (3.1)
where \( c_0 \) and \( c_1 \) are model parameters.

Equation (3.1) is the ARCH(1) model. The GARCH model is an extension of the ARCH model [24] in the sense that in the GARCH model, the forecast for the next period’s variance also has something to do with previous variance, or:

\[
\sigma_{t+1}^2 = c_0 + c_1 (\sigma_t \varepsilon_t)^2 + \beta \sigma_t^2
\]  

Later, long-memory factor is included and fractionally integrated GARCH (FIGARCH) and fractionally integrated Exponential GARCH (FIEGARCH) were proposed [27]. The FARIMA model has also been used in the modeling of volatility. But these models containing fractional differencing are complex and hard to estimate.

Corsi [28] proposed a simple model: the heterogeneous autoregressive model of realized volatility, or HAR-RV for short. He found the sum of several autoregressive processes aggregated on different levels can produce a similar behavior to that of long-memory processes [28, 29]. The author tested the HAR-RV model on USD/CHF, S&P 500 futures and T-bond data, and performed one-step-ahead forecast – predicting one-day ahead in the future [29]. It turns out that the simple HAR-RV model can capture the long-memory behavior, and produce similar accuracy [29] compared with the cumbersome FARIMA model. The HAR-RV model has gained attention ever since it was first developed. Later it is generalized to the HAR-RV-J model to incorporate the jump components [23]. Andersen et al. [23] studied the jump measurements and jump dynamics in detail, and then proposed a continuous sample path variability measure \( C \). The HAR-RV-J model is then extended to the HAR-RV-CJ model, and the realizations and forecasts are found to be quite coherent [23].

### 3.2 Self-similarity related research

Other than time series analysis and prediction, there is also research related to fractals and self-similar processes, which mainly lies in the following areas: estimation of the self-similar parameter \( H \), fractional Brownian motion and its application, and fractals in human physiology.

#### 3.2.1 Estimation of the self-similar parameter \( H \)

Hurst parameter can characterize a self-similar process. The estimation of the Hurst parameter has long been studied and still remains a challenging topic.
Chapter 3. Literature Review

The first and most well-known estimator is the R/S estimator by Hurst [12] and has already been discussed previously. According to Equation (2.2), plot log(R/S) versus log(n) for each n, and fit a straight line, then we can get the Hurst parameter H. This plot is called the pox plot [30]. Usually a low cutoff limit and a high cutoff limit for n are chosen, and only the points between the cutoff limits are used in order to get a reasonable estimation.

Aggregate variance method is another estimator, and was proposed by Beran [14]. He utilizes the aggregate process of a LRD time series. According to Equation (2.9), \(m^{1-H}Y^{(m)}\) has the same distribution as Y. So their variances should satisfy \(m^{2-2H}Var(Y^{(m)}) = Var(Y)\). The algorithm can be described as: first divide the process into blocks of size m, and compute the variances of each block; choose different block sizes m and repeat the process. The Hurst parameter can be found by plotting \(Var(Y^{(m)})\) versus m on the log-log plot, fitting a straight line and finding the slope [18].

Higuchi method was proposed by Higuchi [31]. This method makes use of the relationship between the fractal dimension \(D_F\) and the Hurst parameter H. He first computes the length L of a normalized curve respect to various block sizes m, and then fits a straight line to the log-log plot of L versus m [30]. The slope is the fractal dimension \(D_F\), and the Hurst parameter can be derived using \(H = 2 - D_F\).

The wavelet estimator was introduced by Abry et al. [32]. The authors found that the wavelet coefficients of a self-similar process have zero mean and variance \(\sigma^2 2^{(2H+1)}\), where j is the scale [18]. Based on this, they perform wavelet decomposition to a self-similar process and get the wavelet coefficients on different scales, and then the Hurst parameter is obtained by fitting a straight line to the log-log plot of variance versus scale j.

Above we have listed several prevailing estimation methods for the Hurst parameter. Other methods include periodogram method, which is also referred to the method of Geweke and Porter-Hudak (GPH) [33]; Peng estimator, also known as the detrended fluctuation analysis (DFA), which takes advantage of the variance of residuals [34]; Whittle estimator, which attempts to maximize the likelihood function [14]; etc. There are also methods that are modifications of the existing ones. Interested readers are referred to Beran [14], Dieker [18], and Taqqu et al. [30] for detailed studies. Different estimation methods often result in different Hurst parameters, and hence the estimation of the Hurst parameter still remains a challenging topic.
3.2.2 Fractional Brownian motion

Studies on the theory of fBm include its self-similar property, and stochastic calculus respect to fBm [35]. Another interesting area is the simulation of fBm processes [18]. Interested readers are referred to Dieker [18] for a detailed study. We focus more on the application of fBm and fGn.

Stochastic processes such as the white noise are often used to model the input of a system, the network traffic for instance. Previously, the standard Brownian motion is used based on the Markovian assumption. Recent study shows that LRD exists and a long-memory model would better describe the actual condition. fBm is then proposed to replace the standard Brownian motion and act as the input [36]. Norros [36] proposed a fractional Brownian traffic model, which takes advantage of the self-similar property of fBm, and received satisfactory results.

While the standard Brownian motion is widely used in finance, studies have shown that LRD plays an important role in the financial market. If the standard Brownian motion is replaced by an fBm, the model is long-memory and can capture the inherent LRD. As a result, fractional Black-Scholes model is proposed using fBm; in a similar way, a fractional O-U process can be defined and employed [35, 37].

Fractional Brownian motion is not only limited to two-dimensional time series, but can also be used spatially, such as the application in turbulence. The vorticity field is concentrated along a tube centered in a curve, which is assumed to be the trajectory of a three-dimensional fBm [35]. fBm is also used to geometrically describe the cracking in materials, synthesize the spatial scaling of certain permeability fields, track particles by modelling ocean surface drifter trajectories [38], etc.

3.2.3 Fractals in human physiology

Fractals have been observed in human physiology and can be used to diagnose diseases or find abnormal health conditions. The studies related to this field can be divided into two groups. The first group is the fractal-like structure in human bodies, such as blood vessels, nerves, tubes that transport gas, etc. [3]; the second group is that the signals produced by human bodies show a chaotic and fractal behavior.

Fractals and chaos are closely related. They are both non-linear dynamics that are often found in nature. Fractals are often the remnants of a chaotic system [3]. The system represented by a strange attractor is chaotic, and have fractal property.
In the past, researchers and physicians all believed that healthy people would produce regular and periodic physiological signals, while the existence of disease would result in erratic and chaotic signals. However, Goldberger et al. [3] showed that human beat-to-beat interval data is more like a strange attractor after phase space reconstruction, suggesting an inherent chaotic behavior. Recent studies also revealed that human gait, blood pressure, heart rate, etc. also exhibit fractal feature [39-41], though the underlying mechanism is not clear yet.

Interestingly, the degree of fractal has something to do with the healthiness of people. The physiological signals for young and healthy people tend to be more erratic and vary more dynamically, while those for aged and diseased ones usually look more regular and lack variability [3, 42, 43]. In fact, this observation has been used to detect certain diseases. For example, Stanley et al. [44] found that the cardiac interbeat interval signal is patchy and made up of erratic fluctuations for normal person, while it appears more periodic for those with sleep apnea; the heart rate time series shows more variability for normal person, while there is a loss of complexity for those suffering from a congestive heart failure [44].

Detrended fluctuation analysis (DFA) by Peng as mentioned earlier is the most widely used tool for the analysis of chaotic and fractal behavior in physiological signals. With the help of DFA, Goldberger et al. [40] and Stanley et al. [44] pointed out that a single scaling parameter $H$ cannot fully describe physiological signals, but different parts of the signal have different scaling laws, or different Hurst parameters. This multifractal phenomenon exists for healthy people, and a breakdown of multifractality may suggest heart failure [40, 44].

Research that utilizes the fractal and chaotic behavior of human physiological signals as a diagnostic tool for diseases have been quite popular these days, and we believe this will remain a hot topic in the future.
Chapter 4 Algorithms and Models

Self-similar signals have different statistical properties from those ordinary ones, and the existence of long memory makes it harder to analyze. The unique scaling property, however, provides us with other approaches to analyze and model them. In this chapter, several models that can describe self-similar and LRD processes are discussed, i.e. FARIMA, HAR-RV, wavelet-based and average-VAR. Some models have been proposed for years, some proposed in recent years. At the end of this chapter, a new model is proposed based on the scaling property of LRD processes.

4.1 FARIMA model

4.1.1 Definition

The FARIMA model has been mentioned in Chapter 3, and it is an extension of the ARIMA model. The ARIMA model we commonly use can model processes for which the ACF converges to zero rapidly or after proper differencing, the ACF converges [20]. However, for LRD signals, the ACF decays slowly, resulting in a strong correlation between two observations lying far apart. The FARIMA model proposed by Hosking, which allows the degree of differencing to take fractional numbers, however, can capture this strong autocorrelation [21].

The general FARIMA (p, d, q) model can be defined as [21]:

\[ \Phi(B)(1 - B)^d X_t = \Theta(B)e_t \] (4.1)

where

\[ \Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots \]
\[ \Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots \]

\(B\) is the backward shift operator, \(d \in (-0.5, 0.5)\) is the degree of differencing, \(X_t\) is the process to be modelled, and \(e_t\) is the error term. Recall that ARIMA (p, d, q) model has the same form. The only
difference between the two models is that $d$ can take fractional numbers in the FARIMA model, while in the ARIMA model, $d$ can only take integers. Studies have shown that the differencing parameter $d$ is related to the Hurst parameter $H$ by $d = H - \frac{1}{2}$ [21]. For $d \in (0, 0.5)$, the process $X_t$ in Equation (4.1) has long-memory and can model the LRD time series [22]. The fractional differencing operator can be computed as [21]:

$$
(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = 1 - dB - \frac{1}{2} d (1 - d) B^2 - \frac{1}{6} d (1 - d) (2 - d) B^3 - \cdots \quad (4.2)
$$

### 4.1.2 Parameter estimation

Parameter estimation is of great importance in the application of the FARIMA model. Given a time series with LRD, the first step is to estimate the Hurst parameter $H$ and obtain $d$ using $d = H - \frac{1}{2}$. There are several methods in estimating the Hurst parameter, including the R/S analysis, the Higuchi method, the aggregate variance method, the wavelet based approach, etc. Interested readers are referred to Chapter 3 and Dieker [18].

After the differencing parameter $d$ is obtained, the fractional differencing needs to be performed to the time series to remove the persistence and transform the original long-memory process into a short-memory one. Finally time series parameter estimation algorithms can be applied, such as the maximum likelihood estimation, the least squares estimation, etc.

### 4.1.3 Advantages and drawbacks

The FARIMA model has gained attention ever since it was first proposed. This is because of its ability to model the persistence of a LRD time series.

However, the FARIMA model has some drawbacks that must not be neglected:

1. The differencing parameter $d$ needs to be estimated via the Hurst parameter. However, existing methods in estimation $H$ have shown inconsistency. Different methods give different results, causing a large error in estimating $d$. In addition, the estimation of $H$ requires the use of all the data points, which is not feasible in real-time application [29].
ii. Fractional differencing operator represented in Equation (4.2) is an infinite sum, which is not applicable in real application. When applying Equation (4.2), one often chooses a cutoff limit, resulting in errors and poor accuracy.

iii. Parameter estimation is highly complicated and time-consuming for the whole process.

4.2 HAR-RV model

The HAR-RV model proposed by Corsi [28, 29] has been mentioned in Chapter 3 as an effective model for realized volatility. He found that the aggregation of several autoregressive models can produce a process which shows a rather similar behavior to the LRD processes, and the ACF produced by the HAR-RV model decays slowly, which is close to a LRD process [28, 29].

The model utilizes realized volatility aggregated on weekly and monthly bases, and represents the future volatility as a weighted sum of the volatility on the same scale and on higher levels [28, 29]. This model can capture the short-term variations as well as long-term trend [29]. The time series representation of the model can be described as [29]:

\[
RV_{t+1}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + e_{t+1d}
\]

(4.3)

where \( RV_t^{(d)} \) is the daily realized volatility at time \( t \), \( RV_t^{(w)} \) is the weekly realized volatility at time \( t \), \( RV_t^{(m)} \) is the monthly realized volatility at time \( t \), and \( RV_{t+1}^{(d)} \) is the predicted daily realized volatility at time \( t+1 \). Weekly and monthly realized volatilities are computed as the average of recent daily realized volatilities. As a rule of thumb, it is considered that there are 5 trading days in a week, and 22 trading days in a month. Weekly and monthly realized volatilities are therefore aggregated on 5 days and 22 days, respectively. Coefficients can be estimated using the least squares fitting approach.

This simple HAR-RV model can produce quite satisfactory results, compared with the complicated FARIMA model. However, the HAR-RV model has only been used in the context of realized volatility, which has natural aggregation levels (5 days and 22 days). Later it will be extended to the modelling and prediction of blood pressure signals, for which the selection of aggregation levels is a topic for discussion.
4.3 Wavelet-based model

Wavelet-based analysis for self-similar signals has gained popularity due to its similar scaling property, which allows an efficient application [45-48]. The wavelet-based model in this thesis is due to Wang and Lee [49], who study the self-similarity in human arterial blood pressure signal to detect the acute hypotension episode (AHE).

4.3.1 Wavelet basics

Wavelet analysis is a time frequency analysis which has been used widely in the engineering fields, such as image compression, signal denoising, etc. A wavelet is a function $\psi(t)$ that satisfies $\int_{\mathbb{R}} \psi(t)dt = 0$ as well as square-integrable condition [50]. Haar wavelet is the simplest wavelet and it is a piecewise constant function defined as:

$$\psi(t) = \begin{cases} 
1, & \text{if } 0 \leq t < 1/2 \\
-1, & \text{if } 1/2 \leq t < 1 \\
0, & \text{otherwise}
\end{cases} \quad (4.4)$$

The function $\psi(t)$ is called mother wavelet and its child wavelet $\psi_{j,k}(t)$ can be obtained by a translation and dilation of the mother wavelet [50]:

$$\psi_{j,k}(t) = \frac{1}{2^{j/2}} \psi(2^{-j} t - k), \quad j, k \in \mathbb{Z} \quad (4.5)$$

A discrete wavelet transform (DWT) of the process $X(t)$ can be defined as [50]:

$$d_{j,k} = \int_{\mathbb{R}} X(t) \psi_{j,k}(t)dt, \quad j, k \in \mathbb{Z} \quad (4.6)$$

where $d_{j,k}$ are called wavelet coefficients, or details (see Fig. 4.3).

For each mother wavelet, there is a corresponding scaling function $\phi(t)$, which allows us to generate a sequence of spaces $V_j$ that can approximate functions from $L^2(\mathbb{R})$. The space generated by the wavelet function $\psi(t)$ is denoted as $W$, which can also be viewed as the error incurred when approximating functions in $V_{j+1}$ by functions in $V_j$ [51]. Please refer to Fig. 4.1 for a clearer explanation.
In Fig. 4.1, $V_0 = V_1 \oplus W_1 = V_2 \oplus W_2 \oplus W_1 = V_3 \oplus W_3 \oplus W_2 \oplus W_1 = \cdots$. For any function $f(t) \in V_0$, we can decompose it into two parts: one is the approximation in coarser scales ($V_j$), the other is the details in finer scales ($W_j$). This is called multiresolution analysis (MRA) [51]:

$$f(t) = \sum_{k=-\infty}^{\infty} a_{j,k} \phi_{j,k}(t) + \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t)$$

where $d_{j,k}$ can be obtained by Equation (4.4), and $a_{j,k}$ are called approximation coefficients and can be obtained by the following equation:

$$a_{j,k} = \int_{\mathbb{R}} f(t) \phi_{j,k}(t) dt, \quad j, k \in \mathbb{Z}$$  \hspace{1cm} (4.7)

Fig. 4.2 is an illustrative diagram of the three-scale MRA.

For the Haar wavelet mentioned in Equation (4.4), the corresponding scaling function is:

$$\phi(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{otherwise} \end{cases}$$

Fig. 4.3 is an example of the three-scale MRA performed on a blood pressure signal when Haar wavelet is used.
Vanishing moments $R$ is an important property of wavelet functions. A wavelet function has vanishing moments $R$ if:
Equation (4.8) suggests a wavelet function with vanishing moments $R$ can generate polynomials of degree smaller than $R$ [50, 51].

Daubechies wavelets are a family of orthogonal wavelets with given supports [51, 52]. A Daubechies wavelet with $R$ vanishing moments has support width of $N=2R-1$ [52]. $\text{db}R$ can be used to represent a Daubechies wavelet with $R$ vanishing moments. The Daubechies wavelets generally do not have explicit expressions except for the Haar wavelet, which is a special case of the Daubechies wavelet with vanishing moment 1 and support width of 1 [52]. Fig. 4.4 shows the approximation of some Daubechies wavelet functions.

![Wavelet Functions](image)

**Fig. 4.4 Approximation of some Daubechies wavelet functions**

### 4.3.2 Wavelet transform and self-similarity

Wavelet analysis can be applied to self-similar signals. Research shows that the wavelet coefficients for self-similar signals are covariance stationary at each scale [46, 48]. For instance, the autocorrelation for fBm decays slowly since fBm is non-stationary; however, after the wavelet decomposition, the autocorrelation for the wavelet coefficients on a single scale decays much faster than that of fBm itself. Fig. 4.5 is a comparison between the autocorrelation functions for fBm and its wavelet coefficients when the Haar wavelet is used.

The rate of decay of the wavelet coefficients is closely related to the vanishing moments of the wavelet used. Large vanishing moments $R$ may lead to almost uncorrelated wavelet coefficients [46, 48]. If $R$ is
chosen such that \( H > R - 1/2 \), then divergence occurs for the sum of autocorrelation of wavelet coefficients [46, 48], i.e. ACF decays hyperbolically. If a wavelet with larger vanishing moment is used, the ACF of wavelet coefficients may decay exponentially fast to zero. Fig. 4.6 is a comparison between the autocorrelation functions for wavelet coefficients when different wavelet functions are used.

Fig. 4.5 Comparison of the ACFs before (left) and after (right) wavelet transform

Fig. 4.6 Comparison of the ACFs using different wavelets (R=1 for the left and R=2 for the right)

Fig. 4.5 and Fig. 4.6 show that within-scale correlation is strongly weakened and non-stationary signal can be analyzed in a stationary manner after wavelet transformation. However, across-scale correlation is still maintained [46, 48, 49], which can be described by Equation (4.9):

\[
d_{j,k}^D = 2^{j(H+1/2)} d_{0,k}
\]

(4.9)


**Proof.**

\[
d_{j,k} = \int_{\mathbb{R}} X(t) \psi_{j,k}(t) \, dt
\]

\[
= \int_{\mathbb{R}} X(t) \cdot \frac{1}{2^{j/2}} \psi(2^{-j} t - k) \, dt
\]

\[
= \int_{\mathbb{R}} 2^{j/2} X(t) \cdot 2^{-j} \psi(2^{-j} t - k) \, dt
\]

\[
= \int_{\mathbb{R}} 2^{j/2} X(2^j u) \psi(u - k) \, du
\]

\[
= \int_{\mathbb{R}} 2^{j/2} \cdot (2^j)^H X(u) \psi(u - k) \, du
\]

\[
= 2^{jH + 1/2} \cdot d_{0,k}
\]

By Equation (4.6)

By Equation (4.5)

Let \( u = 2^{-j} t \)

By Equation (2.5)

Equation (4.9) can be explained as: the wavelet coefficients on different scales exhibit the same statistical property after proper scaling. The strong correlation across scales can also be illustrated by the wavelet coefficient plots in Fig. 4.3 (d_1-d_3).

**4.3.3 Wavelet-based algorithm**

The wavelet-based algorithm is motivated by the fact that wavelet decomposition for self-similar signals can reduce the within-scale correlation yet maintain the across-scale correlation. In theory, the larger vanishing moments the wavelet function has, the weaker the within-scale correlation will be. However, the Haar wavelet is chosen because it has a clear cutoff and only uses the current and past observations, while the Daubechies wavelets with vanishing moments greater than 1 have long tails (see Fig. 4.4) and use future observations, which is not applicable. For a self-similar signal, the algorithm can be described as: first decompose the signal using the Haar wavelet, and then apply a vector autoregressive (VAR) model to capture the across-scale correlation [49]; the forecasted wavelet coefficients are then transformed back to get the desired forecast for the original time series.

When applying VAR model to the wavelet coefficients, it is found that the number of wavelet coefficients are not identical on different scales. However, the VAR model requires the same number of wavelet coefficients at all scales. Wang and Lee [49] used a down sampling approach to overcome this issue (see Fig. 4.7). They first grouped the time series into blocks of 4 numbers, and then created vectors of wavelet coefficients based on a three-scale wavelet decomposition. Each time, the values of the next vector are forecasted and only the wavelet coefficient on the top level is used in order to obtain a forecast for the
average of the future four time periods. For the Haar wavelet, the wavelet coefficients on scale \( j+1 \) satisfy [49]:

\[
d_{j+1,n} = 2^{-\frac{j-1}{2}} (X_{2n-1}^{(2^j)} - X_{2n}^{(2^j)})
\]

(4.10)

where \( X_{2n}^{(2^j)} \) is the aggregate process defined by Equation (2.8). After the wavelet coefficient \( d_{3,n+1} \) on scale 3 is obtained, the forecast for the average of the future four time periods can be computed as:

\[
Ave_{n+1} = Ave_n - \frac{d_{3,n+1}}{2^{1/2}}
\]

(4.11)


Fig. 4.7 Across-scale prediction algorithm (down-sampling approach)

### 4.4 Average-VAR model

Recall the statistical properties of self-similar signals. By Equation (2.9), the increment process of a self-similar signal also has the scaling property:

\[
m^{1-H} Y^{(m)} = Y
\]

where \( \{Y_i^{(m)}, i = 1, \cdots, n\} \) is the aggregate process of \( Y_n \). Note this increment process \( Y_n \) is stationary and LRD.
Based on this scaling property, a new model is proposed: the average-VAR model. Aggregate process is obtained by averaging the time series. There should be a strong correlation between processes aggregated on different scales, so a vector autoregressive (VAR) model is applied to describe this across-scale correlation. Fig. 4.8 is an illustration of the three-scale average-VAR model.

![Diagram of the three-scale average-VAR model]

**Fig. 4.8 An illustration of the three-scale average-VAR model**

This average-VAR model is quite straightforward in theory and simple to use in practice. For a stationary time series with LRD, the general steps in realizing this model can be described as: first average the time series on three different time scales to incorporate both short-term and long-term trend; then apply the VAR(1) model based on Equations (4.12) and (4.13). \( V_{t+1} \) is our target forecast.

\[
V_{t+1} = c + AV_t + e_t
\]  
\[
\begin{bmatrix}
V_{1,t+1} \\
V_{2,t+1} \\
V_{3,t+1}
\end{bmatrix} =
\begin{bmatrix}
c_1 \\
c_2 \\
c_3
\end{bmatrix} +
\begin{bmatrix}
A_{11} & A_{12} & A_{13} \\
A_{21} & A_{22} & A_{23} \\
A_{31} & A_{32} & A_{33}
\end{bmatrix}
\begin{bmatrix}
V_{1,t} \\
V_{2,t} \\
V_{3,t}
\end{bmatrix} +
\begin{bmatrix}
e_1 \\
e_2 \\
e_3
\end{bmatrix}
\]  
(4.13)

**4.5 Summary**

The time series satisfying conditions for the above four models are summarized in Table 4.1. Remember some of the models are used for stationary time series with LRD, while some are used for non-stationary time series with self-similarity. If we want to apply a model to a time series that doesn’t satisfy the
condition, the time series must be first transformed according to the relationship between LRD and self-similarity.

Table 4.1 Summary of the models and their time series satisfying conditions

<table>
<thead>
<tr>
<th>Models</th>
<th>Time series satisfying conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>Time series to be stationary &amp; LRD</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>Time series to be stationary &amp; LRD</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>Time series to be non-stationary &amp; self-similar</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>Time series to be stationary &amp; LRD</td>
</tr>
</tbody>
</table>
Chapter 5 Numerical Studies and Results

In this chapter numerical studies on several real-life signals will be presented. The signals include arterial blood pressure signals and volatility indexes, both of which exhibit LRD. The four models discussed in Chapter 4 are then modified and applied to the data sets. We focus on the out-of-sample forecast, more specifically, one-step-ahead forecast. Several error metrics as well as the ROC curve are used to compare the performance of the models.

5.1 Data description

Many signals and processes in real life exhibit self-similar or LRD property, such as internet traffic, volatility in finance, physiological signals, etc. However, some data sets are really hard to collect. In this thesis, two groups of data sets are used, both of which are LRD and easy to acquire: arterial blood pressure signals and volatility index time series.

5.1.1 Arterial blood pressure signal

PhysioNet [43] is funded by the National Institutes of Health of the U.S. It is an open source online forum that provides free exchange of biomedical signals and also related analyzing software. In addition, it provides access to PhysioBank, a large archive that stores physiological signals that can be used in biomedical research [43]. These signals include cardiovascular signals, ECG signals, human gait signals, as well as physiological signals that are recorded from people with certain diseases, and so on. In order to analyze and model the signals, the number of data points in each signal should be at least 4000 to be considered as sufficient. However, too many data points would be time-consuming, so we decided to use signals of length less than 8000. Since blood pressure signal is known to be LRD, we chose 25 arterial blood pressure (ABP) signals from the database as our test signals, each of which has a number of data points ranging from 4000 to 8000, and numbered them as ABP1, ABP2, ABP3, … .
5.1.2 Volatility index

There are two kinds of volatilities as mentioned in Chapter 3, i.e. realized volatility and implied volatility. Realized volatility has been discussed already, so in this section implied volatility is introduced.

Implied volatility is often related to option pricing and can be computed as the volatility that makes the price obtained by option pricing model (such as Black-Scholes model) equal to its current market value [53]. Since computing realized volatility requires large amount of daily high frequency trading data that are not easy to acquire, we decided to use implied volatility as our test data.

VIX is the ticker symbol for Chicago Board Options Exchange Market (CBOE) volatility index [54]. It is a measure of the implied volatility of S&P 500 index options, and can be obtained by computing the expected volatility of the market over the next 30-day period [53, 54]. We also use other implied volatility indexes, refer to Table 5.1 [54] for a full list. These volatility index data can be acquired on CBOE website.

Table 5.1 Volatility indexes [54]

<table>
<thead>
<tr>
<th>Ticker</th>
<th>Index</th>
<th>Website</th>
<th>Time span</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIX</td>
<td>CBOE Volatility Index (S&amp;P 500)</td>
<td><a href="http://www.cboe.com/VIX">www.cboe.com/VIX</a></td>
<td>1990.1.2-2014.2.28</td>
</tr>
<tr>
<td>VXO</td>
<td>CBOE S&amp;P 100 Volatility Index</td>
<td><a href="http://www.cboe.com/VXO">www.cboe.com/VXO</a></td>
<td>1986.1.2-2014.2.28</td>
</tr>
<tr>
<td>VXD</td>
<td>CBOE DJIA Volatility Index</td>
<td><a href="http://www.cboe.com/VXD">www.cboe.com/VXD</a></td>
<td>1997.10.7-2014.2.28</td>
</tr>
</tbody>
</table>

5.2 Preliminary study

In this section, a preliminary study on the data sets mentioned above is conducted and some of the key properties are studied. Given a time series, it is essential to first analyze its stationarity, the Hurst parameter, and the ACF in order to determine whether it is self-similar, LRD, or neither. Afterwards, the models described in Chapter 4 are modified to better fit our data sets.

5.2.1 Properties

In order to obtain the Hurst parameter, R/S analysis is performed to the arterial blood pressure signals as well as the volatility index time series. Recall that R/S analysis can only determine the Hurst parameter for a stationary time series. If the result obtained from the R/S analysis is between 0 and 1, then it is the
Hurst parameter for the time series; however, if the result is greater than 1, then the analyzed time series is non-stationary, and must be differenced first before R/S analysis is performed again to get the true Hurst parameter. After the Hurst parameter is obtained, we can determine whether the process is stationary or not. If the Hurst parameter is between 0.5 and 1, and the time series is non-stationary, then it is self-similar; if the Hurst parameter is between 0.5 and 1, and the time series is stationary, then it is LRD. Fig. 5.1 is an illustrative diagram of the previous procedure.

Fig. 5.1 An illustrative diagram of the analyzing procedure

Following the above steps, the Hurst parameters for the data sets described in section 5.1 are found to be in the range of (0.5, 1), and the signals are all stationary according to R/S analysis, thus are LRD (not self-similar). The Hurst parameters obtained by R/S analysis are summarized in Table 5.2. Note that using other estimation approaches may result in slightly different Hurst parameters. Fig. 5.2 is the
autocorrelation functions for the ABP5 and VIX signals, respectively. We can clearly observe slowly decaying ACFs, which further indicates the existence of LRD.

Table 5.2 The Hurst parameters for the analyzed signals

<table>
<thead>
<tr>
<th>Signals</th>
<th>H</th>
<th>Signals</th>
<th>H</th>
<th>Signals</th>
<th>H</th>
<th>Signals</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABP1</td>
<td>0.8173</td>
<td>ABP9</td>
<td>0.8240</td>
<td>ABP17</td>
<td>0.8577</td>
<td>ABP25</td>
<td>0.8803</td>
</tr>
<tr>
<td>ABP2</td>
<td>0.8283</td>
<td>ABP10</td>
<td>0.9667</td>
<td>ABP18</td>
<td>0.8739</td>
<td>VIX</td>
<td>0.9619</td>
</tr>
<tr>
<td>ABP3</td>
<td>0.9365</td>
<td>ABP11</td>
<td>0.8674</td>
<td>ABP19</td>
<td>0.9266</td>
<td>VXO</td>
<td>0.9450</td>
</tr>
<tr>
<td>ABP4</td>
<td>0.8205</td>
<td>ABP12</td>
<td>0.8791</td>
<td>ABP20</td>
<td>0.7158</td>
<td>VXD</td>
<td>0.9606</td>
</tr>
<tr>
<td>ABP5</td>
<td>0.9437</td>
<td>ABP13</td>
<td>0.9069</td>
<td>ABP21</td>
<td>0.8815</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABP6</td>
<td>0.7993</td>
<td>ABP14</td>
<td>0.9692</td>
<td>ABP22</td>
<td>0.8647</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABP7</td>
<td>0.7962</td>
<td>ABP15</td>
<td>0.7960</td>
<td>ABP23</td>
<td>0.8531</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABP8</td>
<td>0.9318</td>
<td>ABP16</td>
<td>0.8937</td>
<td>ABP24</td>
<td>0.8082</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5.2 ACFs for ABP5 (left) and VIX (right)

5.2.2 Models and refinements

Recall the four models we have discussed in Chapter 4-FARIMA, HAR-RV, wavelet-based and average-VAR. According to Table 4.1, FARIMA, HAR-RV and average-VAR models can be directly applied to LRD signals, while the wavelet-based model requires the self-similar condition be satisfied. Benefiting from the close relationship between self-similar processes and LRD processes, we can transform a LRD signal into a self-similar one just by taking its partial sum. Since our data are LRD, the FARIMA, HAR-RV and average-VAR models are applied directly; for the wavelet-based model, we first take the partial sum, and then use the wavelet-based model to make forecasts for the cumulated series, and finally
difference the cumulated series to get forecasts for the original time series. The FARIMA model can be used without further refinements, while the other three models need more discussion.

5.2.2.1 Choosing optimal aggregate levels for the HAR-RV model and the average-VAR model

In the HAR-RV model, Corsi aggregated realized volatility on weekly and monthly basis, and then forecasted future daily realized volatility as a weighted sum of past daily, weekly and monthly volatilities [29]. The aggregate levels 1, 5, and 22 are reasonable because they have practical meanings – day, week and month. For volatility indexes, the same aggregate levels can be used. However, for arterial blood pressure signals, there are no aggregate levels of practical meaning. Therefore optimal aggregate levels need to be numerically searched for. Three scales are used to incorporate both short-term and long-term effects.

Scale 1: the most detailed scale (short-term scale), \( s_1 = 1 \).

Scale 2: mid-term scale, \( s_2 \in [4,12] \).

Scale 3: long-term scale, \( s_3 \in [12,25] \).

For a certain blood pressure signal, the HAR-RV model is applied with different aggregate level combinations. The aggregate levels are selected to lie in the above intervals and satisfy the condition \( s_1 - s_2 \geq 4 \) to guarantee that these two scales are significantly different in the time span. The Mean Squared Errors (MSEs) for different combinations are compared, and the one that has the minimum MSE is selected as the optimal combination. We randomly choose three signals (ABP5, ABP9 and ABP13) and find the optimal aggregate levels, as shown in Tables 5.3-5.5, respectively.

From the tables, the optimal aggregate levels for the tested signals can be obtained: 1, 5 and 25 in Table 5.3; 1, 9 and 23 in Table 5.4; 1, 5 and 25 in Table 5.5. In the meantime, it is easy to observe that the optimal aggregate levels are not always the same for different signals. Based on the fact that running the optimal level selection program for every signal is time-consuming and the MSEs do not differ much for all the combinations essentially, we determine to use the same aggregate levels for all the signals to be analyzed, i.e. 1, 5 and 25.
Table 5.3 MSEs for ABP5 for different combinations

<table>
<thead>
<tr>
<th></th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>22.274</td>
<td>22.223</td>
<td>22.177</td>
<td>22.138</td>
<td>22.105</td>
<td>22.081</td>
<td>22.058</td>
<td>22.043</td>
<td>22.031</td>
<td>22.019</td>
<td>22.011</td>
<td>22.007</td>
<td>22.010</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4 MSEs for ABP9 for different combinations

<table>
<thead>
<tr>
<th></th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
</table>
Table 5.5 MSEs for ABP13 for different combinations

<table>
<thead>
<tr>
<th></th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
</table>
Like the HAR-RV model, the average-VAR model also requires aggregation at the first step. In order to compare these two methods in a more reasonable way, the same aggregation levels are used for the average-VAR model. In theory, the aggregated sequences would be strongly correlated, and cross correlation is one measure of how strongly this correlation is. Therefore, the cross correlation (denoted as R) for the ABP5 and VIX signals between every two aggregated processes is calculated and shown below:

ABP5: \[ R = \begin{bmatrix} 1.0000 & 0.8997 & 0.7812 \\ 0.8997 & 1.0000 & 0.8701 \\ 0.7812 & 0.8701 & 1.0000 \end{bmatrix} \]

VIX: \[ R = \begin{bmatrix} 1.0000 & 0.9832 & 0.9388 \\ 0.9832 & 1.0000 & 0.9633 \\ 0.9388 & 0.9633 & 1.0000 \end{bmatrix} \]

The above results show clearly that there is a strong correlation between the aggregated processes; as a result, the VAR model can be used to describe this across-scale correlation.

### 5.2.2.2 Use of redundant wavelet transform in the wavelet-based model

In Chapter 4, we have mentioned that Wang and Lee [49] used a down sampling approach to deal with the inconsistency existing in the number of wavelet coefficients on different scales, and their algorithm can predict the average of the future four time periods. However, in our case, the time series are LRD rather than self-similar, thus must be cumulated first in order to use the wavelet-based model. So every point of the cumulated series is important and needs to be predicted in order for the differencing in the last step, and only knowing the average of every four points is not enough. Therefore, a redundant wavelet transform is used instead (see Fig. 5.3).

![Fig. 5.3 Across-scale prediction algorithm (redundant wavelet transform approach)](image)
Vectors of wavelet coefficients are created, and then the VAR(1) model is applied to predict the next vector. Instead of using the top level coefficients, the bottom level \( d_{l,n+1} \) is used to make a forecast for the next time period, or

\[
X_{n+1} = X_n - \frac{d_{l,n+1}}{2^{l/2}}
\]  

(5.1)

At last, the forecast for the original time series can be obtained by differencing the predicted cumulated series:

\[
Y_n = X_{n+1} - X_n
\]

After wavelet transform, the across-scale correlation should be strong according to Wang and Lee [49]. This can be tested by computing the cross correlation (denoted as R) between the wavelet coefficients on three scales. Here shows the results for ABP5 and VIX signals, respectively.

**ABP5:**
\[
\begin{array}{cccc}
1.0000 & 0.9358 & 0.8353 \\
0.9358 & 1.0000 & 0.9259 \\
0.8353 & 0.9259 & 1.0000 \\
\end{array}
\]

**VIX:**
\[
\begin{array}{cccc}
1.0000 & 0.9896 & 0.9725 \\
0.9896 & 1.0000 & 0.9882 \\
0.9725 & 0.9882 & 1.0000 \\
\end{array}
\]

It is easy to see that the across-scale correlation is indeed very strong, which allows us to use the VAR model later.

### 5.3 Performance evaluation criteria

#### 5.3.1 Error metrics

Suppose \( \{\hat{Y}_t\} \) is the forecasted sequence, \( \{Y_t\} \) is the actual sequence, and \( n \) is the length of the sequence. In order to compare the performance of the forecasting models, the following error metrics are used.

i. Mean squared error (MSE): \( \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2 \).

ii. Mean absolute percentage error (MAPE): \( \text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100\% \).
iii. Mean absolute error (MAE): \[ \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|. \]

The above three error measurements are commonly used in evaluating forecasting performance. In addition, a new measure is defined, which considers how accurate the model predicts the direction of future movement. In many real-life situations, we are not interested in the exact value of the next observation; instead we care more about the direction of the next movement. The percentage of right direction predictions (PRD) is calculated as the number of right direction predictions over the number of all direction predictions, or

\[ \text{PRD} = \frac{\text{number of right direction predictions}}{\text{number of all direction predictions}}. \]

### 5.3.2 ROC curve

A receiver operating characteristic (ROC) curve is used as a tool to determine the best threshold as well as evaluate the performance of a binary classifier [55]. It is a plot of true positive rate (TPR) versus false positive rate (FPR) as the threshold \( T \) is varying.

To better illustrate how ROC curve is obtained, consider the diagnosis of a certain disease as an example. There are two groups of subjects, and one of the two groups are healthy (negative), and the other group have a certain disease (positive). We want to separate the two groups and send out alarms in case of the disease. A physiological signal is measured, blood pressure for instance, and a threshold \( T \) is set. For persons with blood pressure higher than \( T \), they are treated as having the disease and an alarm will be sent out; for persons with blood pressure lower than \( T \), they are treated as healthy. If a person does suffer from the disease and the test emits an alarm, then this is called true alarm, which is referred to a true positive (TP) [55]. On the contrary, if a person is healthy but the test falsely emits an alarm, then this is called false alarm, which is referred to a false positive (FP) [55]. Then TPR can be obtained by dividing the number of true positives by the total number of actual positives, and FPR can be obtained by dividing the number of false positives by the total number of actual negatives [55]. Each threshold \( T \) corresponds to a pair of \((FPR, TPR)\). Changing thresholds, and a set of \((FPR, TPR)\) pairs can be obtained. The ROC curve is got by plotting \( TPR \) versus \( FPR \) [55]. Table 5.6 and Fig. 5.4 show an intuitive view of the previous procedure.
Table 5.6 An illustrative diagram of how TPR and FPR are calculated

<table>
<thead>
<tr>
<th>Actual condition</th>
<th>Test outcome</th>
<th>Test positive</th>
<th>Test negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (P)</td>
<td>True positive (TP)</td>
<td>False negative</td>
<td>False negative</td>
</tr>
<tr>
<td>Negative (N)</td>
<td>False positive (FP)</td>
<td>True negative</td>
<td></td>
</tr>
</tbody>
</table>

$TPR = \frac{TP}{P}$ $FPR = \frac{FP}{N}$

Fig. 5.4 Role of the threshold

Fig. 5.5 A typical ROC curve

Fig. 5.5 shows a typical ROC curve. The closer a point is to the up left corner, the better performance the point is represented. A random guess will lie on the diagonal line $y=x$ [55]. Area under the curve (AUC), which is a value between 0 and 1, is often used to evaluate the performance of a model. The larger the
area is, the better performance the model provides. A random guess lying on the diagonal line will result in an AUC of 0.5. If a model has an AUC less than 0.5, the inverse decision should be made [55]. For example, if a model has an AUC less than 0.5, and the model classifies a condition as positive, then we should reverse the decision and treat it as negative and do not send out an alarm.

5.4 Results and discussion

In summary, there are four models in total, FARIMA, HAR-RV, wavelet-based, and average-VAR. Each time series contains data points of 4000 or more. For each time series, the first 1000 data points are used as the historical data to derive a model, which is then used to make a one-step-ahead forecast. When a new data point is available, it is added to the historical database and used to update the model for future forecast. Matlab and R are used to realize the models and make predictions. At last, the forecasted results and the actual values are stored in the same file and compared for justifying the feasibility of the models. The numerical results will be shown in this section, including the comparisons of MSE, MAPE, MAE, PRD, as well as the ROC curve discussed in section 5.3.

5.4.1 Time series plot

5.4.1.1 Comparison of the plots

Figures 5.6 - 5.9 are prediction results for blood pressure signal ABP5 using the four models. In each figure, the red curve represents the prediction result and the blue one represents the actual signal. We also include a partially enlarged plot to facilitate the understanding.
Fig. 5.7 HAR-RV results for ABP5

Fig. 5.8 Wavelet-based results for ABP5

Fig. 5.9 Average-VAR results for ABP5
5.4.1.2 Discussion - the “shifting” behavior

From Figures 5.6 - 5.9, we observe a “shifting” behavior with all the four models, or there is always a lag between predictions and actual values. The predicted value is always a shift of the previous value plus a very small modification. If the current observation shows an increase with respect to the previous value, then the forecast for the next period also tends to be an increase, and vice versa. This phenomenon is reasonable since the latest observation and trend should have the strongest influence on the future observations, thus the forecast for the future value should be close to the most recent observation, and we should expect the trend would last into the future in case of no future information.

However, this “shifting” behavior is not desirable. Since we want to find the hidden information from the past observations and make use of it to make forecasts, the most ideal case is that we can foresee the future trend but not just replicate the past. The “shifting” behavior is somewhat a replicate of the past, which provides only slight improvements to the very simple model – directly use the past observation as the forecasted value for the future. Unfortunately, these four models cannot overcome this difficulty.

5.4.2 MSE, MAPE, MAE and PRD

5.4.2.1 Numerical results

As a measure of model complexity, the total computation time (in seconds) it took to perform all the modelling and forecasting is recorded for each data sequence. The error metrics (MSE, MAPE, MAE and PRD) and computation time for ABP5, ABP13 and VIX are listed in Tables 5.7 - 5.9, respectively. For full numerical results on the error metrics and computation time for other signals, please refer to Appendix A-1.

Table 5.7 Error metrics and computation time for ABP5 (H=0.9437)

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>25.5494</td>
<td>2.89</td>
<td>2.0749</td>
<td>0.4876</td>
<td>18224.264</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>24.5877</td>
<td>2.86</td>
<td>2.0561</td>
<td>0.4901</td>
<td>165.405</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>27.9540</td>
<td>2.86</td>
<td>2.0523</td>
<td>0.4799</td>
<td>488.376</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>24.6106</td>
<td>2.86</td>
<td>2.0570</td>
<td>0.4899</td>
<td>292.168</td>
</tr>
</tbody>
</table>
Chapter 5. Numerical Studies and Results

Table 5.8 Error metrics and computation time for ABP13 (H=0.9069)

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>61.4231</td>
<td>5.24</td>
<td>3.9933</td>
<td>0.4688</td>
<td>5143.031</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>59.5287</td>
<td>5.21</td>
<td>3.9776</td>
<td>0.4569</td>
<td>160.821</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>60.7234</td>
<td>5.09</td>
<td>3.9156</td>
<td>0.4552</td>
<td>299.964</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>56.0381</td>
<td>5.19</td>
<td>3.9429</td>
<td>0.4569</td>
<td>166.085</td>
</tr>
</tbody>
</table>

Table 5.9 Error metrics and computation time for VIX (H=0.9619)

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>2.1543</td>
<td>4.28</td>
<td>0.9140</td>
<td>0.4835</td>
<td>7036.005</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>2.1044</td>
<td>4.28</td>
<td>0.9107</td>
<td>0.4828</td>
<td>131.097</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>2.1235</td>
<td>4.27</td>
<td>0.9096</td>
<td>0.4846</td>
<td>410.840</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>2.1041</td>
<td>4.28</td>
<td>0.9105</td>
<td>0.4828</td>
<td>226.918</td>
</tr>
</tbody>
</table>

5.4.2.2 Discussion

From Tables 5.7 - 5.9, we find that the performances of the four models are almost the same in terms of MSE, MAPE, MAE, and PRD. FARIMA is by far the most time-consuming, whereas the HAR-RV model and the average-VAR model require the least computation. Therefore, the FARIMA model is not recommended for real-time application.

MAPE and PRD are used when comparing the forecast accuracy for different signals, since other error metrics are based on absolute values. We want to find out what properties of a signal may lead to a more accurate forecast. For a LRD signal, the Hurst parameter and the across-scale correlation are two most important features. The cross correlation between wavelet coefficients (denoted as $R_1$) and aggregated series (denoted as $R_2$) for ABP5, ABP13, and VIX are shown below:

**ABP5: (H=0.9437)**

$R_1 = 1.0000 \quad 0.9358 \quad 0.8353 \quad R_2 = 1.0000 \quad 0.8997 \quad 0.7812$

$0.9358 \quad 1.0000 \quad 0.9259 \quad 0.8997 \quad 1.0000 \quad 0.8701$

$0.8353 \quad 0.9259 \quad 1.0000 \quad 0.7812 \quad 0.8701 \quad 1.0000$

**ABP13: (H=0.9069)**

$R_1 = 1.0000 \quad 0.8887 \quad 0.7791 \quad R_2 = 1.0000 \quad 0.8552 \quad 0.6867$

$0.8887 \quad 1.0000 \quad 0.9044 \quad 0.8552 \quad 1.0000 \quad 0.8214$

$0.7791 \quad 0.9044 \quad 1.0000 \quad 0.6867 \quad 0.8214 \quad 1.0000$
CHAPTER 5. NUMERICAL STUDIES AND RESULTS

VIX: \( (H=0.9619) \)
\[
R_1 = \begin{pmatrix}
1.0000 & 0.9896 & 0.9725 \\
0.9896 & 1.0000 & 0.9882 \\
0.9725 & 0.9882 & 1.0000
\end{pmatrix}
\]
\[
R_2 = \begin{pmatrix}
1.0000 & 0.9832 & 0.9388 \\
0.9832 & 1.0000 & 0.9633 \\
0.9388 & 0.9633 & 1.0000
\end{pmatrix}
\]

From the above cross correlations as well as corresponding Hurst parameters, it is easy to find that larger Hurst parameters and stronger across-scale correlations do not necessarily result in a more accurate forecast, since the forecast results for ABP5 are the best among all the three signals in terms of MAPE, while its Hurst parameter is not the largest and its across-scale correlations are not the strongest. This observation is somewhat unexpected, because we were expecting that larger Hurst parameters would lead to more accurate forecasts.

5.4.3 ROC curve

5.4.3.1 Results

In order to use the ROC curve as an evaluation tool in our study, positive and negative conditions must be first defined. In the first case, we try to detect large increase. If the actual signal increases by 10%, it is considered as a large increase (positive). In the second case, the target is large decrease. Similarly, if the actual signal decreases by 10%, it is considered as a large decrease (positive). Afterwards, the ROC curves are obtained and their AUCs are compared. Figures 5.10 - 5.13 are the ROC curves and the corresponding AUCs for the ABP5 and VIX signals. Both increase detection and decrease detection are tested and compared.

Fig. 5.10 The ROC curve for ABP5 – increase detection
CHAPTER 5. NUMERICAL STUDIES AND RESULTS

Fig. 5.11 The ROC curve for VIX – increase detection

Fig. 5.12 The ROC curve for ABP5 – decrease detection

Fig. 5.13 The ROC curve for VIX – decrease detection
5.4.3.2 Discussion

From the ROC curves and the corresponding AUCs, it is observed that the HAR-RV model and the average-VAR model consistently outperform the other two models. In the meantime, in case of decrease detection, the AUCs well passed 0.5, while in case of increase detection, the AUCs barely exceeds 0.5, which indicates that decrease detection is more accurate than increase detection.

We want to find out the reason for this observation. Here is one possible explanation: since the forecast results always show a “shifting” behavior, if for these signals large decrease always occur after a previous decrease, then this may lead to the result that decrease detection is more accurate. To test this assumption, we computed the percentage of decreases followed by a large decrease (denoted as Dec-%), as well as percentage of increases followed by a large increase (denoted as Inc-%), and results can be found in Table 5.10.

Table 5.10 Percentage of increases followed by a large increase and decreases followed by a large decrease

<table>
<thead>
<tr>
<th></th>
<th>Dec-% (%)</th>
<th>Inc-% (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABP5</td>
<td>0.53</td>
<td>1.03</td>
</tr>
<tr>
<td>VIX</td>
<td>1.57</td>
<td>1.93</td>
</tr>
</tbody>
</table>

From Table 5.10, we can see the percentage of decreases followed by a large decrease is in fact less than that of increases followed by a large increase, which is against our assumption, thus this property does not have a direct impact on the observation that decrease detection is more accurate than increase detection. The reason for this still remains unknown and can be a future research topic.

5.5 Summary

In this chapter, we perform numerical studies to several real-life data sets – arterial blood pressure signals and volatility indexes. These signals are found to be LRD with Hurst parameters between 0.5 and 1. Then the four models discussed in the previous chapter are refined and applied to model the signals.

Through numerical studies, we have the following observations:

i. For all the four models, we have observed a “shifting” behavior – the forecasted trend for the future is mostly a replicate of the previous movement, which is not desirable.
ii. The performances of the four models are nearly the same in terms of MSE, MAPE, MAE, and PRD, while at times one model may slightly outperform the other three. However, FARIMA is so complicated and time-consuming that its computation time far exceeds the other three models, thus not recommended.

iii. From the ROC curves and the corresponding AUCs, we can find out that decrease detection is more accurate than increase detection, and HAR-RV model and average-VAR model consistently outperform the other two models.
Chapter 6 Conclusions and Future Work

6.1 Conclusions

In this thesis, self-similar and LRD processes are studied and some of the main properties are explored. One of the main contributions of this thesis is that it provides the readers a detailed and clear understanding of the concepts on self-similarity and LRD. These two concepts are closely related yet different in theory. Self-similar signals are non-stationary, while LRD processes are stationary. They are related by the same Hurst parameter, and LRD processes are the increment series of self-similar processes with Hurst parameters between 0.5 and 1. Clearly distinguishing these two kinds of processes and understanding their differences and relationship are of great importance in the application of these two processes.

After the theory of self-similarity and LRD processes is illustrated, several models are presented to analyze some real-life signals as well as forecast their future values. The data we use include arterial blood pressure signals obtained from PhysioNet [43], and volatility indexes obtained from CBOE website [54]. These signals are closely related to people’s everyday life and making forecast for them is quite useful. There are four models in total, FARIMA, HAR-RV, wavelet-based and average-VAR. The FARIMA model has been widely used in modelling LRD signals, but it is very complicated and time-consuming, and its accuracy is not superior to the other three models. The HAR-RV model was originally proposed to model the LRD in realized volatility, and when extended to model physiological signals, the performance was found very good with little computation time. The wavelet-based model was first proposed to model self-similar processes, but unfortunately, many signals in real practice are actually LRD rather than self-similar. Thus some refinements are made and a modified model is generalized for better analysis of LRD signals. Though the results seem not so satisfactory, this provides the researchers with some insight for future study. The average-VAR model is a new model proposed in this thesis that utilizes the scaling property of LRD signals. This model is theoretically straightforward and can provide quite good performance compared with other models. In summary, the performances of these four models are similar, and one model may perform slightly better than other models in terms of one of the error metrics, while worse than the others in terms of another error metric, for unknown reasons. However, the FARIMA model is not recommended for its complexity and large computation time.
We also compared the ROC curves and found that detecting large decrease is more accurate than detecting large increase, with the HAR-RV model and average-VAR model performing the best, though the reasons are unknown yet.

### 6.2 Future work

The analysis and forecast of self-similar and LRD signals are very important. Future work remains to be done in the following areas:

i. When forecasts and actual signals are compared in the same figure in Chapter 5, we observe that there is always a lag between the forecasted result and the actual value. The four models discussed in this thesis cannot overcome this problem. This is considered as a future research topic in terms of preventing the “shifting” behavior in forecasting self-similar and LRD signals.

ii. When performing optimal aggregate level selection program to different arterial blood pressure signals, we find that the optimal aggregate levels for different patients are not always the same. Since for realized volatility, the cycle lengths are 1, 5, and 22, which are of practical meaning, we believe the optimal cycle lengths for human beings should also have practical use, for example, it may have something to do with the persons’ health conditions. Further exploring the meaning of optimal cycle lengths for human beings can be a useful research topic.

iii. We have tried to understand the inherent relationship between the Hurst parameter, across-scale correlation and prediction accuracy, but failed. Future studies can learn from our failure and continue our research to find out their relationship that can improve the forecast accuracy.

iv. We still do not know the reason why decrease detection is more accurate than increase detection. In the future, more properties of self-similar and LRD signals can be explored, and the intrinsic feature hidden behind that leads to this result can thus be found out.
References


REFERENCES


REFERENCES


REFERENCES


Appendix A-1 Numerical studies for other signals

In Appendix A-1, numerical studies for other signals (excluding ABP5, ABP13 and VIX) are presented, including the cross correlation between wavelet coefficients (denoted as $R_1$) and aggregated series (denoted as $R_2$), error metrics and computation time.

**ABP1:**

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>31.0527</td>
<td>3.43</td>
<td>2.6042</td>
<td>0.5736</td>
<td>12579.114</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>30.8263</td>
<td>3.40</td>
<td>2.5781</td>
<td>0.5749</td>
<td>180.348</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>30.7955</td>
<td>3.02</td>
<td>2.2974</td>
<td>0.5724</td>
<td>326.285</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>30.8394</td>
<td>3.40</td>
<td>2.5787</td>
<td>0.5752</td>
<td>182.372</td>
</tr>
</tbody>
</table>

**ABP2:**

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>127.2004</td>
<td>3.91</td>
<td>3.3595</td>
<td>0.3903</td>
<td>3927.518</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>126.3399</td>
<td>3.96</td>
<td>3.3901</td>
<td>0.3912</td>
<td>151.576</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>130.6242</td>
<td>3.87</td>
<td>3.3071</td>
<td>0.3881</td>
<td>287.152</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>126.4804</td>
<td>3.96</td>
<td>3.3930</td>
<td>0.3914</td>
<td>149.253</td>
</tr>
</tbody>
</table>

**ABP3:**

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>25.9199</td>
<td>4.28</td>
<td>2.2433</td>
<td>0.4879</td>
<td>6941.422</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>25.7779</td>
<td>4.25</td>
<td>2.2191</td>
<td>0.4839</td>
<td>198.191</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>26.7133</td>
<td>4.23</td>
<td>2.2038</td>
<td>0.4875</td>
<td>352.119</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>25.7899</td>
<td>4.25</td>
<td>2.2200</td>
<td>0.4839</td>
<td>192.554</td>
</tr>
</tbody>
</table>
### ABP4:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>24.5403</td>
<td>9.69</td>
<td>2.1807</td>
<td>0.4338</td>
<td>4432.367</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>24.5790</td>
<td>9.59</td>
<td>2.1721</td>
<td>0.4346</td>
<td>201.677</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>25.3349</td>
<td>9.72</td>
<td>2.1978</td>
<td>0.4330</td>
<td>365.029</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>24.6005</td>
<td>9.60</td>
<td>2.1713</td>
<td>0.4343</td>
<td>203.141</td>
</tr>
</tbody>
</table>

### ABP6:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>218.6123</td>
<td>4.40</td>
<td>4.8324</td>
<td>0.4325</td>
<td>13111.155</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>226.9618</td>
<td>4.37</td>
<td>4.8451</td>
<td>0.4340</td>
<td>191.558</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>224.4905</td>
<td>4.43</td>
<td>4.7662</td>
<td>0.4278</td>
<td>312.117</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>227.1848</td>
<td>4.38</td>
<td>4.8488</td>
<td>0.4334</td>
<td>179.426</td>
</tr>
</tbody>
</table>

### ABP7:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>111.1230</td>
<td>4.05</td>
<td>3.2289</td>
<td>0.4484</td>
<td>7923.359</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>110.7397</td>
<td>4.07</td>
<td>3.2295</td>
<td>0.4459</td>
<td>294.096</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>115.3516</td>
<td>4.16</td>
<td>3.3050</td>
<td>0.4440</td>
<td>515.449</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>110.7776</td>
<td>4.08</td>
<td>3.2302</td>
<td>0.4461</td>
<td>294.759</td>
</tr>
</tbody>
</table>

### ABP8:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>295.4171</td>
<td>5.98</td>
<td>6.9722</td>
<td>0.4297</td>
<td>16272.521</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>302.8970</td>
<td>5.89</td>
<td>6.8770</td>
<td>0.4365</td>
<td>156.117</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>288.9746</td>
<td>5.31</td>
<td>6.1951</td>
<td>0.4434</td>
<td>331.460</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>302.8971</td>
<td>5.89</td>
<td>6.8770</td>
<td>0.4360</td>
<td>184.560</td>
</tr>
</tbody>
</table>
### Appendix A-1

#### ABP9:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>15.0458</td>
<td>3.37</td>
<td>1.6937</td>
<td>0.4500</td>
<td>5248.774</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>14.8616</td>
<td>3.32</td>
<td>1.6554</td>
<td>0.4425</td>
<td>98.691</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>15.7767</td>
<td>3.32</td>
<td>1.6471</td>
<td>0.4435</td>
<td>267.3180</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>14.8696</td>
<td>3.32</td>
<td>1.6559</td>
<td>0.4425</td>
<td>149.608</td>
</tr>
</tbody>
</table>

#### ABP10:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>20.8981</td>
<td>2.89</td>
<td>2.4168</td>
<td>0.5081</td>
<td>3577.794</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>20.7130</td>
<td>2.85</td>
<td>2.3881</td>
<td>0.5107</td>
<td>136.501</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>22.0433</td>
<td>2.93</td>
<td>2.4434</td>
<td>0.5074</td>
<td>265.601</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>20.7336</td>
<td>2.85</td>
<td>2.3895</td>
<td>0.5107</td>
<td>148.597</td>
</tr>
</tbody>
</table>

#### ABP11:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>59.1668</td>
<td>3.17</td>
<td>2.3954</td>
<td>0.4477</td>
<td>6068.300</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>57.5167</td>
<td>3.20</td>
<td>2.4061</td>
<td>0.4435</td>
<td>207.085</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>148.3635</td>
<td>3.17</td>
<td>2.5647</td>
<td>0.4452</td>
<td>1737.454</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>57.4682</td>
<td>3.13</td>
<td>2.3969</td>
<td>0.4436</td>
<td>229.592</td>
</tr>
</tbody>
</table>

#### ABP12:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>45.3002</td>
<td>2.72</td>
<td>2.4227</td>
<td>0.5358</td>
<td>17038.745</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>40.7972</td>
<td>2.72</td>
<td>2.4207</td>
<td>0.5375</td>
<td>191.689</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>41.7143</td>
<td>2.81</td>
<td>2.4855</td>
<td>0.5302</td>
<td>327.814</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>40.5100</td>
<td>2.71</td>
<td>2.4124</td>
<td>0.5376</td>
<td>202.599</td>
</tr>
</tbody>
</table>
## Appendix A-1

### ABP14:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>15.7986</td>
<td>1.74</td>
<td>1.6449</td>
<td>0.4808</td>
<td>4513.548</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>15.3416</td>
<td>1.70</td>
<td>1.6098</td>
<td>0.4766</td>
<td>197.642</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>16.2635</td>
<td>1.73</td>
<td>1.6309</td>
<td>0.4756</td>
<td>343.670</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>15.3489</td>
<td>1.70</td>
<td>1.6105</td>
<td>0.4766</td>
<td>189.699</td>
</tr>
</tbody>
</table>

### ABP15:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>55.3460</td>
<td>3.25</td>
<td>2.6240</td>
<td>0.4433</td>
<td>10425.935</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>57.2400</td>
<td>3.31</td>
<td>2.6682</td>
<td>0.4281</td>
<td>271.586</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>60.8650</td>
<td>3.01</td>
<td>2.4944</td>
<td>0.4374</td>
<td>425.150</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>57.2812</td>
<td>3.31</td>
<td>2.6693</td>
<td>0.4280</td>
<td>272.875</td>
</tr>
</tbody>
</table>

### ABP16:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>6.9450</td>
<td>2.04</td>
<td>1.5491</td>
<td>0.4402</td>
<td>14308.108</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>6.9252</td>
<td>2.00</td>
<td>1.5200</td>
<td>0.4375</td>
<td>276.705</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>7.1378</td>
<td>2.03</td>
<td>1.5490</td>
<td>0.4376</td>
<td>486.510</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>6.9252</td>
<td>2.00</td>
<td>1.5200</td>
<td>0.4375</td>
<td>284.473</td>
</tr>
</tbody>
</table>

### ABP17:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>39.1270</td>
<td>3.44</td>
<td>2.5835</td>
<td>0.4757</td>
<td>23204.128</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>39.9816</td>
<td>3.23</td>
<td>2.4363</td>
<td>0.4751</td>
<td>280.371</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>44.7255</td>
<td>3.00</td>
<td>2.3048</td>
<td>0.4662</td>
<td>439.137</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>39.9922</td>
<td>3.23</td>
<td>2.4360</td>
<td>0.4751</td>
<td>267.963</td>
</tr>
</tbody>
</table>
### ABP18:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>20.4930</td>
<td>2.98</td>
<td>1.7897</td>
<td>0.4585</td>
<td>47560.931</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>19.9426</td>
<td>2.92</td>
<td>1.7493</td>
<td>0.4563</td>
<td>319.0100</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>22.4982</td>
<td>2.81</td>
<td>1.6579</td>
<td>0.4537</td>
<td>568.963</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>19.9483</td>
<td>2.92</td>
<td>1.7494</td>
<td>0.4563</td>
<td>351.037</td>
</tr>
</tbody>
</table>

### ABP19:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>12.5166</td>
<td>2.12</td>
<td>1.7259</td>
<td>0.5277</td>
<td>9652.555</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>11.7999</td>
<td>2.11</td>
<td>1.7157</td>
<td>0.5344</td>
<td>262.717</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>12.2865</td>
<td>2.11</td>
<td>1.7222</td>
<td>0.5356</td>
<td>525.385</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>11.8069</td>
<td>2.11</td>
<td>1.7154</td>
<td>0.5345</td>
<td>298.135</td>
</tr>
</tbody>
</table>

### ABP20:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>50.0799</td>
<td>2.61</td>
<td>1.9490</td>
<td>0.5023</td>
<td>23605.208</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>49.2468</td>
<td>2.49</td>
<td>1.8707</td>
<td>0.5107</td>
<td>286.449</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>55.6287</td>
<td>2.56</td>
<td>1.9240</td>
<td>0.5111</td>
<td>602.167</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>49.2744</td>
<td>2.49</td>
<td>1.8707</td>
<td>0.5107</td>
<td>351.155</td>
</tr>
</tbody>
</table>

### ABP21:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>32.7030</td>
<td>3.79</td>
<td>3.0335</td>
<td>0.4801</td>
<td>16462.217</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>32.5337</td>
<td>3.76</td>
<td>3.0100</td>
<td>0.4787</td>
<td>347.445</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>34.3381</td>
<td>3.60</td>
<td>2.9115</td>
<td>0.4757</td>
<td>650.020</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>32.3133</td>
<td>3.73</td>
<td>3.0030</td>
<td>0.4785</td>
<td>408.832</td>
</tr>
</tbody>
</table>
### ABP22:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>80.0292</td>
<td>3.14</td>
<td>3.1507</td>
<td>0.5552</td>
<td>10830.181</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>79.8128</td>
<td>3.00</td>
<td>3.0344</td>
<td>0.5528</td>
<td>270.570</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>83.9977</td>
<td>2.96</td>
<td>2.9674</td>
<td>0.5532</td>
<td>572.203</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>79.8128</td>
<td>3.00</td>
<td>3.0344</td>
<td>0.5528</td>
<td>330.772</td>
</tr>
</tbody>
</table>

### ABP23:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>38.6545</td>
<td>2.91</td>
<td>2.7875</td>
<td>0.5186</td>
<td>26785.252</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>38.8853</td>
<td>2.77</td>
<td>2.6834</td>
<td>0.5172</td>
<td>309.629</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>39.5761</td>
<td>2.78</td>
<td>2.6817</td>
<td>0.5176</td>
<td>629.394</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>38.8958</td>
<td>2.77</td>
<td>2.6837</td>
<td>0.5174</td>
<td>386.321</td>
</tr>
</tbody>
</table>

### ABP24:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>51.4489</td>
<td>6.66</td>
<td>4.5174</td>
<td>0.5274</td>
<td>16242.871</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>51.4828</td>
<td>6.55</td>
<td>4.4683</td>
<td>0.5245</td>
<td>334.539</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>51.0191</td>
<td>5.84</td>
<td>4.3517</td>
<td>0.5265</td>
<td>678.355</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>51.4639</td>
<td>6.55</td>
<td>4.4665</td>
<td>0.5244</td>
<td>406.898</td>
</tr>
</tbody>
</table>

### ABP25:

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>129.4227</td>
<td>10.71</td>
<td>4.6933</td>
<td>0.5281</td>
<td>7547.323</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>129.7750</td>
<td>10.34</td>
<td>4.1977</td>
<td>0.5307</td>
<td>476.021</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>128.1970</td>
<td>10.64</td>
<td>4.5201</td>
<td>0.5320</td>
<td>287.783</td>
</tr>
</tbody>
</table>

65
### Appendix A-1

#### VXD:

\[
\begin{array}{ccc}
R_1 &=& 1.0000 \quad 0.9892 \quad 0.9717 \\
0.9892 &=& 1.0000 \quad 0.9879 \\
0.9717 &=& 0.9879 \quad 1.0000 \\
\end{array}
\]

\[
\begin{array}{ccccc}
\text{R}_2 &=& 1.0000 \quad 0.9840 \quad 0.9397 \\
0.9840 &=& 1.0000 \quad 0.9637 \\
0.9397 &=& 0.9637 \quad 1.0000 \\
\end{array}
\]

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>2.5382</td>
<td>4.63</td>
<td>0.9613</td>
<td>0.4825</td>
<td>5491.320</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>2.4758</td>
<td>4.58</td>
<td>0.9499</td>
<td>0.4845</td>
<td>105.055</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>2.5243</td>
<td>4.60</td>
<td>0.9622</td>
<td>0.4788</td>
<td>234.311</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>2.4788</td>
<td>4.58</td>
<td>0.9507</td>
<td>0.4845</td>
<td>127.174</td>
</tr>
</tbody>
</table>

#### VXO:

\[
\begin{array}{ccc}
R_1 &=& 1.0000 \quad 0.9801 \quad 0.9532 \\
0.9801 &=& 1.0000 \quad 0.9802 \\
0.9532 &=& 0.9802 \quad 1.0000 \\
\end{array}
\]

\[
\begin{array}{ccccc}
\text{R}_2 &=& 1.0000 \quad 0.9710 \quad 0.9049 \\
0.9710 &=& 1.0000 \quad 0.9435 \\
0.9049 &=& 0.9435 \quad 1.0000 \\
\end{array}
\]

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>MAPE (%)</th>
<th>MAE</th>
<th>PRD</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARIMA</td>
<td>3.2226</td>
<td>5.12</td>
<td>1.1165</td>
<td>0.4737</td>
<td>21016.184</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>2.9189</td>
<td>4.88</td>
<td>1.0601</td>
<td>0.4676</td>
<td>247.093</td>
</tr>
<tr>
<td>Wavelet-based</td>
<td>2.9456</td>
<td>4.74</td>
<td>1.0483</td>
<td>0.4703</td>
<td>554.363</td>
</tr>
<tr>
<td>Average-VAR</td>
<td>2.9207</td>
<td>4.88</td>
<td>1.0605</td>
<td>0.4676</td>
<td>319.879</td>
</tr>
</tbody>
</table>