STATISTICAL SELF-SIMILARITY IN TIME SERIES FROM FINANCIAL DATA & CHAOTIC DYNAMICAL SYSTEMS

BY

PANPAN ZHANG

A Thesis Submitted to the Graduate Faculty of

WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF ARTS

Mathematics

May 2012

Winston-Salem, North Carolina

Approved By:

Miaohua Jiang, Ph.D., Advisor

James Norris, Ph.D., Chair

Jennifer Erway, Ph.D.

Acknowledgments

There are many people who helped me with this thesis. Many thanks to my advisor, Dr. Miaohua Jiang, for picking up the topic and helping me with the difficulties in the research every week; thanks to Dr. James Norris and Dr. Kenneth Berenhaut for helps with probability theory and statistics parts in the thesis; thanks to Dr. Jennifer Erway for helps with coding in MATLAB; thanks to Steven Wicker, Katie Novacek for helps with LaTex.

Table of Contents

Abstract

Panpan Zhang

In this paper, I am going to introduce statistical self-similarity for discrete time series. My thesis is divided into three parts:

In the first part, I will give a mathematical definition of self-similarity and detect the main properties of self-similar processes. At the end of this part, fractional Brownian motion(fBm) will be used as an example to check these properties.

In the second part, Discrete Wavelet Transform(DWT) and wavelet coefficients will be introduced. The way to construct wavelet coefficients and their natural properties which are especially related to self-similarity will also be presented. Daubechies wavelet family will be proposed as an example, and the Haar wavelet from Daubechies family is used for further study.

Finally, I will use wavelet transform methods to determine if certain time series from financial data and chaotic dynamical systems are self-similar processes with stationary increments. The whole process of check will be run in MATLAB.

Chapter 1: Introduction

In general, self-similarity occurs when the shape of an object is similar, or approximately similar to the part of itself. That is each portion of the self-similar object can be considered as a reduced scale of the whole. In natural science, there are a lot of examples of self-similarity, such as fern, snowflake, mountain ridges, coastlines etc. From a view of statistics, self-similarity is defined as two subsets of the whole set are invariant in statistical distribution on different scales. In other words, self-similarity implies being rescaled or translated on the original signal keeps the same statistical distribution. As a fresh and interesting topic, many studies on self-similarity have been done in different fields, such as magnetohydrodynamics(MHD) in physics, human nervous system in biology, stellar fragments in astronomy and option pricing in finance etc. In historical mathematics, the idea of self-similarity was first introduced in 1967 by Mandelbrot[2], when he was analysing the shape of British coast. In 1975, he solidified many researches on "Fractal"and illustrated the mathematical definition of self-similarity. However, self-similarity is referred to as self-affinity sometimes if the fractal is scaled by different amount in different directions. In this paper, self-similarity can be detected by the Discrete Wavelet Transform (DWT). Wavelet analysis began with Fourier with his theories of Fourier synthesis, and first wavelet was declared by Haar in 1909. Levy explored the Haar wavelet basis function and investigated Brownian motion in 1930. In 1980s, Mallat discovered some relations between wavelet bases and filters and introduced the wavelet decomposition algorithm and basic Multiresolution Analysis (MRA). In the early 1990's, Daubechies constructed a series orthonormal basis functions based on Mallat's work results. The Haar wavelet from Daubechies wavelet family will be used to determine self-similarity of time series in this paper.

Chapter 2: Self-similarity and its properties

2.1 Definition of self-similarity

We begin by defining self-similarity.

Definition 1. *Let $Z = \{Z(t)\}_{t\in\mathbb{R}}$ be a real-valued stochastic process on the real line, we say $Z = \{Z(t)\}_{t\in\mathbb{R}}$ is self-similar with parameter $H > 0$, if*

$$\{Z(at)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{a^H Z(t)\}_{t\in\mathbb{R}} \tag{2.1}$$

for any $a > 0$, where $\stackrel{d}{=}$ denotes the equality of the finite-dimensional distribution. And $Z = \{Z(t)\}_{t\in\mathbb{R}}$ is also called as H-ss process.

Definition 2. *A real-valued process $\{Z(t)\}_{t\in\mathbb{R}}$ has stationary increments, if*

$$\{Z(t+h) - Z(h)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{Z(t) - Z(0)\}_{t\in\mathbb{R}} \tag{2.2}$$

for all $h \in \mathbb{R}$

Remark 1. *If $\{Z(t)\}_{t\in\mathbb{R}}$ is self-similar, then:*
(1) $\{Z(t)\}_{t\geqslant 0}$ is defined, and the finite-dimensional distribution of $\{Z(t)\}_{t\geqslant 0}$ can be determined on the intervals of finite length on the positive real line.
(2) A non-degenerate H-ss process $\{Z(t)\}_{t\in\mathbb{R}}$ is not stationary.

In part(2) of **Remark 1**, a stochastic process $Z = \{Z(t)\}_{t\in\mathbb{R}}$ is called *stationary* if $\{Z(t)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{Z(t+h)\}_{t\in\mathbb{R}}$ for all $h \in \mathbb{R}$.

Proposition 1. *If the process $\{X(t)\}_{t>0}$ is H-ss, then the process $Y(t) = e^{-tH}X(e^t)$ is stationary, where $t \in \mathbb{R}$. Conversely, if the process $\{Y(t)\}_{t\in\mathbb{R}}$ is stationary, then the process $X(t) = t^H Y(\ln t)$ is H-ss, where $t > 0$.*

For a proof of **Proposition 1**, see Taqqu[3]. From **Proposition 1**, we can see that there exist different types of self-similar processes, and stationary process is one of them. In this paper, I am focusing on the self-similar processes with stationary increments.

If there exists a process $\{Z(t)\}_{t\in\mathbb{R}}$ satisfying both Relation (2.1) and (2.2), we call it *H-sssi* process, with definition as follows:

Definition 3. *If a real-valued process $Z = \{Z(t)\}_{t\in\mathbb{R}}$ is self-similar with parameter H and has stationary increments, it is called H-sssi process.*

In the rest of this paper, instead of working on simple self-similar process, I will focus more on $H - sssi$ process. $H - sssi$ process has more nice properties and they will be presented in the next section.

2.2 Properties of self-similarity

In this section, I will deduce some important properties of $H - sssi$ process, and they are the keys to construct such process and detect self-similar patterns of different time series in the future work. In addition, I will simplify notation $\{Z(t)\}_{t\in\mathbb{R}}$ into $Z(t)$. Before attaining the properties of $H - sssi$ processes, we need to assume that the process $Z(t)$ has finite variance.

Definition 4. *A process $Z(t)$ has finite variance if $E[Z^2(t)] < \infty$ for each $t \in \mathbb{R}$.*

Here, the condition $E[Z^2(t)] < \infty$ is strong enough to guarantee that the variance of process $Z(t)$ is finite, since:

$$\begin{aligned} E[Z(t)] &= \int_\Omega Z(t)d\omega \\ &\leqslant (\int_\Omega Z^2(t)d\omega))^{1/2}(\int_\Omega 1d\omega)^{1/2} \\ &< \infty \end{aligned}$$

for each $t \in \mathbb{R}$, where Ω is a given probability space. Hence, $Var[Z(t)] = E[Z^2(t)] - (E[Z(t)])^2 < \infty$. That is to say, my study on $H - sssi$ process will be restricted in l^2 space,

Proposition 2. *If $Z(t)$ is an H-sssi process with finite variance, then:*
(1) $Z(0) = 0$
(2) If $H \neq 1$, then $E[Z(t)] = 0$ for all $t \in \mathbb{R}$
(3) $Z(-t) = -Z(t)$
(4) $E[Z^2(t)] = |t|^{2H}\sigma^2$, where $\sigma^2 = E[Z^2(1)]$
(5) The covariance function is given by

$$\Gamma_H(s,t) = \frac{\sigma^2}{2}\{|s|^{2H} + |t|^{2H} - |t-s|^{2H}\}$$

for all $s, t \in \mathbb{R}$.
(6)The self-similarity parameter H of $Z(t)$ satisfies $H \leqslant 1$; and it is trivial when $H = 1$.

I will give some brief explanations of each part in **Proposition 2**: Part(1) and (2) can be directly checked by **Definition 1** and **2**; Part(3) is obvious from the result of Part(1) and the definition of stationary increments; Part(4) is mainly induced from Part(3), as well as the definition of self-similarity; Part(5) can be easily got from the result of Part(4), and it is the key to construct an $H - sssi$ process, which will be introduced in the later section and the proof will also be given; Part(6) gives a restriction of index H, and it is obvious that it is trivial when $H = 1$, since it follows $Z(t) = tZ(1)$, for all $t \in \mathbb{R}$.

So far, we have known the basic properties of self-similarity. However, we still need to show that such $H - sssi$ process does exist. We have the following proposition to support the existence of $H - sssi$ process.

Proposition 3. *If $0 < H \leqslant 1$, then the function*

$$R_H(s,t) = |s|^{2H} + |t|^{2H} - |t-s|^{2H}, s, t \in \mathbb{R}$$

is non-negative definite.

Proof. We can see that $R_H(s,t)$ is the main part of covariance function Relation(2.3) and we want to show $R_H(s,t) = |s|^{2H} + |t|^{2H} - |t-s|^{2H}$,$s, t \in \mathbb{R}$ is non-negative definite. It is equivalent to show that $\sum_{i=1}^{n}\sum_{j=1}^{n} R_H(t_i,t_j)u_iu_j \geqslant 0$, for $\forall t_1, t_2, ..., t_n$ and $u_1, u_2, ..., u_n \in \mathbb{R}$. The proof can be divided into three main steps.

Step 1:

Let the origin $t_0 = 0$ and $u_0 = -\sum_{i=1}^{n} u_i$, so it follows $\sum_{i=0}^{n} u_i = u_0 + \sum_{i=1}^{n} u_i = 0$ at t_0. Then we have:

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} |t_i|^{2H} u_i u_j &= \sum_{i=1}^{n} |t_i|^{2H} u_i \sum_{j=1}^{n} u_j \\
&= -\sum_{i=1}^{n} |t_i|^{2H} u_i u_0 \\
&= -\sum_{i=1}^{n} |t_i - t_0|^{2H} u_i u_0.
\end{aligned}$$

Similarly, $\sum_{i=1}^{n}\sum_{j=1}^{n} |t_j|^{2H} u_i u_j = -\sum_{j=1}^{n} |t_j - t_0|^{2H} u_j u_0$.

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} R_H(t_i,t_j)u_iu_j &= \sum_{i=1}^{n}\sum_{j=1}^{n}(|t_i|^{2H} + |t_j|^{2H} - |t_i - t_j|^{2H})u_iu_j \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n} |t_i - t_j|^{2H} u_i u_j - \sum_{i=1}^{n} |t_i - t_0|^{2H} u_i u_0 - \sum_{j=1}^{n} |t_j - t_0|^{2H} u_j u_0 \\
&= -\sum_{i=0}^{n}\sum_{j=0}^{n} |t_i - t_j|^{2H} u_i u_j.
\end{aligned}$$

Step 2: Then, let $\forall c > 0$, consider

$$\begin{aligned}\sum_{i=0}^{n}\sum_{j=0}^{n} e^{|t_i-t_j|^{2H}} u_i u_j &= \sum_{i=0}^{n}\sum_{j=0}^{n} e^{|t_i-t_j|^{2H}} u_i u_j - \sum_{i=0}^{n}\sum_{j=0}^{n} u_i u_j \\ &= \sum_{i=0}^{n}\sum_{j=0}^{n} (e^{|t_i-t_j|^{2H}} - 1) u_i u_j.\end{aligned}$$

Then, according to Taylor Expansion, $e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + ...$, so we have

$$\begin{aligned}\sum_{i=0}^{n}\sum_{j=0}^{n} (e^{|t_i-t_j|^{2H}} - 1) u_i u_j &= \sum_{i=0}^{n}\sum_{j=0}^{n} (1 - c|t_i - t_j|^{2H} + o(t_i, t_j, u_i, u_j)) - 1) u_i u_j \\ &= -c \sum_{i=0}^{n}\sum_{j=0}^{n} |t_i - t_j|^{2H} u_i u_j \\ &= c \sum_{i=1}^{n}\sum_{j=1}^{n} R_H(t_i, t_j) u_i u_j.\end{aligned}$$

That is, our target turns to show $\sum_{i=0}^{n}\sum_{j=0}^{n} e^{|t_i-t_j|^{2H}} u_i u_j \geqslant 0$.

Step 3:

Since $e^{c|t|^{2H}}$, where $0 < H \leqslant 1$, is the characteristic function of a $2H$ stable random variable ξ with scaling $c^{1/2H}$(see Morlan[5]), so we have:

$$E[e^{it\xi}] = e^{-c|t|^{2H}}, for \forall t \in \mathbb{R}.$$

Hence, let $t_i > t_j$, and we can get

$$\begin{aligned}
\sum_{i=0}^{n}\sum_{j=0}^{n} e^{|t_i-t_j|^{2H}} u_i u_j &= \sum_{i=0}^{n}\sum_{j=0}^{n} E[e^{i|t_i-t_j|\xi}] u_i u_j \\
&= E[\sum_{i=0}^{n}\sum_{j=0}^{n} e^{i|t_i-t_j|\xi} u_i u_j] \\
&= E[\sum_{i=0}^{n}\sum_{j=0}^{n} e^{it_i\xi} e^{-it_j\xi} u_i u_j] \\
&= E[\sum_{i=0}^{n} u_i e^{it_i\xi} \sum_{j=0}^{n} u_j e^{-it_j\xi}] \\
&= E[|\sum_{i=0}^{n} u_i e^{it_i\xi}|^2] \geqslant 0.
\end{aligned}$$

The last step above is true since $e^{it_i\xi}$ and $e^{-it_j\xi}$ are conjugate. □

2.3 fractional Brownian motion

In the last section, we have attained a series of properties of $H - sssi$ process, and also proved $H - sssi$ in fact exists. In this section, I will introduce a typical and famous $H - sssi$ process, fractional Brownian motion (fBm). Here, I need to give the definition of Gaussian process first, since it turns to be a part of the definition of fBm.

Definition 5. *Let $Z(t)$ be a stochastic process, if any finite linear combination of the sample data in $Z(t)$ are normally distributed, $Z(t)$ is called as Gaussian process.*

If a Gaussian process has mean zero, it is true that such Gaussian process is determined entirely by its variance and covariance function. That is, an $H - sssi$ zero mean Gaussian process is entirely differed by relation (2.3). Now, I can give the definition of fBm as follows:

Definition 6. *Fractional Brownian motion (fBm) is a Gaussian H-sssi zero mean process with parameter H satisfying $0 < H < 1$, and it is notated by $\{B_H(t)\}_{t\in\mathbb{R}}$.*

Proposition 4. *Suppose a stochastic process $\{Z(t)\}_{t\in\mathbb{R}}$ satisfies the following conditions:*
(1) Gaussian process with mean zero and $Z(0) = 0$,
(2) $E[Z^2(t)] = \sigma^2|t|^{2H}$ for some $\sigma > 0$, and $0 < H < 1$,
(3) Stationary increments.
Then $\{Z(t)\}_{t\in\mathbb{R}}$ is fractional Brownian motion.

Proof. The definition of fBm has been given in **Definition 6**, so our proof turns to check if the three conditions above are equivalent the conditions in **Definition 6**. In addition, the check of self-similarity part needs to be related with **Proposition 2** in the last section. I will check the conditions one by one:
Condition (1):
Let $Z(t)$ be a stochastic process, the first part of condition(1) indicates that $Z(t)$ is a Gaussian having mean zero, satisfying the basic construction of fBm. Moreover, mean zero corresponds to part(2) in **Proposition 2**, while the second part $Z(0) = 0$ corresponding to part(1). Furthermore, as we discussed previously, condition(1) implies the mean function of Gaussian process $Z(t)$ is fixed, so the process now is uniquely characterized by its variance and covariance function.
condition (2):
Firstly, it is obvious that the latter part $0 < H < 1$ corresponds to part(6) in **Proposition 2**. Meanwhile, the trivial case of $H = 1$ is excluded over here. Secondly, the statement $E[Z^2(t)] = \sigma^2|t|^{2H}$ corresponds to part(4) in **Proposition 2**, which is the variance function of self-similar process. Thirdly, upon the variance function, I will deduce the covariance function(part(5) in **Proposition 2**) and use it to confirm

self-similarity. According to condition (2), for $\forall s, t \in \mathbb{R}$, we have:

$$\begin{aligned} E[Z^2(s)] &= \sigma^2 |s|^{2H}, \\ E[Z^2(t)] &= \sigma^2 |t|^{2H}, \\ E[Z^2(s-t)] &= \sigma^2 |s-t|^{2H}. \end{aligned}$$

So we have:

$$\begin{aligned} E[Z(s)Z(t)] &= E[\frac{1}{2}\{Z^2(s) + Z^2(t) - Z^2(s-t)\}] \\ &= \frac{1}{2}\{E[Z^2(s)] + E[Z^2(t)] - E[Z^2(s-t)]\} \\ &= \frac{1}{2}(\sigma^2|s|^{2H} + \sigma^2|t|^{2H} - \sigma^2|s-t|^{2H}) \\ &= \frac{\sigma^2}{2}(|s|^{2H} + |t|^{2H} - |s-t|^{2H}). \end{aligned}$$

The proof above shows that part(5) in **Proposition 2** is well satisfied. Lastly, I will use the covariance function to confirm self-similarity. For any $a > 0$, we have:

$$\begin{aligned} E[Z(as)Z(at)] &= \frac{\sigma^2}{2}(|as|^{2H} + |at|^{2H} - |a(s-t)|^{2H}) \\ &= \frac{\sigma^2}{2}(|a|^{2H}|s|^{2H} + |a|^{2H}|t|^{2H} - |a|^{2H}|s-t|^{2H}) \\ &= |a|^{2H}\frac{\sigma^2}{2}(|s|^{2H} + |t|^{2H} - |s-t|^{2H}) \\ &= |a|^{2H}E[Z(s)Z(t)] \\ &= E[|a|^{2H}Z(s)Z(t)] \\ &= E[a^H Z(s) a^H Z(t)]. \end{aligned}$$

Since s and t are arbitray, we can conclude that $\{Z(at)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{a^H Z(t)\}_{t\in\mathbb{R}}$ for

$\forall a > 0, 0 < H < 1$.

Condition (3):

The condition directly gives the feature of stationary increments of the process $\{Z(t)\}_{t\in\mathbb{R}}$. To sum up, the three conditions given in **Proposition 2** reflect that the stochastic process $\{Z(t)\}_{t\in\mathbb{R}}$ is a Gaussian process with mean zero, self-similar and has stationary increments, respectively. In addition, the certain mean function, variance and covariance functions of $Z(t)$ match the counterparts in **Proposition 2** very well, which confirms that such process $\{Z(t)\}_{t\in\mathbb{R}}$ is self similar. Defer to **Definition 6**, the stochastic process $\{Z(t)\}_{t\in\mathbb{R}}$ is fractional Brownian motion. $\square$

To conclude, in Chapter 2, I give the definitions of self-similarity, stationary increments and $H-sssi$ process when combining the former two definitions together. The existence of $H-sssi$ process is proved and its properties are listed in **Proposition 2**. Finally, the famous $H-sssi$ process, fractional Brownian motion is introduced and the following **Proposition 4** in fact offers a way to check if a process is $H-sssi$. In the next chapter, I will introduce a very modern topic: Wavelet analysis.

Chapter 3: Wavelet analysis

3.1 Wavelet overview

Having studied the definition of self-similarity, we know that the constant a in Relation (2.1) is arbitrary, but it is still difficult to check if sample data has self-similar features for any the positive constant a, especially when the sample size is large. Even if we can determine the mean, variance and covariance functions of sample data, it is still hard to convince that Relation(2.3) holds for every two moments s and t. In modern mathematics, wavelet analysis, as one of the most efficient ways to detect self-similarity, is very popular on researching self-similar problems. The basic idea of wavelet analysis is to represent and analyse data through scaling. The brief history and development of wavelet has been presented in Chapter 1 and more details are introduced by Graps[7]. First, we consider the mathematical definition of a wavelet:

Definition 7. *If a function $\psi(t)$ satisfies the admissibility condition,*

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{3.1}$$

and integrability conditions, $\psi(t) \in L^1(\mathbb{R}) \bigcap L^2(\mathbb{R})$ (i.e. $\int_{-\infty}^{+\infty} |\psi(t)|dt < \infty$ and $\int_{-\infty}^{+\infty} \psi^2(t)dt < \infty$), then we call $\psi(t)$ a wavelet.

Remark 2. *If $\psi(t)$ is a wavelet, it also needs to meet the following requirements:*

(1) $\psi(t)$ *is bounded;*

(2) $\psi(t)$ *is centred around the origin;*

(3) $\psi(t)$ *is either finite or decreases to zero rapidly when time $t \to \infty$.*

Definition 8. *Given an positive integer N, if for all integers $0 \leqslant k < N$, a wavelet*

$\psi(t)$ *satisfies:*

$$\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0 \tag{3.2}$$

then we say the wavelet $\psi(t)$ *has N vanishing moments.*

Vanishing moment sometimes is thought as the "degree" of wavelet. We can see it is trivial when $N = 1$, since the only choice of k in Relation (3.2) is zero, which leads to the fundamental condition of a wavelet. That is, any wavelet has at least 1 vanishing moment.

Various wavelet functions have been developed and frequently used in different mathematical analysis. Some well-known wavelets include *Daubechies wavelet family*, *Spline wavelet family*, *coiflet wavelet family*, *Mexican hat wavelet* etc.

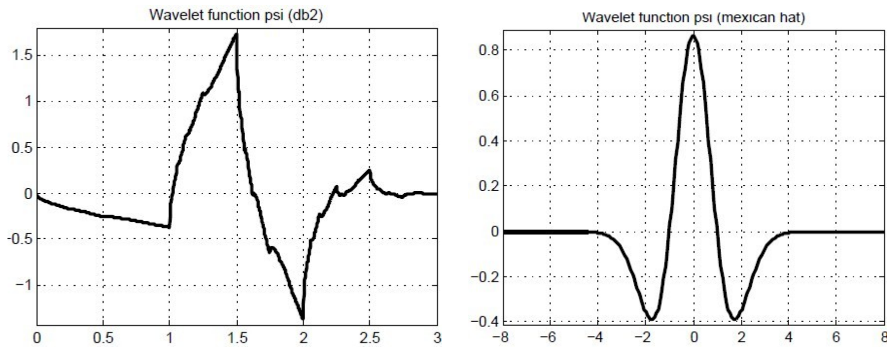


Figure 3.1: Wavelet function examples

Visually, the two typical wavelet function examples in Figure(3.1) roughly reflect basic properties of the wavelets, including the two fundamental conditions in **Definition 7** and three extended conditions in **Remark 2**. Here, I will check whether the *Haar wavelet*, also known as *Daubechies 1*, satisfies the above properties. Another reason to use *Haar wavelet* is that I will use it for further study on self-similarity.

Definition 9. *Haar wavelet function:*

$$\psi(t) = \begin{cases} 1, & 0 \leqslant t < \frac{1}{2} \\ -1, & \frac{1}{2} \leqslant t < 1 \\ 0, & otherwise \end{cases} \tag{3.3}$$

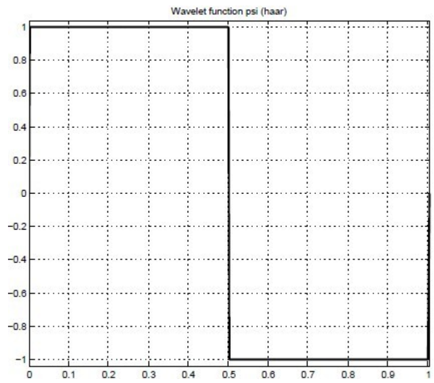


Figure 3.2: Haar wavelet function

Claim: Relation (3.3) defines a wavelet.

Proof. I will check the conditions in **Definition 7** and **Remark 2** one by one:

(1) Admissibility:

$$\begin{aligned} \int_{-\infty}^{\infty} \psi(t)dt &= \int_0^{\frac{1}{2}} 1dt + \int_{\frac{1}{2}}^{1} (-1)dt + 0 \\ &= t|_0^{\frac{1}{2}} + (-t)|_{\frac{1}{2}}^{1} + 0 \\ &= (\frac{1}{2} - 0) + (-1 + \frac{1}{2}) + 0 \\ &= 0 \end{aligned}$$

(2) Integrability:

It has been proved in part(1) that $\int_{-\infty}^{+\infty} \psi(t)dt = 0 < \infty$; Similarly, we can integrate $\psi^2(t)$ from $-\infty$ to ∞ and get $\int_{-\infty}^{+\infty} \psi^2(t)dt = 1 < \infty$

(3) Conditions in **Remark 2**:

Obviously, $\psi(t)$ is bounded within [-1,1]; from figure(3.2), the graph of $\psi(t)$ is symmetric to $x = \frac{1}{2}$, which is not far away from the origin; $\psi(t)$ is finite, and the values of $\psi(t)$ equal to zero other than [-1 1], which satisfies part(3) in **Remark 2**. □

As mentioned above, *vanishing moment* appears as a very important characteristic of wavelet, so it is meaningful to determine the *vanishing moment* of *Haar wavelet*.

Claim: *Haar wavelet* has one vanishing moment. Let $\psi(t)$ be *Haar* wavelet, ,

Proof. It is equivalent to prove that $\int_{-\infty}^{+\infty} t^k\psi(t)dt \neq 0$, where $\psi(t)$ is *Haar wavelet function* and $k = 0, 1, 2, ..., N-1$, if positive integer $N > 1$. In other words, $k = 0$ is the unique value of non-negative integer k to hold Relation (3.2).

Assume not, i.e. $\exists k > 0$ holds Relation (3.2) with *Haar wavelet* $\psi(t)$. Let $\psi(t)$ be *Haar wavelet function* as Relation (3.1), and integrate the left side of Relation (3.2) when plugging in $\psi(t)$:

$$
\begin{aligned}
\int_{-\infty}^{+\infty} t^k\psi(t)dt &= \int_0^{\frac{1}{2}} t^k dt - \int_{\frac{1}{2}}^{1} t^k dt + 0 \\
&= \frac{1}{k+1}t^{k+1}|_0^{\frac{1}{2}} - \frac{1}{k+1}t^{k+1}|_{\frac{1}{2}}^{1} \\
&= \frac{1}{k+1}(\frac{1}{2})^{k+1} - [\frac{1}{k+1} - frac1k+1(\frac{1}{2})^{k+1}] \\
&= \frac{1}{k+1}(\frac{1}{2})^{k} - \frac{1}{k+1} \\
&= \frac{1}{k+1}[(\frac{1}{2})^{k} - 1] \\
&= 0.
\end{aligned}
$$

Notice that $k > 0$, so the factor $\frac{1}{k+1} \neq 0$, which implies $(\frac{1}{2})^k - 1 = 0$. However, the unique real root of this equation is $k = 0$, and it is a contradiction to the assumption $k > 0$.

Hence, the *Haar wavelet* has one vanishing moment. □

3.2 Wavelet coefficients

3.2.1 Introduction of wavelet functions and coefficients

As mentioned in the last section, the main idea of wavelet analysis is to change scales. The scales are considered to change in two directions upon the original wavelet function, where the two directions are called *dilation* and *translation*. In order to distinguish the original wavelet function from the scale-changed wavelet function, we name the original wavelet function *mother wavelet*. For example, Relation (3.3) is called as the *mother wavelet* of *Haar wavelet*.

Definition 10. *Let $\psi(t)$ be a mother wavelet, for $j \in \mathbb{Z}$, $k \in \mathbb{Z}$, denote the function*

$$\psi_{j,k}(t) = \frac{1}{2^{\frac{j}{2}}}\psi(2^{-j}t - k) \tag{3.4}$$

as the dilation and translation of $\psi(t)$.

In Relation (3.4), it is obvious that the index j refers to *dilation*, and $\psi(2^{-j}t)$ corresponds to the expansion in scale of $\psi(t)$ by a factor 2^j; the index k refers to *translation*, and $\psi(t-k)$ corresponds to shifting $\psi(t)$ towards right by k units. Again, we take *Haar wavelet* as an example to get a clearer picture of *dilation* and *translation*.

Let $\psi(t)$ be the mother wavelet of *Haar*, and it is defined on the interval $[0, 1)$.

Let $j = 2$ and $k = 1$, then the new wavelet:

$$\begin{aligned}\psi_{2,1}(t) &= \frac{1}{2}\psi(2^{-2}t - 1) \\ &= \frac{1}{2}\psi(2^{-2}(t-4))\end{aligned}$$

The new wavelet function $\psi_{j,k}(t)$ should be defined on $[0, 1) \xrightarrow{\times 2^2} [0, 4) \xrightarrow{+4} [0, 8)$. So, we have:

$$\psi_{2,1}(t) = \begin{cases} \frac{1}{2}, & 0 \leqslant t < 6 \\ -\frac{1}{2}, & 6 \leqslant t < 8 \\ 0, & otherwise \end{cases}$$

From Relation (3.4) and the example above, we can roughly conclude the defined interval of wavelet function is dilated by positive j, but contracted by negative j; it is shifted rightwards by positive k, but leftwards by negative k. Though wavelet $\psi_{j,k}(t)$ is scaled from *mother wavelet* $\psi(t)$, it still preserves $\| \cdot \|_{L^2}$, and this proposition is shown as below:

Proposition 5. *Let $\psi_{j,k}(t)$ be scaled from the mother wavelet $\psi(t)$, then*

$$\textstyle\int_{-\infty}^{+\infty} \psi_{j,k}^2(t)dt = \int_{-\infty}^{+\infty} \psi^2(t)dt.$$

Proof. According to **Definition 10**, for $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$,

$$\begin{aligned}
\int_{-\infty}^{+\infty} \psi_{j,k}^2(t)dt &= \int_{-\infty}^{+\infty} [2^{-\frac{j}{2}}\psi(2^{-j}t - k]^2 dt \\
&= \int_{-\infty}^{+\infty} 2^{-j}\psi^2(2^{-j}t - k)dt \\
&= 2^{-j}\int_{-\infty}^{+\infty} \psi^2(2^{-j}t - k)d(2^{-j}t - k) \cdot 2^j \\
&= \int_{-\infty}^{+\infty} \psi^2(2^{-j}t - k)d(2^{-j}t - k) \\
&= \int_{-\infty}^{+\infty} \psi^2(t)dt.
\end{aligned}$$

□

If we use $\{\psi_{j,k}(t), j, k \in \mathbb{Z}\}$ as a *filter* set, we can define Discrete Wavelet Transform (DWT) of a stochastic sample path as follows:

Definition 11. *Let $\{Z(t)\}_{t\in\mathbb{R}}$ be a stochastic process, for $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$,*

$$d_{j,k} = \int_{-\infty}^{+\infty} Z(t)\psi_{j,k}(t)dt \tag{3.5}$$

are the wavelet coefficients of $\{Z(t)\}_{t\in\mathbb{R}}$ through DWT.

Here, wavelet coefficients $d_{j,k}$ are also called *details* in contrast to the scaling coefficients $a_{j,k}$, *approximations*, which will be introduced and constructed in the next subsection. In addition, I will focus on $d_{j,k}$ and use these *detail* coefficients to detect self-similarity and $H - sssi$ process in the next chapter. In fact, the adjective "Discrete "in DWT does not refer to the integration in Relation (3.5) with respect to time t, but the indices j and k taking integers.

3.2.2 Multiresolution Analysis(MRA)

In **Definition 11** in the last subsection, we learned that *detail* wavelet coefficients are constructed by a set of scaled wavelet functions - these are called *filters*. In wavelet analysis, those *filters* are also called multiresolution wavelets, and the study on such *filters* are called Multiresolution Analysis (MRA).

Definition 12. *Multiresolution analysis is a method relevant to the Discrete Wavelet Transform that consists of a sequence of closed subsets $\{V_n\}_{n\in\mathbb{Z}}$ in $L^2(\mathbb{R})$, which satisfies the following conditions:*
(1) $... \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} ... \subset L^2(\mathbb{R})$.
(2) $\cap_{-\infty}^{+\infty} V_j = \{0\}$ and $\cup_{-\infty}^{+\infty} V_j$ is dense in $L^2(R)$, for $\forall j \in \mathbb{Z}$.
(3) V spaces are self-similar, which means $f(2^j t) \in V_j \Leftrightarrow f(t) \in V_0$.
(4) There exists a scaling function $\phi \in V_0$ whose translations span the space: $V_0 = \{f(x) \in L^2(\mathbb{R}) | f(x) = \sum_k c_k \phi(x-k)\, k \in \mathbb{Z}\}$ for some constants c_k, and the set $\{\phi(\cdot - k)\}$ is an orthonormal basis.

In part(4), ϕ is called *mother scaling function*. In fact, wavelet function ψ also can also be defined through ϕ. We have the so-called two-scaled equations as follows: For $n \in \mathbb{Z}$,

$$\phi(t/2) = \sqrt{2} \sum_n u_n \phi(t-n) \tag{3.6}$$

$$\psi(t/2) = \sqrt{2} \sum_n v_n \phi(t-n). \tag{3.7}$$

In the two-scaled function, u_n and v_n are some coefficients to form the basis of the scaling and wavelet coefficients, and they vary in different types of wavelets.
MRA was first introduced by Mallat in 1989, and Mallat also proposed the fast

recursive pyramidal decomposition algorithm based on MRA. And the fast recursive pyramidal decomposition, which is known as wavelet decomposition now, illustrates the way of calculating scaling coefficients and wavelet coefficients in different scales. Here, a new name, scaling coefficients, are the coefficients from the last subsection. They have the very similar constructions as wavelet coefficients, the main difference is that they are basically constructed on scaling functions ϕ instead of wavelet functions ψ.

Definition 13. *Let $Z(t)$ be a stochastic process and $\phi(t)$ be mother scaling function, the scaling coefficient $a_{j,k}$ is defined as:*

$$a_{j,k} = \int_{-\infty}^{+\infty} Z(t)\phi_{j,k}(t)dt \tag{3.8}$$

where $\phi_{j,k}(t) = \frac{1}{2^{\frac{j}{2}}}\phi(2^{-j}t - k)$.

We can see that scaling function $\phi_{j,k}$ is similarly defined as scaled wavelet function $\psi_{j,k}$ in Relation (3.4) and scaling coefficients $a_{j,k}$ are also called as *approximation* coefficients mentioned in the last subsection. In Relation (3.6) from two-scaled equations, we have the connection between $\phi(t)$ and $\phi(t/2)$. If combining it with Relation

(3.8), we have:

$$
\begin{aligned}
a_{j,k} &= \int_{-\infty}^{+\infty} Z(t)2^{-\frac{j}{2}}\phi(2^{-j}t-k)dt \\
&= \int_{-\infty}^{+\infty} Z(t)2^{-\frac{j}{2}}\sqrt{2}\sum_n u_n\phi(2^{-j+1}t-2k-n)dt \\
&= \int_{-\infty}^{+\infty} Z(t)2^{\frac{-j+1}{2}}\sum_n u_n\phi(2^{-j+1}t-2k-n)dt \\
&= \sum_n u_n\int_{-\infty}^{+\infty} Z(t)2^{\frac{-j+1}{2}}\phi(2^{-j+1}t-(2k+n))dt \\
&= \sum_n u_n a_{j-1,2k+n} \\
&= \sum_n u_{-n} a_{j-1,2k-n}
\end{aligned}
$$

Similarly, if plugging Relation (3.7) into Relation (3.8) instead of Relation (3.6), we can get the result of *detail* coefficients in the same way:

$$d_{j,k} = \sum_n v_{-n} a_{j-1,2k-n}.$$

With the *approximation* coefficient $a_{j,k}$, *detail* coefficient $d_{j,k}$ and their relations with one lower level *approximation* coefficient $a_{j-1,k}$ indicated above, we can derive the following conclusion:

$$a_{j-1,k} \stackrel{\downarrow 2, u_{-n}}{\longrightarrow} a_{j,k} \tag{3.9}$$

$$a_{j-1,k} \stackrel{\downarrow 2, v_{-n}}{\longrightarrow} d_{j,k}. \tag{3.10}$$

There is a lot of information of *approximation* and *detail* coefficients in Relation (3.9) and (3.10). First of all, we can see both *approximation* and *detail* coefficients at j-th level are yielded by *approximation* coefficient at $(j-1)$-th level for fixed $j \in \mathbb{Z}$,

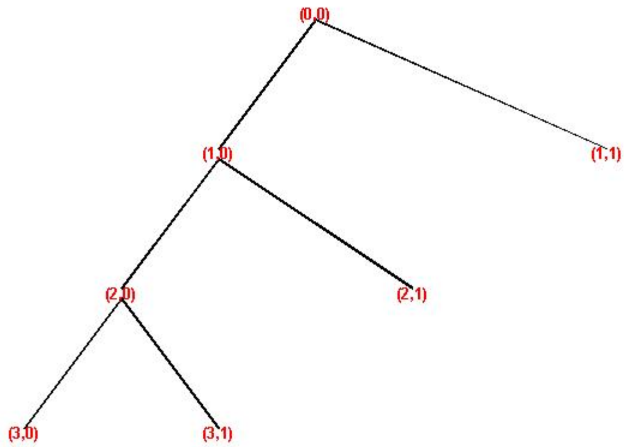


Figure 3.3: Tree decomposition

which is also shown in the tree decomposition in Figure (3.3); In addition, the length, sometimes also called frequency, of $a_{j,k}$ and $d_{j,k}$ are halved by the length of $a_{j-1,k}$, for $j \in \mathbb{Z}$. If the length of $a_{j-1,k}$ is odd, the length of $a_{j,k}$ and $d_{j,k}$ are round up to the closest integer. A more detailed explanation and algorithm are given by Mallat[25] and Vishwanath[9].

Now we have determined the scales of *approximation* and *detail* coefficients. The next question coming up is how to calculate these coefficients. From Relation (3.8), we can see that those coefficients basically depend on original signal $Z(t)$ and scaling and wavelet functions, $\phi_{j,k}$ and $\psi_{j,k}$. Let $j = 0$ in Relation (3.8), we have zero-level *approximation* coefficient $a_{0,k} = \int_{-\infty}^{+\infty} Z(t)\phi(t-k)dt$. The later terms $a_{j,k}$ can be computed from $a_{0,k}$ and constants u_{-n}. Similarly, *detail* coefficients $d_{j,k}$ also can be iterated by $a_{0,k}$ and constants v_{-n}. Since both *mother wavelet* and *mother scaling* functions can be obtained once the certain wavelet is determined, then the constants u_n and v_n become the keys to calculate *approximation* and *detail* coefficients.

In fact, u_n is called low-pass filter, denoted by $L(z)$, and it generates *approximation* coefficients, which are high-scale and low-frequency components of original signal; v_n is called high-pass filter, denoted by $H(z)$, and it generates *detail* coefficients, which are low-scale and high-frequency components of original signal; u_n and v_n consist a filter bank. There is one very important relation between u_n and v_n:

$$v_n = (-1)^n u_{1-n}. \tag{3.11}$$

The relation above can be got when taking the Fourier transform of scaling function and comparing the structure of two-scaled functions. The comparison and proof are given by Vidakovic[15]. And more detailed exploration of how filter bank working in DWT and what *approximation* and *detail* coefficients look like is given by Misiti[13].

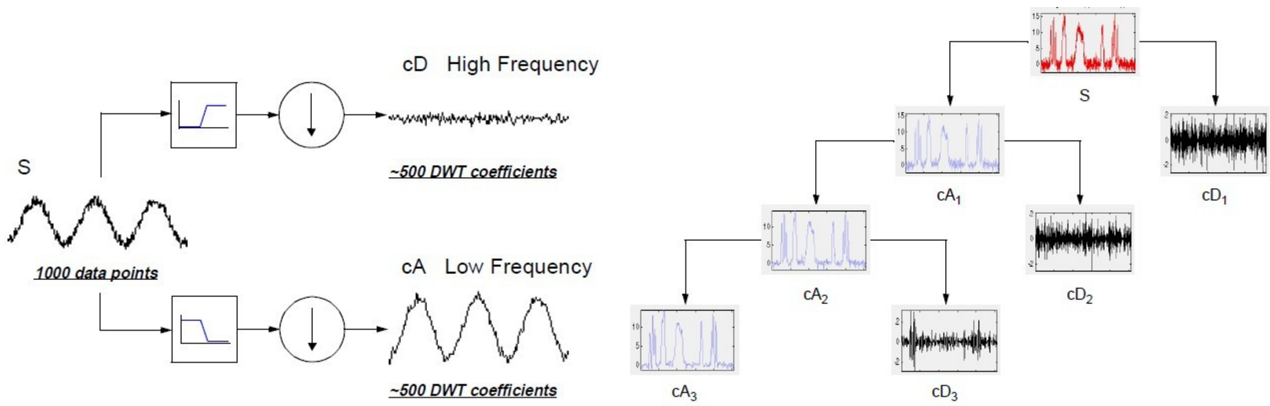


Figure 3.4: Filter bank and Wavelet decomposition

Figure (3.4) gives a very straightforward version of wavelet decomposition. The left one illustrates what filters look like and the scale information of each iteration of decomposition; The right one is a better illustration based on Figure (3.3). We see that *approximation* coefficients are exact the approximation of the original signal, while *detail* coefficients more reflect the noises of the original signal, and it is also true that the more level of decomposition taken, the *approximation* coefficients of the signal is coarser, however, *detail* coefficients show the natural features of the signal. Besides, *approximation* and *detail* coefficients also can be used to reconstruct the signal, where the reconstruction process is called *synthesis*. This process is to upsample the size of *approximation* and *wavelet* coefficients, so it is thought as an inverse process of DWT. Wavelet synthesis are frequently applied in inverse problems, especially inverse image problems. There are a lot of introductions of wavelet synthesis offered by Misiti[13].

3.2.3 Daubechies Wavelet

Daubechies wavelet, named after Ingrid Daubechies, is a family of orthogonal wavelets defining on DWT. Basically, the construction of Daubechies wavelet is formed by the first N terms of binomial expansion:

$$P_N(y) = (1-y)^{-N} = \sum_{k=0}^{N-1} \binom{k+N-1}{k} y^k. \tag{3.12}$$

It is proved by Temme[16]. There can be different orders of Daubechies wavelets if taking different N values, so Daubechies wavelets are always denoted by "dbN". Haar wavelet is "db1", which has been shown in the previous work, however, higher order of Daubechies wavelets are not easy to describe, such as "db2"in Figure 3.1. Relation (3.12) is used to determine the values of high-pass and low-pass filters of Daubechies Wavelets when taking Fourier transform to the scaling function and using some nice properties of Daubechies wavelets. The exploration and proof are given by Narcowich[12] and Teeme[16], respectively. Here, I will only list some important properties and their corollaries of Daubechies wavelet family:

Proposition 6. *Let ϕ be mother scaling function and $\psi(t)$ be mother wavelet function of Daubechies wavelets, for fixed integer N, we have:*
(1)$\int_{-\infty}^{+\infty} \phi(t) = 1$
(2)$\int_{-\infty}^{+\infty} \phi(t-j)\phi(t-k)dt = \delta_{j,k}$*, for* $\forall j, m < N, j, m \in \mathbb{Z}$.

Let us take a look at these properties. If we integrate Relation (3.6) both sides with respect to t and use the result of part(1), we will have a very important result of Daubichies filter coefficients:

$$\begin{aligned}
\int_{-\infty}^{+\infty} \phi(t/2)dt &= \sqrt{2}\sum_n u_n \int_{-\infty}^{+\infty} \phi(t-n)dt \\
2\cdot\int_{-\infty}^{+\infty} \phi(t/2)d(t/2) &= \sqrt{2}\sum_n u_n \int_{-\infty}^{+\infty} \phi(t-n)d(t-n) \\
2 &= \sqrt{2}\sum_n u_n \\
\sqrt{2} &= \sum_n u_n.
\end{aligned}$$

Part(2) mainly comes from the orthogonality of the filters in Daubechies wavelets and it results another important corollary of filter coefficients of Daubechies wavelet family. Let $\phi(t)$ and $\phi(t-l)$ are two scaling functions, where $l \in \mathbb{Z}$:

$$\begin{aligned}
\phi(t)\phi(t-l) &= \sqrt{2}\sum_n u_n\phi(2t-n)\phi(t-l) \\
&= \sqrt{2}\sum_n u_n\phi(2t-n)\sqrt{2}\sum_m u_m\phi(2t-2l-m).
\end{aligned}$$

Integrate the equation above both sides:

$$\begin{aligned}
\delta_{0,l} &= 2\sum_n u_n \sum_m u_m \int_{-\infty}^{+\infty} \phi(2t-n)\phi(2t-2l-m)dt \\
&= 2\sum_n u_n \sum_m u_m \frac{1}{2}\int_{-\infty}^{+\infty} \phi(2t-n)\phi(2t-2l-m)d(2t) \\
&= \sum_n\sum_m u_n u_m \delta_{n,2l+m} \\
&= \sum_n u_n u_{n-2l}.
\end{aligned}$$

If setting $l = 0$, we can attain another important result of filter coefficients:

$$\sum u_n^2 = 1.$$

Now, let us focus on *Haar wavelet(db1)*, which we will use to detect self-similarity in the next section. Its mother wavelet function has been given in Relation (3.3), and here, let us take a look at its mother scaling function:

Definition 14. *The mother scaling funtion of Haar wavelet is the indicator function on [0,1]:*

$$\phi(t) = \begin{cases} 1, & 0 \leqslant t \leqslant 1 \\ 0, & otherwise \end{cases}. \tag{3.13}$$

And the figure of the scaling function is shown in Figure 3.5:

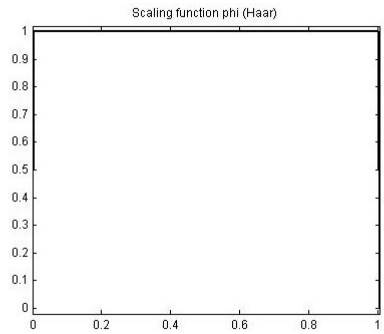


Figure 3.5: Haar scaling function

As we have obtained both mother scaling and wavelet function, we can try to calculate the filter coefficients. By Figure 3.5, it is obvious the mother scaling function of *Haar* has the following simple properties:

$$\begin{aligned} \phi(t) &= \phi(2t) + \phi(2t-1) \\ &= \sqrt{2}(\frac{1}{\sqrt{2}}\phi(2t) + \frac{1}{\sqrt{2}}\phi(2t-1)). \end{aligned}$$

If compared with Relation (3.6) in two-scaled function, it automatically yields the low-pass filter coefficients $u_0 = u_1 = \frac{1}{\sqrt{2}}$. By relation(3.11), the high-pass filter coefficients $v_0 = \frac{1}{\sqrt{2}}$ and $v_1 = -\frac{1}{\sqrt{2}}$, all of them shown in Figure(3.6):

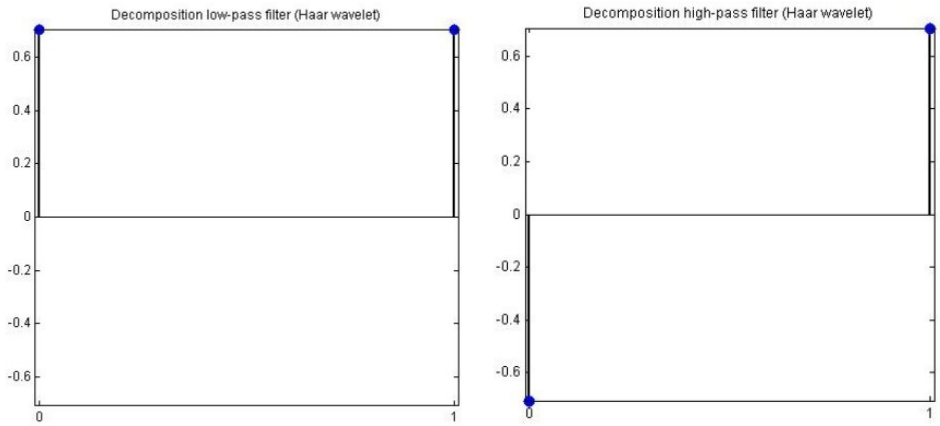


Figure 3.6: Haar low-pass and high-pass filters

If we take a quick check of the filter coefficients of *Haar* wavelet, we can see $u_0 + u_1 = \sqrt{2}$ and $u_0^2 + u_1^2 = 1$, which satisfy the requirements of filter coefficients from Daubichies wavelet family. Though we can determine the filters for *Haar wavelet* by sticking two other scaling functions, it is very hard to imitate the calculation of *Haar* to obtain obtain filter coefficients of higher-order wavelets in *Daubechies* family. However, Narcowich[12] and Temme[16] offer two nice ways to calculate general *Daubichies* wavelet filter coefficients by taking Fourier transform of the scaling function and two-scaled functions (sometimes they are also called Transform functions sometimes). Here, I will omit the detail proofs. Now, we can conclude that *Haar wavelet* is definitely a member of *Daubechies wavelet family.* If we try to combine the filter coefficients and mother wavelet function together, we can obtain the j-th iteration of filters of *Haar*:

$$\psi_{j,k}(t) = \begin{cases} 2^{-j/2}, & 0 \leqslant t < 2^{j-1} \\ -2^{-j/2}, & 2^{j-1} \leqslant t < 2^{j} \\ 0, & otherwise \end{cases} .$$

It is obvious that the period of the wavelet function is 2^j, and each positive value of k will shift the interval rightwards by $k \cdot 2^j$ units. In fact, it is clearer if we can write the filter coefficients of *Haar* in matrix representation. Let $u_0 = u_1 = v_0 = \frac{1}{\sqrt{2}}$ and $v_1 = -\frac{1}{\sqrt{2}}$, we have filter matrix as follows:

$$C = \begin{pmatrix} u_0 & u_1 & 0 & 0 & \cdots & 0 & 0 \\ v_0 & v_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & u_0 & u_1 & \cdots & 0 & 0 \\ 0 & 0 & v_0 & v_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & u_0 & u_1 \\ 0 & 0 & 0 & 0 & \cdots & v_0 & v_1 \end{pmatrix} .$$

If we have the signal vector:

$$\overrightarrow{S} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ \vdots \\ \vdots \end{pmatrix} .$$

Then, the output vector should be:

$$\overrightarrow{O} = \begin{pmatrix} low_1 \\ high_1 \\ low_2 \\ high_2 \\ low_3 \\ high_3 \\ \vdots \\ \vdots \end{pmatrix}.$$

In each iteration, $\overrightarrow{O} = C \cdot \overrightarrow{S}$. In the follwoing iteration, the previous output $\overrightarrow{O}$ becomes the new signal and the size of filter matrix C should be changed to match the length of $\overrightarrow{O}$ to attain a new output. We also can find that the matrix representation of filter coefficients persuasively reconfirms that the size of output is halved in each decomposition process of DWT.

3.3 Wavelet coefficients of self-similar process

3.3.1 Properties of wavelet coefficients

In this section, we try to calculate wavelet coefficients of self-similar processes and detect their statistical properties. First of all, regardless of self-similarity, let us take a look at some natural properties of *wavelet* coefficients $d_{j,k}$ themselves.

Proposition 7. *Let $\{Z(t)\}_t \in \mathbb{R}$ be a stochastic process, if the wavelet ψ has N vanishing moments, then the wavelet coefficients $d_{j,k}$ are the same for $Z(t)$ and $Z(t)+P(t)$, where $P(t)$ is a polynomial of $N-1$ degree.*

Proof. Recall in Relation (3.5), we know $d_{j,k} = \int_{-\infty}^{+\infty} Z(t)\psi_{j,k}(t)dt$. Here, we want to show:

$$
\begin{aligned}
d_{j,k} &= \int_{-\infty}^{+\infty} (X(t) + P(t))\psi_{j,k}(t)dt \\
&= \int_{-\infty}^{+\infty} X(t)\psi_{j,k}(t)dt + \int_{-\infty}^{+\infty} P(t)\psi_{j,k}(t)dt \\
&= d_{j,k} + \int_{-\infty}^{+\infty} P(t)\psi_{j,k}(t)dt.
\end{aligned}
$$

So it is equivalent to proving $\int_{-\infty}^{+\infty} P(t)\psi_{j,k}(t) = 0$. According to Relation (3.4): $\psi_{j,k} = 2^{-j/2}\psi(2^{-j}t - k)$, let $s = 2^{-j}t - k$, and we have:

$$
\begin{aligned}
\int_{-\infty}^{+\infty} P(t)\psi_{j,k}(t) &= \int_{-\infty}^{+\infty} P(2^j(s+k))2^{-j/2}\psi(s)2^j ds \\
&= 2^{j/2}\int_{-\infty}^{+\infty} P(2^j(s+k))\psi(s)ds.
\end{aligned}
$$

Since $P(2^j(s+k))$ is a polynomial of $N-1$ degree but the wavelet ψ has N vanishing moments, according to **Definition 8**, we can obtain the conclusion that $\int_{-\infty}^{+\infty} P(t)\psi_{j,k}(t) = 0$. □

Proposition 8. *Let $\{Z(t)\}_{t\in\mathbb{R}}$ be a stochastic process, if $\{Z(t)\}_{t\in\mathbb{R}}$ has stationary increments, then for fixed $j \in \mathbb{Z}$, the sequence $\{d_{j,k}, k \in \mathbb{Z}\}$ is stationary.*

Proof. We want to show the sequence $\{d_{j,k}\}_{k\in\mathbb{Z}}$ is stationary, that is, to prove $d_{j,k+k_0} \overset{d}{=} d_{j,k}$. In other words, the sequence $\{d_{j,k}, k \in \mathbb{Z}\}$ does not depend on $k_0 \in \mathbb{Z}$.

$$
\begin{aligned}
d_{j,k+k_0} &= \int_{-\infty}^{+\infty} Z(t)\psi_{j,k+k_0}(t)dt \\
&= \int_{-\infty}^{+\infty} Z(t)2^{-j/2}\psi(2^{-j}t - k - k_0)dt.
\end{aligned}
$$

Let $s = t - 2^j k_0$, and plug into the formula above, combining with Relation (3.1) and the definition of stationary increments, we can obtain:

$$\begin{aligned}
d_{j,k+k_0} &= \int_{-\infty}^{+\infty} Z(s + 2^j k_0) 2^{-j/2} \psi(2^{-j}s + k_0 - k - k_0) ds \\
&= \int_{-\infty}^{+\infty} [Z(s + 2^j k_0) - Z(2^j k_0)] 2^{-j/2} \psi(2^{-j}s - k) ds \\
&\stackrel{d}{=} \int_{-\infty}^{+\infty} [Z(s) - Z(0)] 2^{-j/2} \psi(2^{-j}s - k) ds \\
&= \int_{-\infty}^{+\infty} Z(s) \psi_{j,k}(s) ds \\
&= d_{j,k}.
\end{aligned}$$

$\square$

Proposition 8 shows that *wavelet* coefficients at the same level of decomposition have the same statistical distribution. That is, if two *wavelet* coefficients have the same index j, coefficients themselves $d_{j,k}$ do not depend on translation k.

Having known the definition, way of calculation, and basic properties of *wavelet* coefficients, we can try to detect the properties of *wavelet* coefficients of self-similar process.

Proposition 9. *Let $\{Z(t)\}_{t \in \mathbb{R}}$ be an H-ss stochastic process, then for fixed $j \in \mathbb{Z}$, we have*

$$d_{j,k} \stackrel{d}{=} 2^{j(H+\frac{1}{2})} d_{0,k}. \tag{3.14}$$

Proof. Since $Z(t)$ is self-similar and it is obvious that $2^j > 0$ for $j \in \mathbb{Z}$, we have:

$$Z(2^j t) \stackrel{d}{=} 2^{jH} Z(t).$$

Let $u = 2^{-j}t - k$ and $s = u + k$,

$$\begin{aligned}
d_{j,k} &= \int_{-\infty}^{+\infty} Z(t)\psi_{j,k}(t)dt \\
&= \int_{-\infty}^{+\infty} Z(t)2^{-j/2}\psi(2^{-j}t - k)dt \\
&= 2^{-j/2}\int_{-\infty}^{+\infty} Z(2^j(u+k))\psi(u)\frac{du}{2^{-j}} \\
&= 2^{j/2}\int_{-\infty}^{+\infty} Z(2^j s)\psi(s-k)ds \\
&\overset{d}{=} 2^{j/2}\int_{-\infty}^{+\infty} 2^{jH}Z(s)\psi(s-k)ds \\
&= 2^{j(H+\frac{1}{2})}\int_{-\infty}^{+\infty} Z(s)\psi(s-k)ds \\
&= 2^{j(H+\frac{1}{2})}d_{0,k}.
\end{aligned}$$

□

Once again, here $\overset{d}{=}$ means having the same finite dimensional distributions. **Proposition 9** states that *wavelet* coefficients $d_{j,k}$ and $2^{j(H+\frac{1}{2})}d_{0,k}$ from self-similar process have the same distribution, though they have different scales. However, it is a question how to determine if two sets of data have the same probability distribution. To answer this question, we can ues Two-sample *Kolmogrov-Smirnov* Test to compare data samples and detect their probability distributions.

Definition 15. *Suppose $X_1, X_2, ..., X_n$ are i.i.d. random variables, the empirical distribution function $F_n(x)$ is defined as the proportion of the observed values of the sample that are less than or equal to x:*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \leqslant x},$$

where $I_{X_i \leqslant x}$ is the indicator function.

Definition 16. *The Kolmogorov-Smirnov Test for two samples: Let a random sample of m observations $X_1, X_2, ..., X_n$ with distribution function $F(x)$; and another random sample of n observations $Y_1, Y_2, ..., Y_n$ with distribution function $G(x)$,*

$$H_0 \quad : F(x) \stackrel{d}{=} G(y)$$

$$H_1 \quad : F(x) \stackrel{d}{\neq} G(y)$$

Use test statistic D_{mn}:

$$D_{mn} = \sup_{-\infty < x < \infty} |F_m(x) - G_n(x)|.$$

The test procedure is rejecting H_0 when

$$(\frac{mn}{m+n})^{\frac{1}{2}} D_{mn} \geqslant K_\alpha,$$

where K_α is Kolmogorov statistic, whose cumulative distribution function is $Pr(K \leqslant x) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2x^2}$, and α is significance level.

The more detailed explanation and testing procedures are given by DeGroot[26]. In MATLAB, there exists a built in function **kstest2** for Two-sample Kolmogorov-Smirnov Test, for example:

```
>> h = kstest2(rand(100,1),rand(1000,1))
h =
     0
```

When two samples of data come from the same distribution, **kstest2** will return 0, otherwise it will return 1. The default significance level α in function **kstest2** is 5%. However, using *Kolmogorov-Smirnov* test in MATLAB will fail sometimes, since the test itself is taking supreme norm and sample data are generated randomly by computer. Especially when the sample size is very large, there are more chances to get big errors in *Kolmogorov-Smirnov* statistic with respect to supreme norm. So based on the idea of *Kolmogorov-Smironov* test, there is another hypothesis test - *Cramer Von Mises* test, can be used to check if two sample data come from the same distribution. Instead of supreme norm, *Cramer Von Mises* test is using L^2 norm statistic $\omega^2 = \frac{mn}{m+n}\int_{-\infty}^{+\infty}[F_m(x) - G_n(x)]^2 d(F_m(x) + G_n(x))$ to check. Though the fact that data are randomly generated also influences *Cramer Von Mises* test to some extent, the accuracy and approximation of *Cramer Von Mises* test is better than *Kolmogorov-Smironov* test, which is obvious and also explained explicitly by Hodges[28]. More detail introduction and exploration of *Cramer Von Mises* test is given by Anderson[27]. Unfortunately, *Cramer Von Mises* test has no built-in function in MATLAB, but the M-file is offered in **Appendix A** by Cardelino (http://www.mathworks.com/matlabcentral/fileexchange/13407-two-sample-cramer-von-mises-hypothesis-test/content/cmtest2.m). In the future work of this paper, I will find another more convincing way to check if such wavelet coefficients from given sample data come from the same distribution. But I will also use Two-sample *Kolmogorov-Smironov* test and *Cramer Von Mises* test to convince my conclusion when the new method is ambiguous.

Proposition 10. *Let $Z(t)$ be an H-sssi Gaussian process with mean zero and finite variance, then the wavelet coefficients should have the following properties:*
(1)$E[d_{j,k}] = 0$;
(2)$E[d_{j,k}^2] = 2^{j(2H+1)}E[d_{0,0}^2]$;
(3)$E[d_{j,k_1}, d_{j,k_2}] \leqslant C|k_1 - k_2|^{2(H-N)}$ *for fixed* $j \in \mathbb{Z}$, *when* N *is sufficiently large, where* C *is some constant.*

In **Proposition 10**, part(1) is obviously true since $E[Z(t)] = 0$.

From **Proposition 8** and Relation(3.13), we can easily obtain part(2):

$$\begin{aligned} E[d_{j,k}^2] &= E[d_{j,k}d_{j,k}] \\ &\stackrel{d}{=} E[2^{j(H+\frac{1}{2})}d_{0,k}2^{j(H+\frac{1}{2})}d_{0,k}] \\ &= 2^{j(2H+1)}E[d_{0,k}^2] \\ &= 2^{j(2H+1)}E[d_{0,0}^2]. \end{aligned}$$

Additionally in part(2), if we take the logarithm of both sideswith respect 2, it yields a line with slope $(2H+1)$ and interception $\log_2 E[d_{0,0}^2]$. This observation is the basis to estimate the Hurst parameter H through wavelet analysis.

Part(3) refers to the covariance structure of the *wavelet coefficients* in the fixed level j. Tewfik and Kim[29] proposed and proved that d_{j,k_1} and d_{j,k_2} are almost uncorrelated when $N > H + \frac{1}{2}$ and this correlation decays hyperbolically fast when N is sufficiently large. In fact, even in different level j, the correlation of *wavelet coefficients* are extremely small if the condition $N > H + \frac{1}{2}$ holds. It is true that $E[d_{j_1,k_1}d_{j_2,k_2}] \leqslant C|2^{-j_1}k_1 - 2^{-j_2}k_2|^{2(H-N)}$ for some constant C, which was proved by Flandrin[17] and introduced to public by Arby[18].

So far, we have studied that the properties of *wavelet* coefficients of $H - sssi$ process, and these properties will help us to check statistical self-similarity for time series from financial data and chaotic dynamical systems. In other words, **Proposition 9** and **10** are the keys for us to check and study self-similarity for time series from various aspects. However, when we apply the method to real data, all calculation will be no longer theoretically, but numerically. It automatically lead to a question: How much reliability of this method to detect self-similarity in numerical calculation? To answer this question, we can take a look at the following example.

3.3.2 Wavelet analysis on fractional Brownian motion

As already been proved, fractional Brownian motion(fBm) is an $H - sssi$ process. We can take fBm as an example to test the method of detecting self-similarity given in the last subsection.

First of all, We need to generate a sample path of fBm. There are a couple of ways to simulate fBm: Hosking Method, Cholesky Method, Fourier Transform Method, Wavelet-based Simulation etc., and the algorithms are introduced by Ton Dieker[19]. In addition, fBm is also available to be generated in various programs, such as Maple and MATLAB. The program we will use in this paper is MATLAB. I will give the codes of generating fBm with Cholesky method in MATLAB here:

```
function [ S ] = fBm( H,n )
if nargin < 2
   n = 256;
end
if H == 0
   S = randn(n,1);
   return
end
t = (1:n)';
a = 2*H;
r = zeros(n,n);
for i = 1:n;
    for j = 1:n;
    r(i,j) = (t(i).^a + t(j).^a-abs(t(i)-t(j)).^a)/2;
    end
end
```

```
l = chol(r);
l = l';
x = randn(n,1);
S = l*x
end
```

However, there is a built-in function **wfbm** in MATLAB to generate fBm based on Wavelet simulation and ARIMA progress, and the algorithm was firstly introduced by Abry and Sellan[20] in 1996. If we want to generate fBm with $H = 0.5$ and length(size) $L = 2^{10}$, we can easily run the following command to get the fBm data and its plot:

```
>> FBM05 = wfbm(0.5, 2^10, 'plot');
```

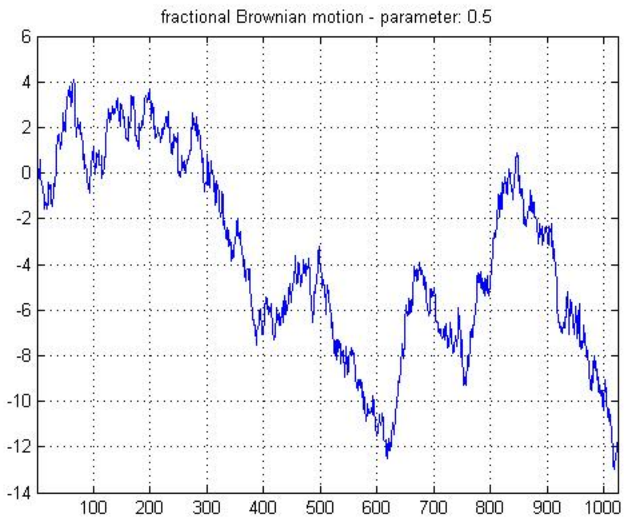


Figure 3.7: fractional Brownian motion with $H = 0.5$

Worthy of mentioning here, we are choosing the size of generated fBm equal to 2^{10}, because we have discussed previously that the wavelet decomposition reduces the size of signal by a factor of $\frac{1}{2}$ in each iteration, and the size which equals to a power of 2 will save time running the program in MATLAB.

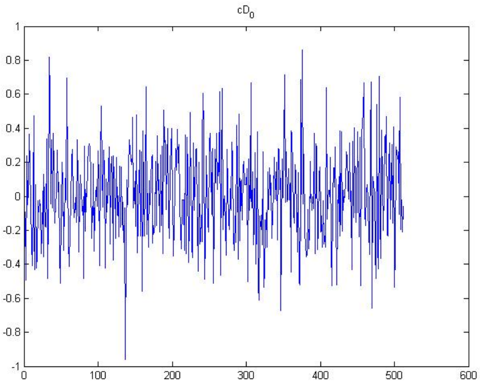
Next step, we need to implement wavelet decomposition to the generated signal. Fortunately, we can use command **wavedec** in MATLAB to make it. The wavelet we are going to use is *Haar* wavelet, which is denoted as "*db1*"in MATLAB. In fact, compared with the wavelets of higher order from *Daubechies* family, *Haar* wavelet, as the simplest one, has the easiest but nicest local properties, and it gives optimum estimation results of self-similar process. The comparison of wavelets from *Daubechies* wavelet family is given by Ciftlikli and Gezer[21]. In addition, we are very familiar with how *Haar* wavelet works now. So it is good to choose *Haar* wavelet as our sample wavelet in wavelet decompositon.

```
>> [C L]=wavedec(FBM05,10,'db1');
```

In the code above, there are two outputs for **wavedec** C and L, where C records the information of *approximation* and *detail* coefficients and L returns the size of wavelet coefficients in each step. Then we can use command **detcoef** to array the wavelet coefficients level by level.

```
>>[cD0,cD1,cD2,cD3,cD4,cD5,cD6,cD7,cD8,cD9]= detcoef(C,L,[1:1:10]);
```

Figure 3.8 below shows some paths of *detail* coefficients. They look like holding the first property in **Proposition 10** in Section 3.2.2, where both their expected values are roughly zero. In order to determine if those *detail* coefficients follow **Proposition 10** and **11**, we calculate their expected values and variance in MATLAB. In MATLAB statistics package, command **mean** is used to evaluate expected value of a string or vector, and **var** for the calculation of variance. Besides, in our case, once part(1) in **Proposition 10** holds, it would be true that $Var[d_{j,k}] = E[d_{j,k}^2]$, since $Var[d_{j,k}] = E[d_{j,k}^2] - (E[d_{j,k}])^2$, where $E[d_{j,k}] = 0$. If checking all the expected value and variance of *wavelet* coefficients and we can obtain the table in Figure 3.9.



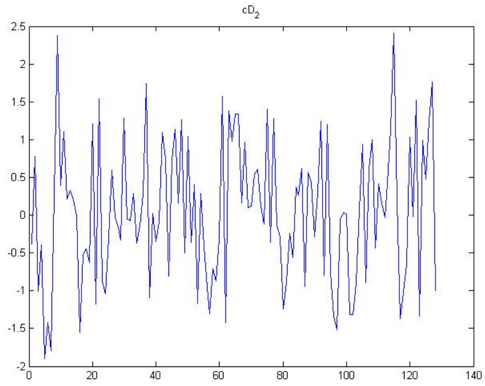


Figure 3.8: Figures of cD_0 and cD_2

	Expected Value	Variance	Var(dj+1,k)/Var(dj,k)	Relative Error	Log2(Variance)
fBm data	-3.697609758	18.14457078			
$d_{0,k}$	0.011562064	0.07543571	2.669945807	0.498157749	-3.728608569
$d_{1,k}$	0.061650239	0.20140926	4.219532359	0.052027652	-2.311798109
$d_{2,k}$	0.047231365	0.84985287	3.619922226	0.104996116	-0.234714992
$d_{3,k}$	0.207529823	3.07640130	4.106707728	0.025983765	1.621243709
$d_{4,k}$	1.081801271	12.63388101	3.481137483	0.149049706	3.659225985
$d_{5,k}$	1.489261812	43.98027673	4.27201055	0.063672724	5.458784777
$d_{6,k}$	6.688686398	187.88420622	5.273093851	0.241432049	7.553699987
$d_{7,k}$	10.06701352	990.73105243	4.08314971	0.02036411	9.952349661
$d_{8,k}$	10.91383461	4045.30320988			11.98203213
$d_{9,k}$	87.79335977				

Figure 3.9: Wavelet coefficients of generated fBm

From the table in Figure 3.9, first of all, we can see the expected value of original fBm data is around -3.7 but not exact 0. This is because the sample we generated is random; $E[X(t)]$ is continuously integrated on $(-\infty, +\infty)$, but our sample is discrete and its size is limited due to the memory of program. Secondly, the expected value of *wavelet* coefficients $d_{j,k}$ are close to zero generally, but they are apparently far away from zero in higher-level. That is because if we take the expected value on both sides in Relation 3.14, we have $E[d_{j,k}] = 2^{j(H+\frac{1}{2})}E[d_{0,k}]$. As the same reason indicated above, numerical computation can not guarantee that $E[d_{0,k}] = 0$, so the $E[d_{j,k}]$ is even farther away from zero when j turns to be very large. This is also why in wavelet toolbox or wavelet packages from different programs, the level of wavelet analysis usually does not exceed 6.

Having analysed the reasons of existing errors, we can start to check self-similarity, even H-sssi features, of fBm. The first part to check is H-ss feature, self similarity with Hurst parameter H, of generated fBm data. The corresponding part is **Proposition 9**. In our case, $H = \frac{1}{2}$, so we need to check if $d_{j,k} \overset{d}{=} 2^j d_{0,k}$. As we have obtained all the *wavelet* coefficients $d_{j,k}$, we can try to compare the histograms between $2^j d_{0,k}$ and $d_{j,k}$ to get the initial conclusion. The histogram of given data can be obtained by command **hist** in MATLAB. Figure 3.10 shows an example of two *wavelet* coefficients of fBm.

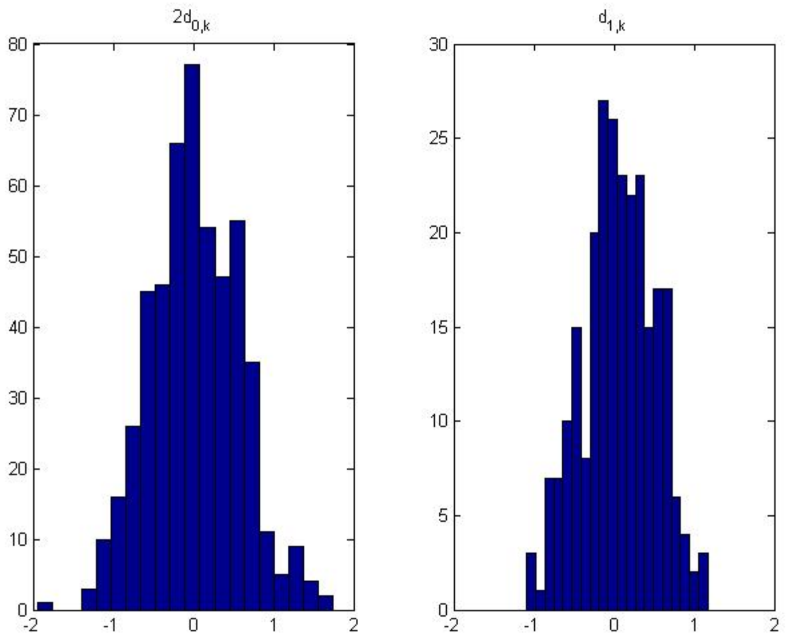


Figure 3.10: Histograms of wavelet coefficients of fBm

Though the trends of two histograms look roughly alike in general, some detail parts are still quite different from each other. So we can further take a look at the cumulative distribution function(c.d.f.) of them. We can use command **dfittool** to implement this method in MATLAB.

Again, we can see the two cumulative distribution functions in the example in

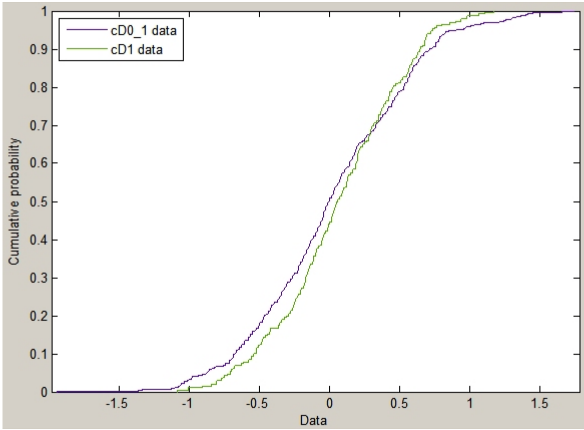


Figure 3.11: C.d.f. of wavelet coefficients of fBm

Figure 3.11 do not match each other perfectly, but we roughly can speculate that they come from the same distribution. To confirm our assumption, we can try to use *Kolmogorov -Smirnov* test to confirm it.

	$d_{1,k}$	$d_{2,k}$	$d_{3,k}$	$d_{4,k}$	$d_{5,k}$	$d_{6,k}$	$d_{7,k}$	$d_{8,k}$	$d_{9,k}$
$d_{0,k}$	0	0	0	0	0	0	0	0	0

Figure 3.12: Return value of **kstest2** of wavelet coefficients

We also can get the return values all zeros when using *Cramer Von Mises* test to reconfirm our result. Now, we are confident to say Relation 3.14 holds and fBm is H-ss. In fact, though we are using hypothesis test to confirm sample data come from the same distribution. However, it does not mean the method of comparison of histograms and cumulative distribution functions are not convincing. If we enlarge the sample size of fBm to 2^{20}, we can see both histograms and c.d.f. of *wavelet* coefficients from each level with proper scaling match each other very well, and related results are shown in **Appendix B**.

Having verified self-similarity and discussed the expected values of *wavelet* coefficients, We can turn to analyse variance structure of them. In the third column of the table in Figure 3.9, I calculated $\frac{Var[d_{j+1,k}]}{Var[d_{j,k}]}$ instead of directly checking part(2) in **Proposition 10**. That is because if $E[d_{j,k}^2] = 2^{j(2H+1)}E[d_{0,0}^2]$ holds for all $j \in \mathbb{Z}$, equivalently,

$$\begin{aligned} E[d_{j-1,k}^2] &= 2^{(j-1)(2H+1)}E[d_{0,0}^2] \\ &= 2^{(j-1)(2H+1)}\frac{E[d_{j,k}^2}{2^{j(2H+1)}} \\ &= \frac{E[d_{j,k}^2]}{2^{2H+1}}. \end{aligned}$$

Surely, the relation above is equivalent to $Var[d_{j,k}] = 2^{2H+1}Var[d_{j-1,k}]$, for all $j \in \mathbb{Z}$. And this relation is clearer and easier to check than part(2) in **Proposition 10**. In our case, H equals to 0.5, which implies that $\frac{Var[d_{j,k}]}{Var[d_{j-1,k}]} = 2^{2\times 0.5+1} = 4$. The relative error of $\frac{Var[d_{j,k}]}{Var[d_{j-1,k}]}$ away from 4 is shown in the fifth column in the table in Figure 3.9. I think the relative errors are small enough that we can agree on part(2) in **Proposition 10**. To confirm it, as I have mentioned before, we can take the logarithm with respect to 2 of part(2) in **Proposition 10**. If so,

$$\begin{aligned} E[d_{j,k}^2] &= 2^{j(2H+1})E[d_{0,k}^2] \\ \log_2(Var[d_{j,k}]) &= (2H+1)j + \log_2(Var[d_{0,k}]). \end{aligned}$$

When $H = \frac{1}{2}$, the slope of the line function with respect to j is $2H+1 = 2$. If plotting it in MATLAB with discrete points when taking different integer values of j, and use command **polyfit** to approximate the line who has the least square distances away from these points.

```
>> polyfit(x,y',1)

ans =

   1.984156285284708  -4.164156855016389
```

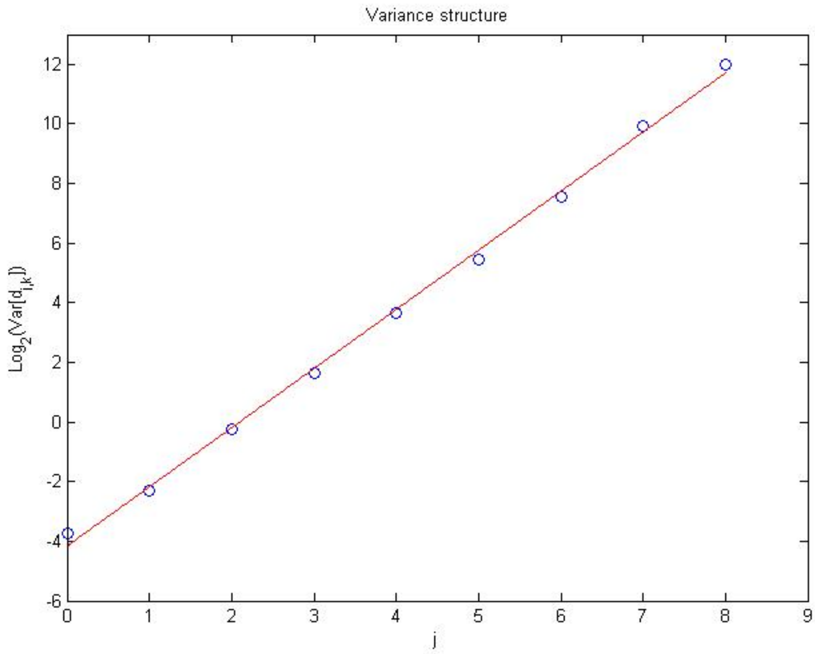


Figure 3.13: Check variance structure by taking logarithm

From the result of **polyfit**, we can see the slope is very close to the supposed value 2 and the interception -4.164156855016389 is also very close to $\log_2(Var[d_{0,k}])$. From Figure 3.13, it is also clear that all the plotting discrete points are almost on the same line, which reflects that fBm has stationary increments, and it is an H-sssi process but not only H-ss. The variance structure of fBm with sample size 2^{20} is also shown in **APPENDIX B**.

Finally, we need to check part(3) in **Proposition 10**. In fact, we are aiming at ensuring the fact that the *wavelet* coefficients are almost uncorrelated and the correlation tends to zero very fast when N is sufficiently large. According to the inequality $E[d_{j,k_1}d_{j,k_2}] \leqslant C|k_1 - k_2|^{2(H-N)}$, for fixed j and $|k_1 - k_2|$, we can calculate $E[d_{j,k_1}d_{j,k_2}]$ as the same as covariance of *wavelet* coefficients, since $Cov[d_{j,k_1}, d_{j,k_2}] = E[d_{j,k_1}d_{j,k_2}] - E[d_{j,k_1}]E[d_{j,k_2}]$ where $E[d_{j,k_1}] = E[d_{j,k_2}] = 0$. Up to covariance calculation, for example, when $j = 0$ and $|k_1 - k_2| = 1$, $E[d_{0,k_1}, d_{0,k_2}]$ should be the covariance between two following columns:

$$\begin{pmatrix} d_{0,1} & d_{0,2} \\ d_{0,2} & d_{0,3} \\ d_{0,3} & d_{0,4} \\ \vdots & \vdots \\ d_{0,n-2} & d_{0,n-1} \\ d_{0,n-1} & d_{0,n} \end{pmatrix}.$$

In order to get all the covariances of *wavelet* coefficients with the same level j but different $|k_1 - k_2|$, we can use the m-file **LINE.m**:

```
function [ U V ] = LINE( S )
n = length(S);
u = 1;
while u <= n-1
    U(u) = u;
    u = u+1;
end
w = 1;
while w <= n-1;
    v = 1;
    while v <= n-w;
        W(v) = (S(v)-mean(S(1:n-w)))*(S(v+w)-mean(S(w+1:n)));
        v = v+1;
    end
    V(w) = abs(mean(W));
    w = w+1;
end
plot(U,V,'o');
```

```
axis auto;
end
```

If we take a look at $d_{0,k}$ from generated fBm as an example: From Figure(3.13),

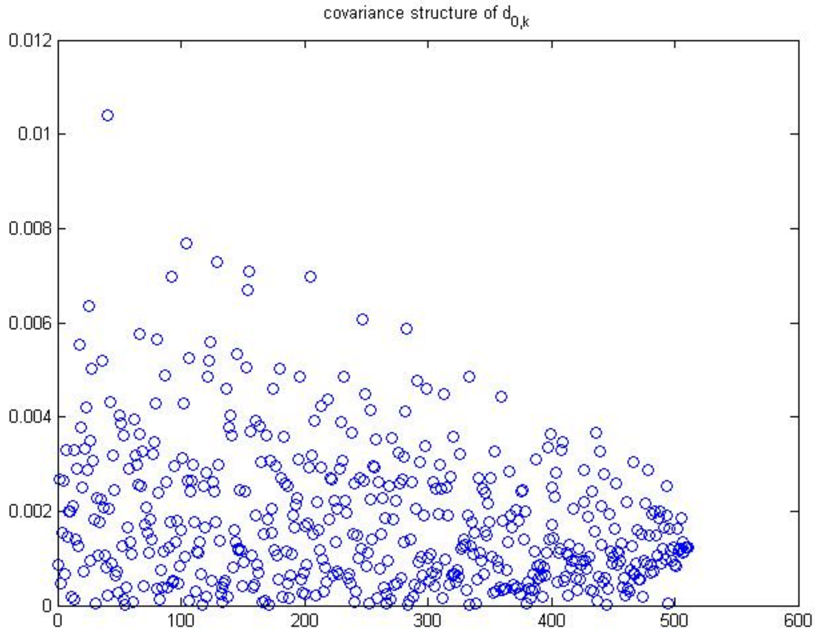


Figure 3.14: covariance structure of wavelet coefficient

we can see that the covariance of *wavelet* coefficients decay to zero when $|k_1 - k_2|$ is getting bigger. It confirms that the correlation of *wavelet* coefficients does decay as $|k_1 - k_2|$ increases and it is almost uncorrelated and decays fast when the lag is large. However, this pattern is not always clear in numerical test when the data are generated randomly, and it turns more unconvincing when the sample size is not big enough. If we enlarge the sample size to 5000, we can see the decaying of the covariance is clearer, which is shown in **APPENDIX C**.

To conclude, based on our example of testing fBm, we are confident to say that the test method of following **Proposition 9** and **10** is valid to check self-similarity and

$H - sssi$ process. In the future work, we can use this method to check self-similarity and stationary increments of the time series from financial data and chaotic dynamic systems.

Chapter 4: Detect self-similarity on financial data

In this chapter, we will focus on detecting self-similarity of financial data. The target financial data will include Dow Jones indices, NASDAQ indices, and several separate quotes from these stock markets.

4.1 Dow Jones indices

First of all, we can download the historical prices of Dow Jones and write them in an Excel file from http://finance.yahoo.com. The data we are going to analyse is the *Close price* column and we can load the target data into MATLAB by using command **xlsread**. Next, we will use command **wavedec** and **detcoef** again to calculate the *wavelet coefficients* of *Close price* of Dow Jones indices. The expected value, variance and related information are shown in the tables in Figure 4.1.

According to the two tables in Figure 4.1, from the upper one, we can see that expected value of Dow Jones indices itself is far away from zero, which results the expected values of *wavelet coefficients* do not follow part(1) in **Proposition 10** very well. That is because one of the conditions of **Proposition 10** is that the stochastic process should be mean-zero Gaussian process, however, we can fix this problem by standardizing the original Dow Jones indices, so that we get the lower table, in which we can see that the expected value of *wavelet* coefficients are relatively much closer to zero. In addition, it is nice to see that the variance structure of *wavelet* coefficients are not ruined when comparing the two $\frac{d_{j+1,k}}{d_{j,k}}$ columns, they turn to be exact the same in both tables. That is we do not need to worry about the expected value of *wavelet* coefficients part in **Proposition 10** any more in our future work since we always can fix it by standardizing the original given data sample, and this process

	Expected Value	Variance	$d_{j+1,k}/d_{j,k}$	Relative Error
Dow Jones Data	2659.870455	14470136.22		
$d_{0,k}$	0.594695297	1462.04412	2.665795779	0.305469742
$d_{1,k}$	0.101232341	3897.511042	3.820992618	0.004501766
$d_{2,k}$	1.969646074	14892.36092	3.269694249	0.148133698
$d_{3,k}$	8.332236641	48693.46686	3.565815495	0.070984066
$d_{4,k}$	28.05862862	173631.9186	4.899234813	0.276416913
$d_{5,k}$	112.4184032	850663.5403	3.240976915	0.155615538
$d_{6,k}$	254.4119052	2756980.897	5.868466031	0.528934534
$d_{7,k}$	513.4669893	16179248.74	2.55854025	0.333413447
$d_{8,k}$	1890.727357	41395259.12	4.898651586	0.276264963
$d_{9,k}$	5384.939375	202780951.7	1.194945815	0.688676067
$d_{10,k}$	6998.36707	242312249.7	4.218998067	0.099192159
$d_{11,k}$	23302.80359	1022314913	39.07574778	9.180558244
$d_{12,k}$	122039.78	39947719702	1.523028885	0.603199295
$d_{13,k}$	174415.4967	60841530991		
MATLAB Estimated H =	3.83827162	MATLAB Estimated H =	0.466547912	
Average of $d_{j+1,k}/d_{j,k}$ =	3.477095042	Estimated H =	0.398941251	
First 5-level $d_{j+1,k}/d_{j,k}$ =	3.644306591	Estimated H =	0.432822168	

	Expected Value	Variance	$d_{j+1,k}/d_{j,k}$
New Dow Jones Data	-6.94E-15	1	
$d_{0,k}$	0.000156336	0.000101039	2.665795779
$d_{1,k}$	0.000026612	0.000269349	3.820992618
$d_{2,k}$	0.000517788	0.001029179	3.269694249
$d_{3,k}$	0.002190409	0.003365101	3.565815495
$d_{4,k}$	0.007376157	0.011999329	4.899234813
$d_{5,k}$	0.029552968	0.058787528	3.240976915
$d_{6,k}$	0.066880748	0.190529022	5.868466031
$d_{7,k}$	0.134982112	1.118113091	2.55854025
$d_{8,k}$	0.49704144	2.860737349	4.898651586
$d_{9,k}$	1.41561289	14.01375555	1.194945815
$d_{10,k}$	1.839756763	16.74567855	4.218998067
$d_{11,k}$	6.125927675	70.64998545	39.07574778
$d_{12,k}$	32.08227126	2760.701012	1.523028885
$d_{13,k}$	45.85099445	4204.627384	
Estimated H =	3.83827162		

Figure 4.1: Wavelet coefficients of Dow Jones and Standardized Dow Jones data

will not influence the distribution of *wavelet* coefficients.

From the upper table in Figure(4.1), the relative error column refers to the relative error between $\frac{d_{j+1,k}}{d_{j,k}}$ and the term 2^{2H+1} with estimated parameter H from MATLAB. In MATLAB, we can use command **wfbmesti** to estimate the Hurst parameter H from H-sssi process. From both the $\frac{d_{j+1,k}}{d_{j,k}}$ and the relative error columns, it is obvious that the value of $\frac{d_{12,k}}{d_{11,k}}$ is far greater than others, which can be treated as an obvious big error. So we can get rid of it and take the average of $\frac{d_{j+1,k}}{d_{j,k}}$ column and this is the way used to evaluate the Hurst parameter H that we mentioned it before. Recall we also have mentioned in the analysis of fBm data that we always take the first several levels of *wavelet coefficients* in wavelet analysis, since their errors are comparatively small. Hence, we can also take the first five levels of $\frac{d_{j+1,k}}{d_{j,k}}$ and obtain another estimated Hurst parameter. We can see such estimation of Hurst parameter is closer to the parameter estimated by MATLAB.

Having confirmed the expected value of **wavelet** coefficients around zero and acknowledged the way of estimating Hurst parameter H, we can keep on checking self-similarity and H-sssi features of Dow Jones indices now. Again, let us take a look at the histograms of *wavelet* coefficients of Dow Jones indices:

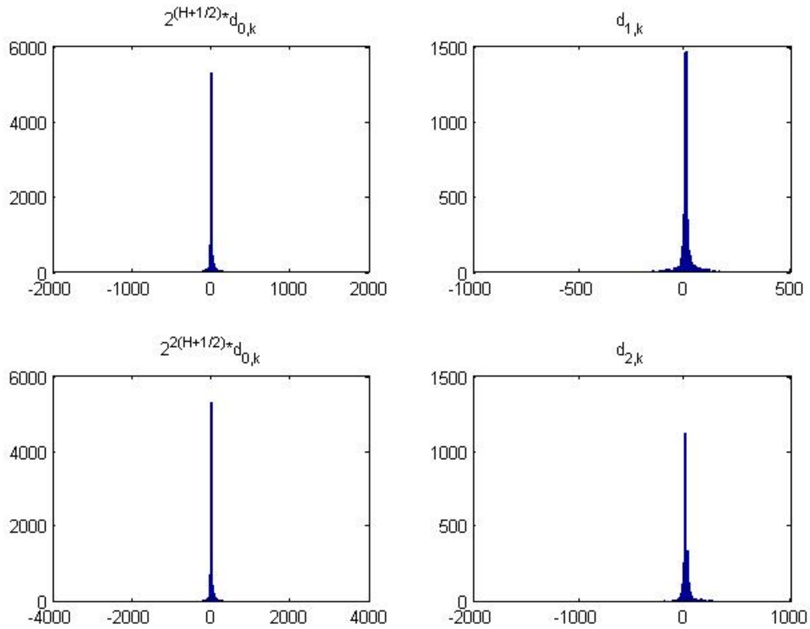


Figure 4.2: Histograms of wavelet coefficients of Dow Jones indices

We can see that the histograms in Figure 4.2 look really alike, even better than the histogram of fBm. It can be understood since the sample size of Dow Jones indices is bigger than 2^{13}, which is much bigger than our example of fBm. So if it is self-similar, the histogram of its *wavelet* coefficients with proper scales should be very similar to each other. We can get the initial conclusion that Dow Jones indices is self-similar in general. Next, let us take a look at c.d.f. of Dow Jones indices to confirm our initial conclusion:

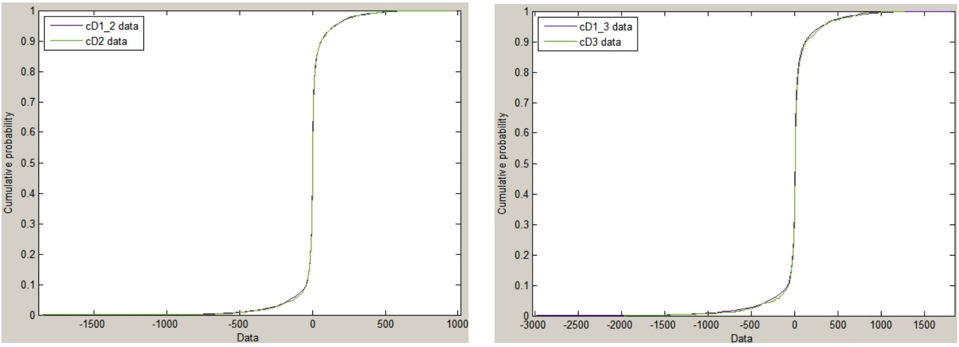


Figure 4.3: Cumulative distribution functions of wavelet coefficients of Dow Jones indices

From the examples in Figure 4.3, we can see the c.d.f. of different *wavelet* coefficients match each other almost perfectly. Thus, it appears that Dow Jones indices has self-similar features.

Next step, let us further check if Dow Jones indices have $H - sssi$ features. It is the same as checking the variance structure of *wavelet* coefficients. Again, if taking the logarithm of the equation in part(2) in **Propostion 10**, we can plot it and get Figure 4.4.

From Figure 4.4, we can see that the logarithm of variance structure of the first nine *wavelet* coefficients are almost on the line, but last several points are comparatively far away the line, which implies that Dow Jones indices do not have stationary increments. We can go back to take a look at Dow Jones indices data figure to try to find out the reason of ambiguity of stationary increments.

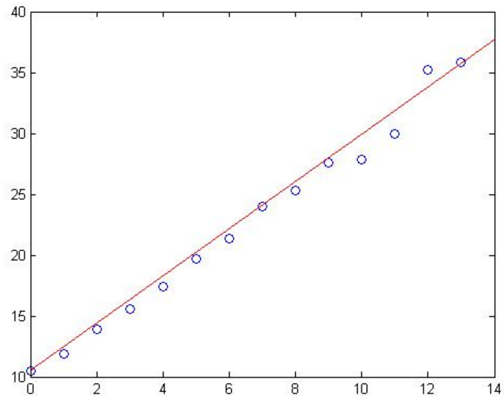


Figure 4.4: Variance structure of dow jones indices

From Figure 4.5, it seems that the increments of Dow Jones indices are quite different. Roughly, from 1 to 5000, the increments look like bigger than the rest of them. So we can divide the entire Dow Jones indices into two main parts, from 1 to 5000 and the rest. We can take a sample from each part, and a sample with data from both of the two parts to compare their variance structure. When choosing the sample size, we can always choose $4096 = 2^{12}$ to speed up the calculation in MATLAB:

From Figure 4.6, we can see in each part,the discrete points are almost on the same line, which means they almost have stationary increments in each sample. However, if we take a further look at the red lines, we realize that the slopes are a little different from each other, which means that the samples have stationary increments under different parameters H in each part. If we try to estimate the Hurst parameter of each part, we can get the following table:

We can see we have different estimation of H for different parts, and such difference results in the fact that increments vary overall. If taking a look at the relative errors of parameters H, we can see though the increments are not stationary, they change slightly. We can reach a more proper conclusion: Dow Jones indices do not have strictly stationary increments, but they are almost stationary.

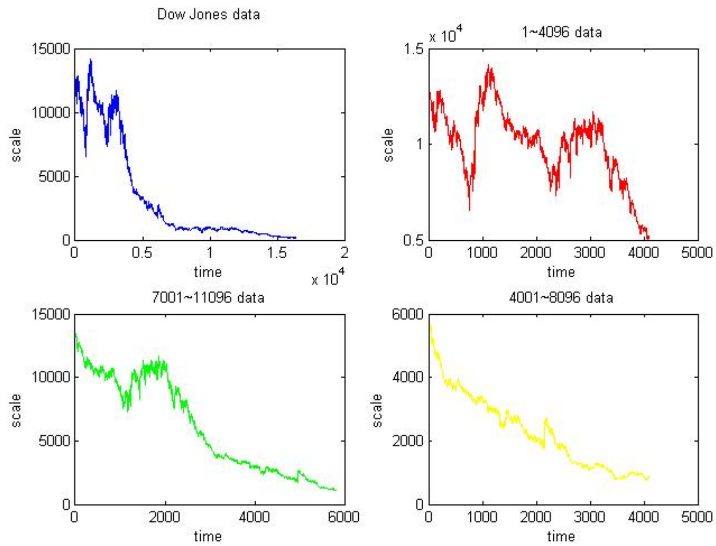


Figure 4.5: Dow Jones indices

Finally, let us take a look at the covariance structure of Dow Jones indices by using m-file **LINE.m**. We can see the decaying of the covariance of *wavelet* coefficients clearly. For example, the covariance structure of $d_{2,k}$ is shown below:

To conclude, from the analysis of *wavelet* coefficients of Dow Jones indices, we know close prices of Dow Jones indices have the features of self-similar processes and almost have stationary increments. However, if we look at individual stocks in the group of 30 companies that Dow Jones index represents instead of analysing the whole index component, it would be a very interesting topic to check if they also have self-similar features or stationary increments. We will discuss this topic in the next section.

4.2 Stocks from Dow Jones

In this section, we will focus on some individual stocks and use the same method to determine if they have self-similar features or stationary increments. Let us review the whole process of testing at first. Remember that we have discussed that we do not

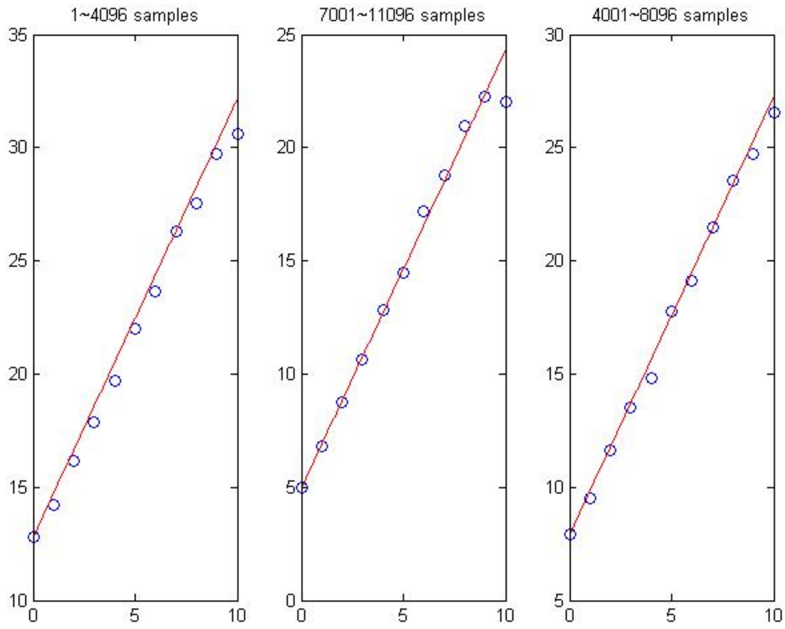


Figure 4.6: Variance structure of different parts of Dow Jones indices

MATLAB Estimated H =	0.466547912			
	Wavelet Analysis H	Relative Error	First 5-level Analysis H	Relative Error
Dow Jones Data	0.398941251	0.144908292	0.432822168	0.072287847
(1~4096 Samples)	0.430605536	0.077038982	0.43124891	0.075659972
(7001~11096 Samples)	0.426592789	0.085639914	0.455499277	0.023681673
(4001~8096 Samples)	0.398597555	0.145644971	0.546075059	0.170458692

Figure 4.7: Comparison of Hurst parameter of different parts of Dow Jones indices

need to worry about the first criterion, $E[d_{j,k}] = 0$, in **Proposition 10**, since it always can be fixed manually. So our test will have the following main steps: (1) check self-similarity by comparing histogram and c.d.f. of *wavelet* coefficients; (2) check variance structure of *wavelet* coefficients to determine stationary increments; (3) construct covariance structure to determine the correlation between *wavelet* coefficients.

Before starting our test, we can pick up four quotes from Dow Jones and download them from the same links we mentioned in the previous section. Here, the stocks we are going to test are Boeing company, Johnson & Johnson, JP Morgan Chase and

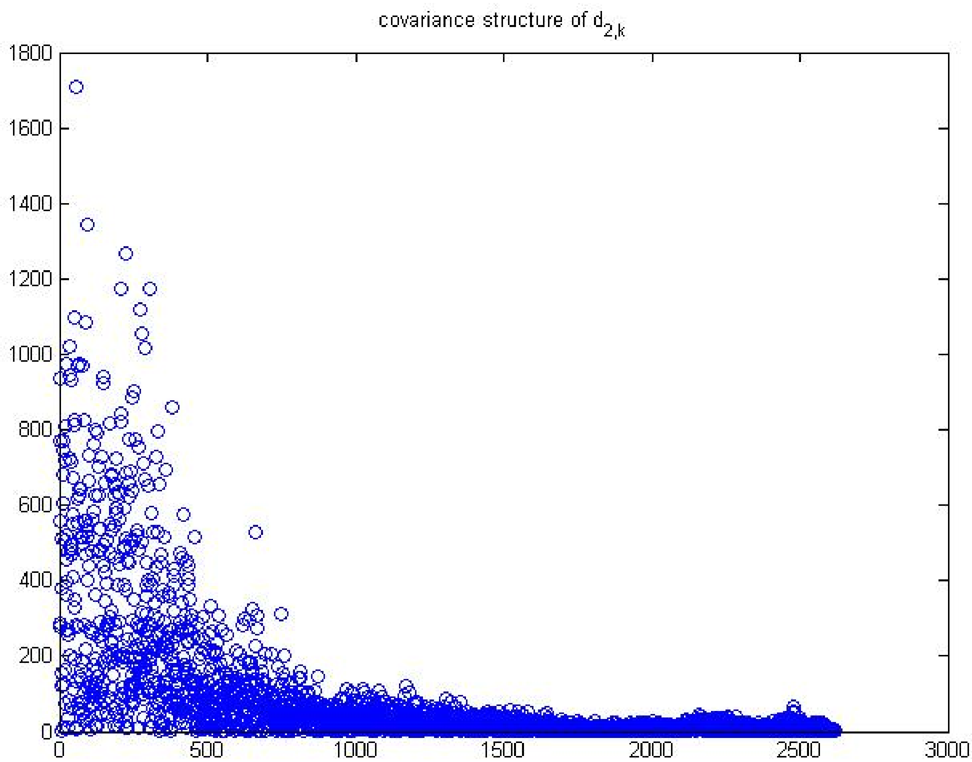


Figure 4.8: Covariance structure of Dow Jones indices

Walt Disney company. These four companies have totally different volumes and their prices vary. As our previous work, we can write the data into excel file, and read them in MATLAB. Then, if running commands **wavedec** and **detcoef**, we can obtain their *wavelet* coefficients. Then we can use command **wfbmesti** to estimate the Hurst parameter for each data set. Once having all these information, we can plot the histograms of them to check if they have self-similar features.

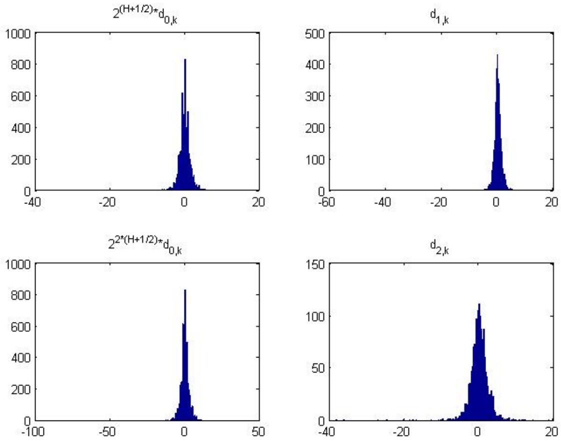


Figure 4.9: Histogram of Boeing company from Dow Jones

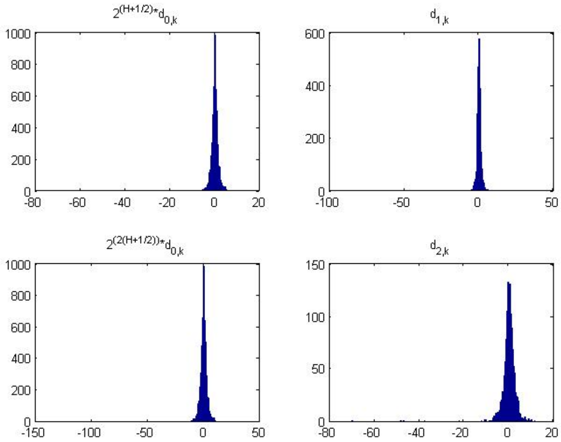


Figure 4.10: Histogram of Johnson & Johnson company from Dow Jones

Here, the histograms of Boeing company and Johnson & Johnson are posted, and we can see that the histograms do not look like perfectly match but roughly the same. We are able to get more convincing conclusion if plotting their c.d.f.

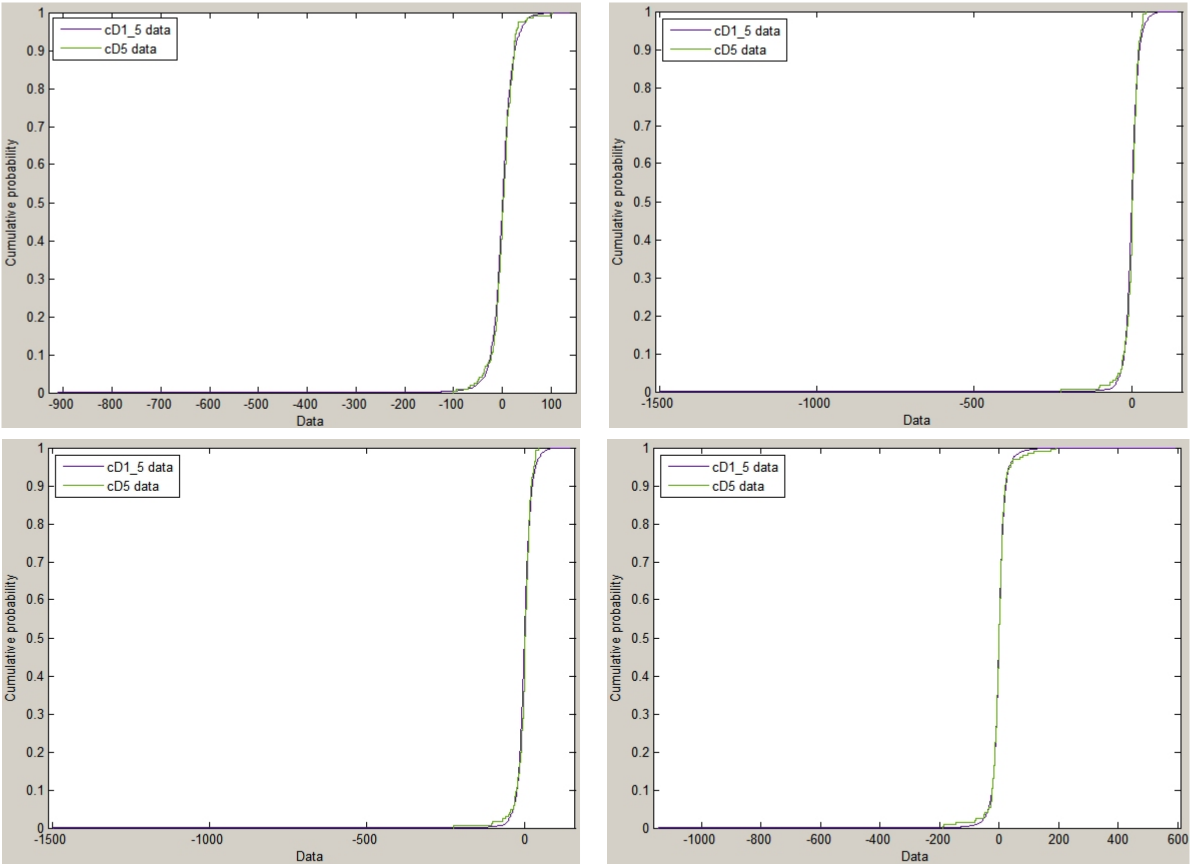


Figure 4.11: c.d.f. of stocks from Dow Jones

We can see that the c.d.f. of *wavelet* coefficients with different levels almost perfectly match each other, which confirms that both of them are self-similar. The same conclusions can be obtained on JP Morgan and Walt Disney company.

Then, we can move to study the variance structure of *wavelet* coefficients of these four companies. We need to get the similar table as in Figure 4.1.

	Variance	$d_{j+1,k}/d_{j,k}$		Variance	$d_{j+1,k}/d_{j,k}$
Boeing Company	517.0639511		Johnson&Johnson	512.300467	
$d_{0,k}$	0.748670145	3.852247974	$d_{0,k}$	0.835551366	9.668918127
$d_{1,k}$	2.884063047	3.447739847	$d_{1,k}$	8.078877746	1.80217796
$d_{2,k}$	9.943499089	5.823866939	$d_{2,k}$	14.55957541	7.180387497
$d_{3,k}$	57.9096156	2.892148613	$d_{3,k}$	104.5433933	2.429638287
$d_{4,k}$	167.4832144	3.421108196	$d_{4,k}$	254.0026309	3.005958782
$d_{5,k}$	572.9781976	3.422686667	$d_{5,k}$	763.5214389	7.016974424
$d_{6,k}$	1961.124838	7.357575638	$d_{6,k}$	5357.610409	4.354771665
$d_{7,k}$	14429.12433	2.426328713	$d_{7,k}$	23331.17	2.691497997
$d_{8,k}$	35009.79867	4.542104058	$d_{8,k}$	62795.79732	1.376392685
$d_{9,k}$	159018.1486	1.078224729	$d_{9,k}$	86431.6761	0.767467867
$d_{10,k}$	171457.3001	1.941837343	$d_{10,k}$	66333.53413	2.160503599
$d_{11,k}$	332942.1882	0.533419865	$d_{11,k}$	143313.8392	0.200520399
$d_{12,k}$	177597.9771		$d_{12,k}$	28737.34825	
	Variance	$d_{j+1,k}/d_{j,k}$		Variance	$d_{j+1,k}/d_{j,k}$
JP Morgan Chase	458.7123096		Walt Disney Company	1201.557602	
$d_{0,k}$	0.700893625	6.608998767	$d_{0,k}$	3.314465451	1.776698309
$d_{1,k}$	4.632205105	1.944782868	$d_{1,k}$	5.888805162	8.165858348
$d_{2,k}$	9.008633132	3.343839675	$d_{2,k}$	48.0871488	1.739113831
$d_{3,k}$	30.12342489	2.779353051	$d_{3,k}$	83.62902556	6.535161243
$d_{4,k}$	83.72363288	3.585704981	$d_{4,k}$	546.5291666	2.868633935
$d_{5,k}$	300.2082474	4.413868292	$d_{5,k}$	1567.792114	2.38535238
$d_{6,k}$	1325.079664	11.79218211	$d_{6,k}$	3739.73665	11.65777423
$d_{7,k}$	15625.58072	1.140309507	$d_{7,k}$	43597.00553	0.867584926
$d_{8,k}$	17817.99824	1.761839705	$d_{8,k}$	37824.10481	8.016405707
$d_{9,k}$	31392.45676	10.37940776	$d_{9,k}$	303213.3697	1.899913727
$d_{10,k}$	325835.1092	0.378994358	$d_{10,k}$	576079.2432	1.134409906
$d_{11,k}$	123489.6681		$d_{11,k}$	653510	7.296618205
			$d_{12,k}$	4768412.963	

Figure 4.12: Variance information about stocks from Dow Jones

We can see that the values in the column $\frac{d_{j+1,k}}{d_{j,k}}$ from each stock vary from each other, and the differences are greater than the counterpart from Dow Jones indices. This implies that stationary increments do not exist in separate stocks. It will be clearer if we try to plot $(j, \log_2(E[d_{j,k}^2]))$, see Figure 4.13. From these variance structure figures, we can see that none of them has very particular features of stationary increments. Especially Walt Disney company, almost no point is on the expected variance line. We can get the conclusion that the assumption of stationary increments is not true for separate stock from Dow Jones.

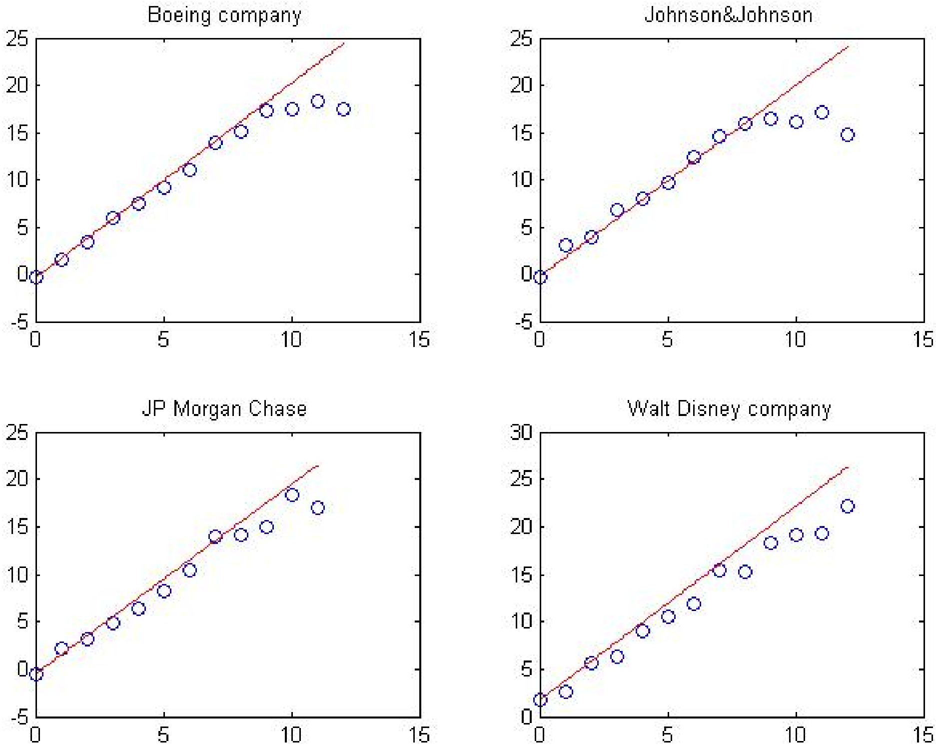


Figure 4.13: Variance structure of wavelet coefficients of stocks from Dow Jones

Finally, we check the covariance structure of *wavelet* coefficients from these four stocks. This time, we are using JP Morgan Chase and Walt Disney Company as examples to detect covariance structure of *wavelet* coefficients. From Figure 4.14, we can see that the covariances of them decay to zero very fast when the lag is large.

From the test above, we can conclude that separate stocks from Dow Jones have features of self-similarity but not stationary increments. The correlation of *wavelet* coefficients exists but decays very fast. In other words, they are self-similar but not $H-sssi$ processes. Besides, as mentioned at the beginning of this section, we also have done similar tests on NASDAQ indices and four stocks(Amazon, Columbia, Ebay, Microsoft) from NASDAQ, and obtained very similar conclusions. NASDAQ

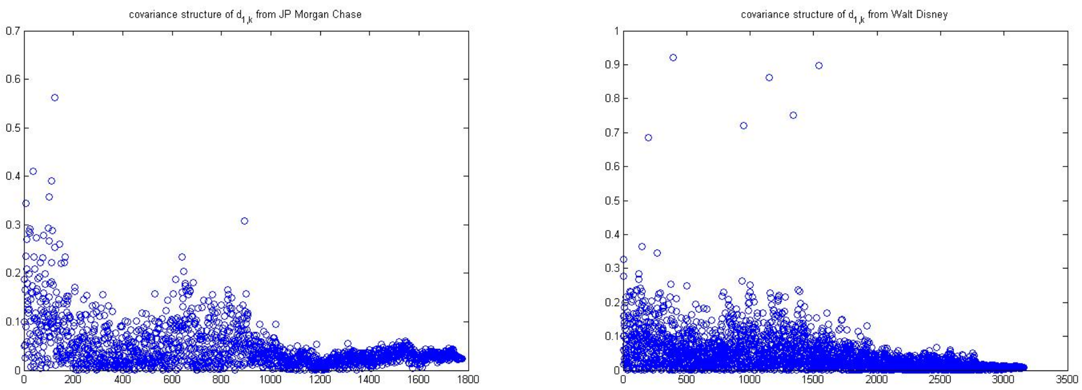


Figure 4.14: Covariance structure of wavelet coefficients of stocks from Dow Jones

indices are self-similar and the increments are almost stationary. However, the stocks from NASDAQ are only self-similar, but do not have stationary increments either. Moreover, the scales of the companies from NASDAQ are smaller than those from Dow Jones comparatively, and most of them are more recently listed in the market. Their increments turn to be more unstable than the larger-scaled companies.

At the end of this section, I want to point out that self-similarity has become one of the most important features of financial data, and fractional Brownian motion is one of the best models in financial analysis. Wavelet analysis is the main method to determine self-similarity in financial analysis, too. So far, the studies on self-similar financial time series have determined that the *wavelet* coefficients are totally uncorrelated and the increments are independent when the estimated Hurst parameter $H = 0.5$. That is the reason why the estimated H from Dow Jones, NASDAQ, and their stocks are close to 0.5. More explanations and applications of wavelet analysis on financial data refer to researches done by Vuorenmaa[22] and Wingi[23].

Chapter 5: Detect self-similarity on chaotic dynamical systems

The dynamics of spatial averages of dynamical systems with chaotic behaviour is one of the most popular topics. The exploration of self-similar patterns mainly depends on the time series simulated by Jiang[30]. Though the process of data generation is not the main topic, I still will give a briefly introduction.

Let $f_\lambda(x) = \lambda x(1-x)$ be a quadratic family with parameter $2 \leqslant \lambda \leqslant 4$, the target is to investigate the dynamics of spatial average for the entire family $A(n,\lambda) = \int_0^1 f_\lambda^n(x)dx$, where f_λ^n is denoted by n-th iteration of composite of f_λ. Since the function $f_\lambda^n(x)$ is very difficult to integrate when the iteration n is getting big. Jiang[30] used Monte Carlo and Quasi Monte Carlo integrations to evaluate $A(n,\lambda)$. More strictly, Jiang[] restricted the range of λ into $[3.5, 4]$ and set the iteration $41 \leqslant n \leqslant 90$; then he set iteration n unchanged, but shrink the range of λ into $[3.95, 4]$, $[3.989, 3.994]$ and $[3.991, 3.9915]$ respectively (i.e. in each step, new interval is a part of previous interval), to explore the tongue-like structure for more details of dynamics. Jiang[] claimed if taking time average of $A(n,\lambda)$: $STA(\lambda) = \lim_{N\to+\infty} \frac{1}{N}\sum_{n=0}^{N-1} A(\lambda,n)$ has self-similar properties on the previous given range of λ.

Let us take a look at the data path simulated on $[3.5, 4]$, shown in Figure(5.1). Take DWT on the data with wavelet *db1*, and we can get such *wavelet* coefficients as usual. First of all, we need to make sure if they follow self-similarity properties. As what we have done previously, we can take commands **hist** and **dfittool** to compare the histograms and c.d.f. of *wavelet* coefficients.

From Figure(5.2), it seems that the histograms do not match each other again. Though c.d.f. are still match pretty well, they still look a little more different between

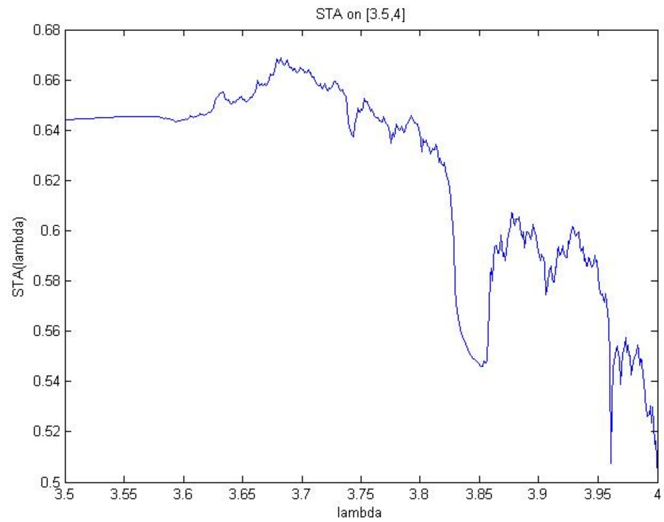


Figure 5.1: STA(λ) on $[3.5, 4]$

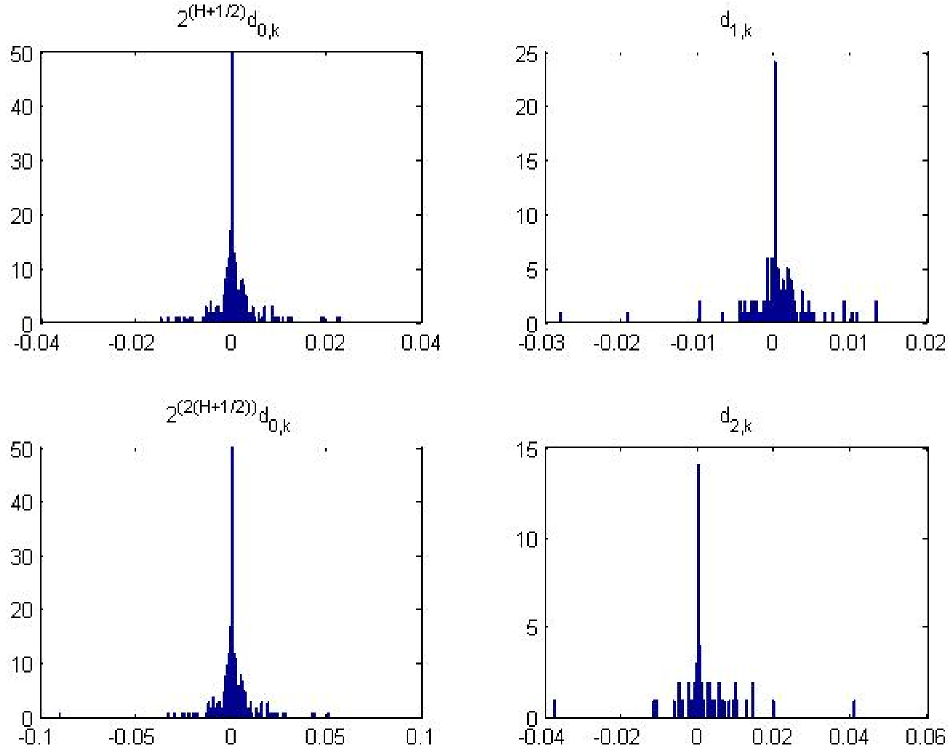


Figure 5.2: Histogram of wavelet coefficients from STA(λ) on $[3.5, 4]$

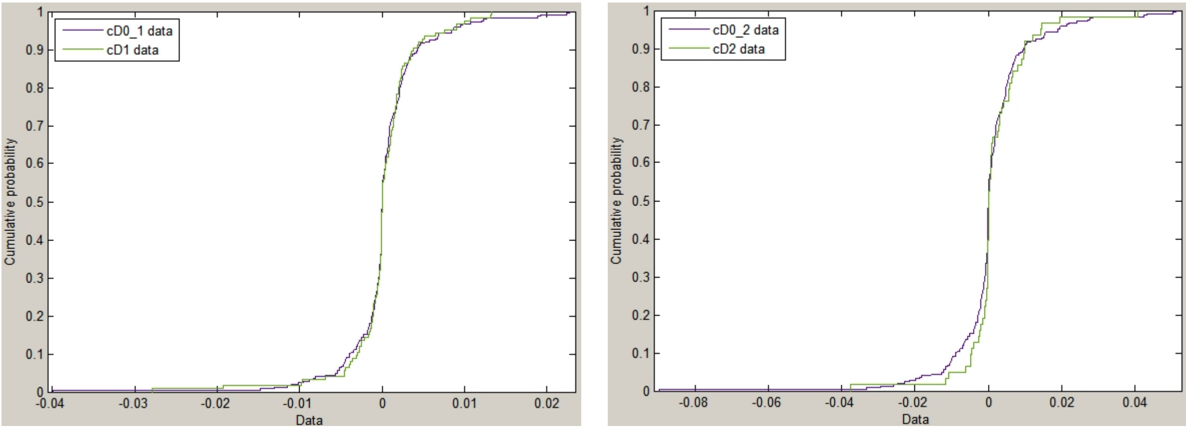


Figure 5.3: c.d.f. of wavelet coefficients from STA(λ) on $[3.5, 4]$

each pair than c.d.f. we plotted in previous cases. At this point, we are probably not as confident as in previous case to assure the data from $STA(\lambda)$ on $[3.5, 4]$ is self-similar. In order to ensure the assumption of self-similarity is correct, we can run Kolmogorov-Smirnov two-sample Test in MATLAB to confirm.

```
>> h = kstest2(cD0_1,cD1)

h =

     0
```

If pairwise testing all the *wavelet coefficients*, we can find that all the outputs h return 0, which impliess to accept the null hypothesis. That is, each pair of *wavelet coefficients* come from the same distribution, and this strongly suggests that data from $STA(\lambda)$ on $[3.5, 4]$ have self-similar properties.

Next, we can take a look at the variance of *wavelet* coefficients and check if it has stationary increments to make it an $H - sssi$ series.

	Expected Value	Variance	dj+1,k/dj,k	Relative Error
STA in [3.5,4]	0.62192776	0.00154909		
$d_{0,k}$	-0.00017440	0.00000525	4.03202931	0.20297494
$d_{1,k}$	-0.00023896	0.00002116	3.75503415	0.25772952
$d_{2,k}$	0.00166978	0.00007945	7.83968049	0.54969653
$d_{3,k}$	0.00442750	0.00062286	2.52224856	0.50141848
$d_{4,k}$	0.01636356	0.00157101	2.05389512	0.59399951
$d_{5,k}$	0.00760844	0.00322670	9.99267169	0.97528568
$d_{6,k}$	0.15844849	0.03224333	2.88435566	0.42983953
$d_{7,k}$	0.12414688	0.09300122		
MATLAB Estimated H =	5.05884885	MATLAB Estimated H =	0.66940457	
Average of dj+1,k/dj,k =	4.72570214	Estimated H =	0.62026435	0.07340885

Figure 5.4: Wavelet coefficients from STA(λ) on $[3.5, 4]$

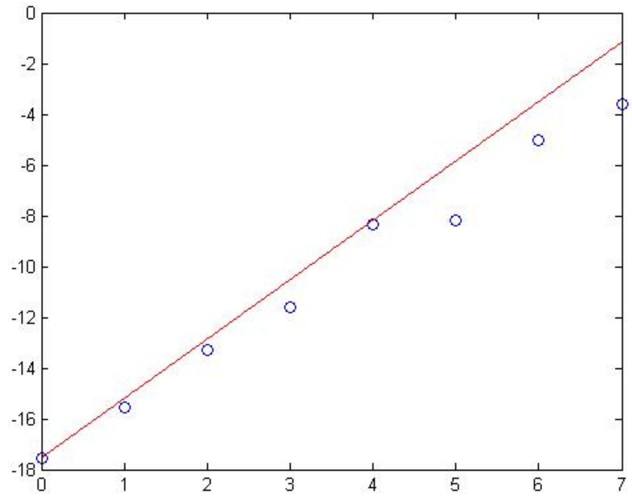


Figure 5.5: Variance structure of wavelet coefficients from STA(λ) on $[3.5, 4]$

From Figure 5.4, we can see that the values in column $\frac{d_{j+1,k}}{d_{j,k}}$ are very different from each other, and not even close to the estimated parameter H from MATLAB. This means the time series does not have stationary increments at all. More clearly, from Figure 5.5, it shows that most blue plots are out of line, and some of them are not even close, which disobeys the patterns of stationary increments.

Finally, I can move to the covariance structure of *wavelet* coefficients, shown in Figure 5.6. We can see that though the trend of decaying is not obvious, the

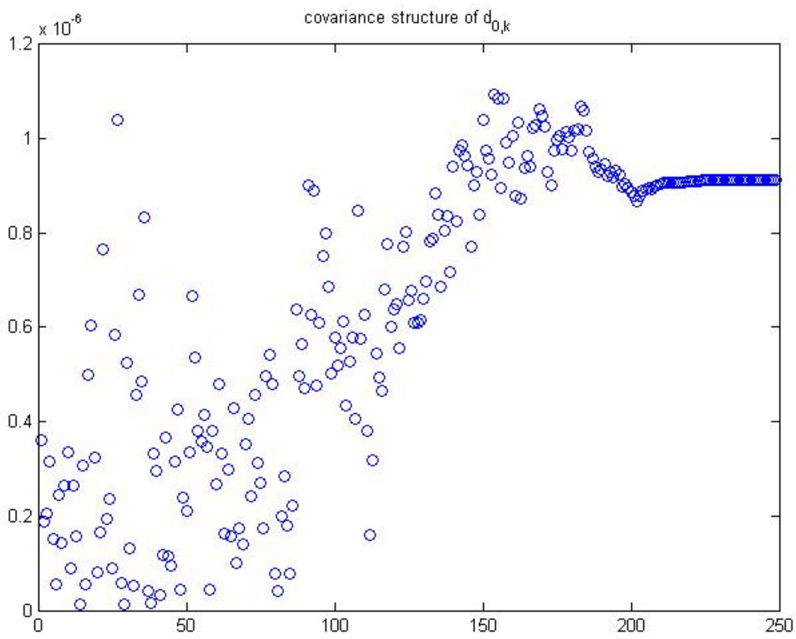


Figure 5.6: Covariance structure of wavelet coefficients from STA(λ) on $[3.5, 4]$

correlation of *wavelet* coefficients is around 10^{-6}, and it is small enough to present that *wavelet* coefficients of spacial average of time series from chaotic dynamic systems are generally uncorrelated.

To sum up, after the test on $STA(\lambda)$ on $[3.5, 4]$, we have the conclusion that these 500 data are self-similar but they do not have stationary increments. So this sample is a self-similar process but not $H{-}sssi$ process. When we test the other spatial averages of time series from smaller intervals, $[3.95, 4]$, $[3.989, 3.994]$ and $[3.991, 3.9915]$, we can obtain very similar conclusion.

Bibliography

[1] Pardo, J. C. "A brief introduction to self-similar processes." Department of Mathematical Sciences, University of Bath, United Kingdom, 2007.

[2] Mandelbrot, B. B. "How Long is the Coast of Britain?." *Science*, Volume 156, No. 3557: 636-638, 1967.

[3] Taqqu, Murad S. "Fractional Brownian Motion and Long-Range Dependence." *Theory and Applications of Long-Range Dependence*. Eds. Paul Doukhan and Georges Oppenheim. Boston: Birkhuser, 2003.

[4] Brummelhuis, Raymond. "Lecture 7: Characteristic functions of random variables." *Mathematical Methods*. Birkbeck, University of London.

[5] Nolan, John P. "Stable Distributions: Models for Heavy Tailed Data." Boston: Birkhuser, 2012.

[6] Embrechts, Paul and Makoto Maejima. "An Introduction to the Theory of Self-similar Stochastic Processes." *International Journal of Modern Physics B*, 14 (2000): 1399-1420. Print.

[7] Graps, Amara. "An Introduction to Wavelets."*IEEE Computational Science and Engineering*, 2.2 (1995). Web.

[8] Heijmans, Henk J. A. M. "Discrete Wavelets and Multiresolution Analysis." *Series In Approximations and Decompositions*. Ed. Tom H. Koornwinder. Vol. 1. Singapore: World Scientific Publishing Co., Inc, 1993.

[9] Vishwanath, Mohan, Robert Michael Owens, and Mary Jane Irwin. "VLSI Architectures for the Discrete Wavelet Transform." *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 42.5 (1995): 305-316. Print.

[10] Walker, James S. *A Primer on Wavelets and Their Scientific Applications*. Boca Raton, Florida: CRC Press, 1999.

[11] Merry, R.J.E. "Wavelet Theory and Applications: A Literature Study." Technical report, Eindhoven University of Technology (2005).

[12] Narcowich, F. J. Class Notes. "Notes on Daubechies' Wavelets." Math 641. University of Texas A&M. Fall 2009.

[13] Misiti, Michel, et al. "Wavelet Toolbox: For Use with MATLAB." Natick, MA: The MathWorks, Inc, 1996.

[14] Ma, Junxing, et al. "A Study of the Construction and Application of a Daubechies Wavelet-Based Beam Element." *Finite Elements in Analysis and Design*, 39.10 (2003): 965-975. Print.

[15] Vidakovic, Brani. Handouts. "Basics of Wavelets." Georgia Tech University. Fall 2004.

[16] Temme, Nico M. "Asymptotics and Numerics of Zeros of Polynomials That Are Related to Daubechies Wavelets." *Appl. Comp. Harmonic Analysis*, 4 (1997): 414-428.

[17] Flandrin, Patrick. "Wavelet Analysis and synthesis of Fractional Brownian Motion." *IEEE Transactions on Information Theory*, 38.3 (1992): 910-917. Print.

[18] Abry, P., Flandrin, P., Taqqu, M.S. and Veitch, D. "Wavelets for the analysis, estimation and synthesis of scaling data." *Self-Similar Network Traffic and Per-*

formance Evaluation. Eds. K. Park and W. Willinger. 39-88. New York: Wiley, 2000. 39-88.

[19] Dieker, Ton. “Simulation of Fractional Brownian Motion.”Thesis. Vrije Universiteit Amsterdam, 2004. Web.

[20] Abry, Patrice and Fabrice Sellan. “The Wavelet-Based Synthesis for Fractional Brownian Motion Proposed by F. Sellan and Y. Meyer: Remarks and Fast Implementation.”*Applied and Computational Harmonic Analysis*, 3 (1996): 377-383.

[21] ÇİFTLİKLİ, Cebrail and Ali Gezer. “Comparison of Daubechies Wavelets for Hurst Parameter Estimation.”*Turk J Elec Eng & Comp Sci*, 18.1 (2010): 117-128.

[22] Vuorenmaa, Tommi. “The Discrete Wavelet Transform with Financial Time Series Applications.”Seminar on Learning Systems. Rolf Nevanlinna Institute, Helsinki, Finland. 4 September 2003.

[23] Sikazwe, Geofrey Wingi. “Statistical Self-Similarity: Fractional Brownian Motion. Abstract.”10 March 2010.

[24] Patrice Abry, Patrick Flandrin, Murad S. Taqqu, Darrl Veitch. “Self-similarity and long-range dependence through the wavelet lens.” *Theory and Applications of Long-Range Dependence*, 527-556, Eds. Paul Doukhan and Georges Oppenheim. Boston: Birkhuser, 2003.

[25] Mallat, Stephane G. “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation.”*IEEE Transaction on pattern analysis and machine intelligence*, VOL. 11, No. 7. July, 1989.

[26] DeGroot, Morris H., Schervish Mark J. *Probability and Statistics(third edition)*, Addison-Wesley, 2002.

[27] Anderson, T. W. “On the distribution of the two Sample Cramer-von Mises Criterion.” *The Annals of Mathematical Statistics*, Vol.33, No. 3, 1148-1159, Sep. 1962.

[28] Hodges Jr., J. L. “The significance probability of the Smirnov two-sample test.” *ARKIV FOR MATEMATIK*, Band 3, Nr 43 June. 1957.

[29] Tewfik, A. H., Kim, M. “Correlation structure of the discrete wavelet coefficients of fractional Brownian motion.” *IEEE Trans. Inf. Theory*, 38(2), 904-909, 1992.

[30] Jiang, M. “Quasi-Monte Carlo studies of Spatial Averages of Quadratic maps.” *Institute of Mathematics and its Applications*, Preprint, August 1998, No. 1584

.

Appendix A: Cramer Von Mises test M-file

From: Juan Cardelino

```
function [H, pValue, CM_limiting_stat] = cmtest2(x1 , x2 , alpha)
if nargin < 2
error('stats:cmtest2:TooFewInputs','At least 2 inputs are required.');
end
[rows1 , columns1]  =  size(x1);
[rows2 , columns2]  =  size(x2);
if (rows1 ~= 1) & (columns1 ~= 1)
error('stats:cmtest2:VectorRequired','Sample X1 must be a vector.');
end
if (rows2 ~= 1) & (columns2 ~= 1)
error('stats:cmtest2:VectorRequired','Sample X2 must be a vector.');
end
x1  =  x1(~isnan(x1));
x2  =  x2(~isnan(x2));
x1  =  x1(:);
x2  =  x2(:);
if isempty(x1)
error('stats:cmtest2:NotEnoughData',...
'Sample vector X1 is composed of all NaN''s.');
end
if isempty(x2)
error('stats:cmtest2:NotEnoughData',...
```

```
'Sample vector X2 is composed of all NaN''s.');
end
if (nargin >= 3) & ~isempty(alpha)
if prod(size(alpha)) > 1
error('stats:cmtest2:BadAlpha',...
'Significance level ALPHA must be a scalar.');
end
if (alpha <= 0 | alpha >= 1)
error('stats:cmtest2:BadAlpha',...
'Significance level ALPHA must be between 0 and 1.');
end
else
alpha  =  0.05;
end
binEdges    =  [-inf ; sort([x1;x2]) ; inf];
binCounts1  =  histc (x1 , binEdges);
binCounts2  =  histc (x2 , binEdges);
sumCounts1  =  cumsum(binCounts1)./sum(binCounts1);
sumCounts2  =  cumsum(binCounts2)./sum(binCounts2);
sampleCDF1  =  sumCounts1(1:end-1);
sampleCDF2  =  sumCounts2(1:end-1);
N1=length(x1);
N2=length(x2);
N=N1+N2;
if verbose
fprintf('m1: %4.3f M1: %4.3f m2: %4.3f M2: %4.3f \n',min(sampleCDF1),
```

```
max(sampleCDF1),min(sampleCDF2),max(sampleCDF2));
end
CMstatistic  =  N1*N2/N^2*sum((sampleCDF1 - sampleCDF2).^2);
z=[
0.00000 0.02480 0.02878 0.03177 0.03430 0.03656 0.03865 0.04061 0.04247 0.04427
0.04601 0.04772 0.04939 0.05103 0.05265 0.05426 0.05586 0.05746 0.05904 0.06063
0.06222 0.06381 0.06541 0.06702 0.06863 ...
0.07025 0.07189 0.07354 0.07521 0.07690 0.07860 0.08032 0.08206 0.08383 0.08562
0.08744 0.08928 0.09115 0.09306 0.09499 0.09696 0.09896 0.10100 0.10308 0.10520
0.10736 0.10956 0.11182 0.11412 0.11647 ...
0.11888 0.12134 0.12387 0.12646 0.12911 0.13183 0.13463 0.13751 0.14046 0.14350
0.14663 0.14986 0.15319 0.15663 0.16018 0.16385 0.16765 0.17159 0.17568 0.17992
0.18433 0.18892 0.19371 0.19870 0.20392 ...
0.20939 0.21512 0.22114 0.22748 0.23417 0.24124 0.24874 0.25670 0.26520 0.27429
0.28406 0.29460 0.30603 0.31849 0.33217 0.34730 0.36421 0.38331 0.40520 0.43077
0.46136 0.49929 0.54885 0.61981 0.74346 ...
1.16786 ]';
Pz=[0:0.01:0.99 0.999]';
T_mean =1/6+1/6/(N);
T_var  =1/45*(N+1)/N^2 * ( 4*N1*N2*N-3*(N1^2+N2^2)-2*N1*N2 ) / (4*N1*N2);
CM_limiting_stat =  ( CMstatistic - T_mean ) / sqrt(45*T_var) + 1/6;
if CM_limiting_stat > z(end)
pValue=1;
elseif CM_limiting_stat < z(1)
pValue=0;
else
```

```
pValue = interp1(z,Pz,CM_limiting_stat,'linear');
end
H  =  alpha > 1-pValue;
if verbose
fprintf('CM_stat: %6.5f CM_lim_stat: %6.5f\n',CMstatistic,CM_limiting_stat);
fprintf('T_mean: %4.3f T_var: %4.3f \n',T_mean,T_var);
end
```

Appendix B: Histogram and c.d.f. of wavelet coefficients of fBm

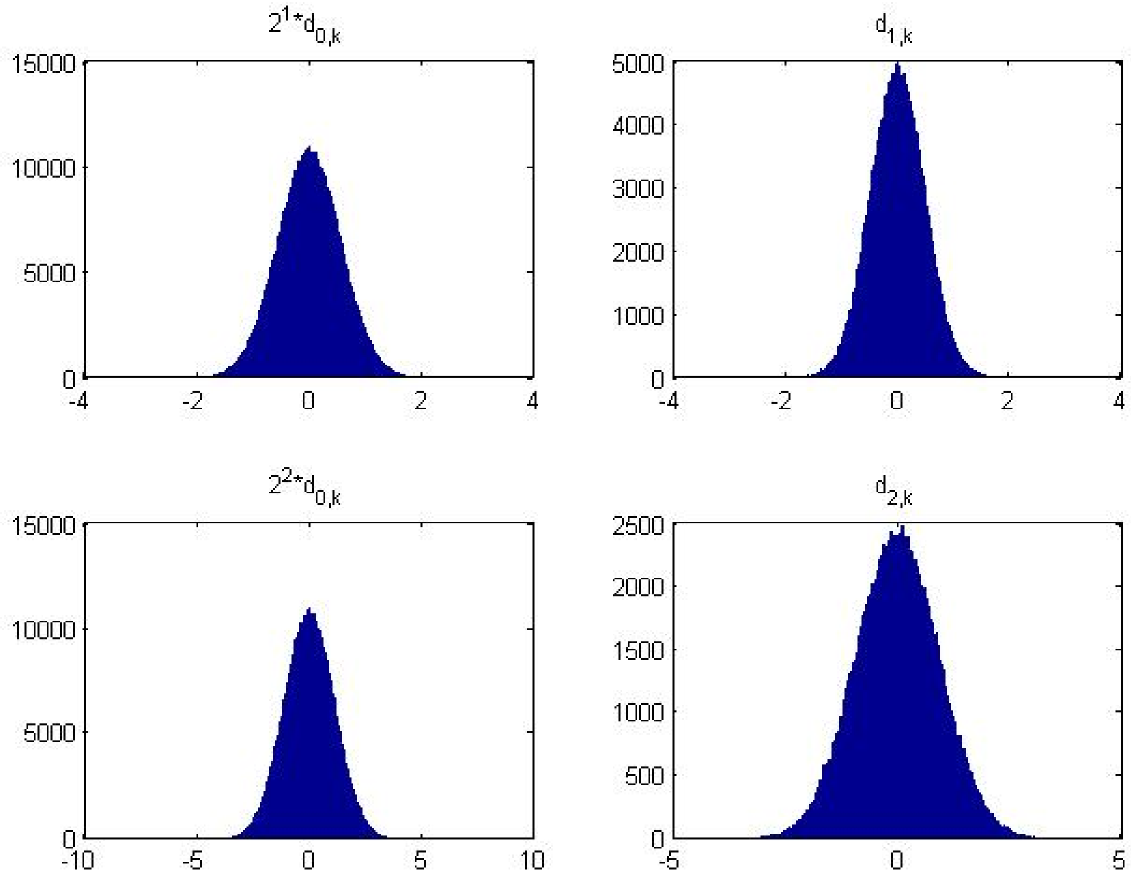


Figure B.1: Histogram of wavelet coefficients of fBm

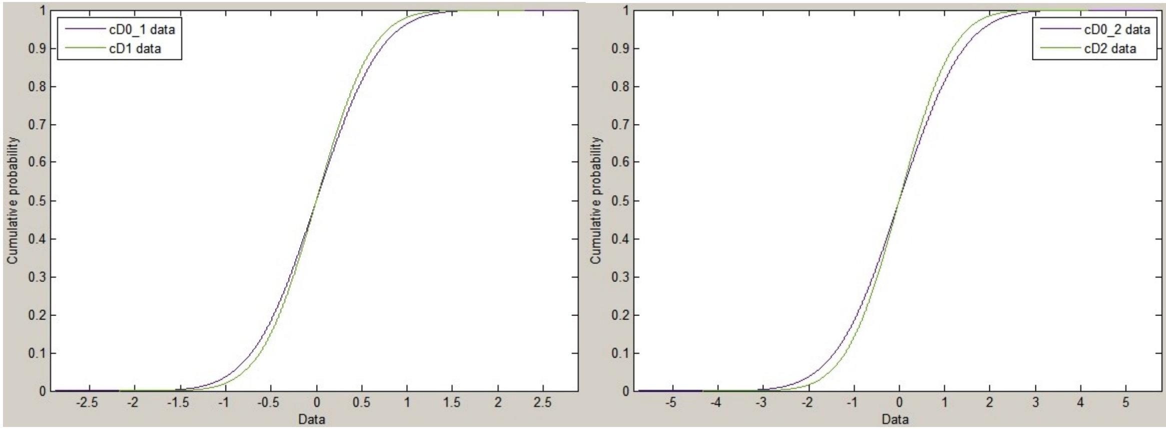


Figure B.2: C.d.f. of wavelet coefficients of fBm

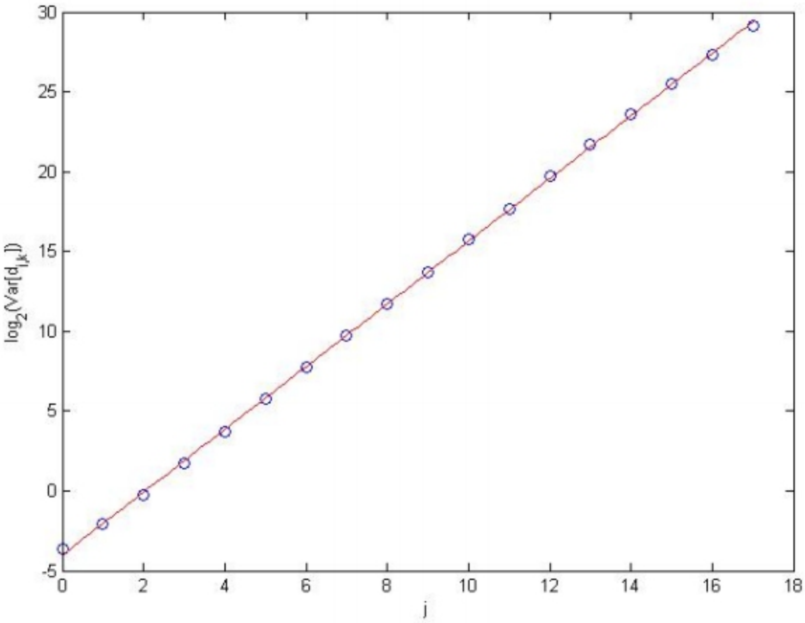


Figure B.3: Variance structure of wavelet coefficients of fBm

Appendix C: Covariance structure of wavelet coefficients of fBm

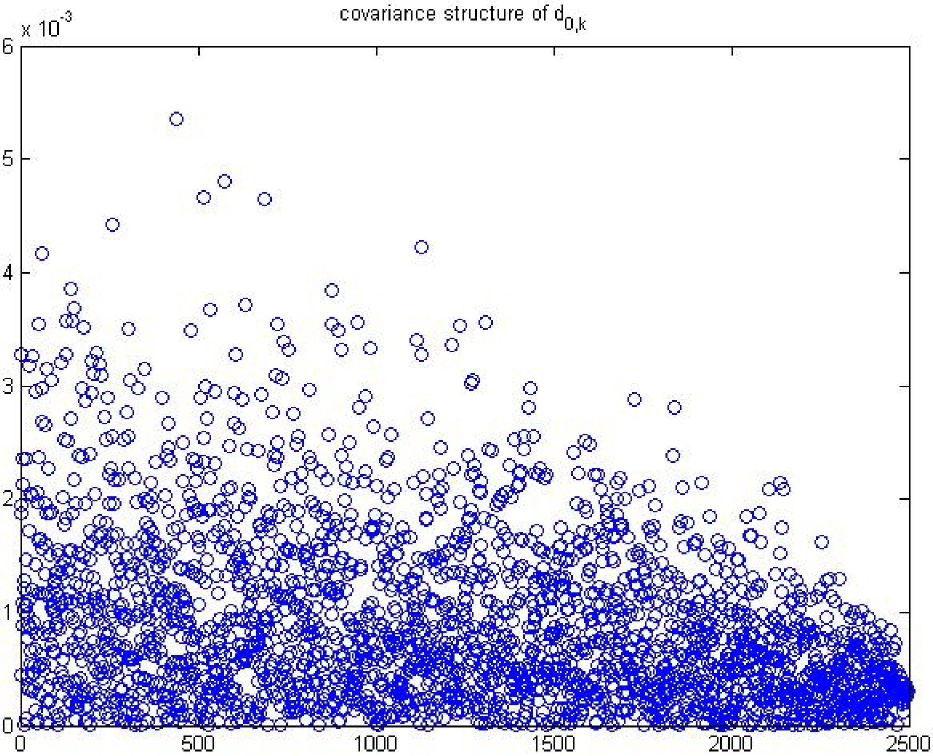


Figure C.1: Histogram of wavelet coefficients of fBm

Vita

Name: Panpan Zhang

Date of Birth: Dec. 5^{th}, 1986

City of Birth: Shanghai, China

Address: 115 Rosedale Circle, Winston Salem, NC 27106

Tel: (347)-285-8737

Email: zhanp0@wfu.edu

Educational experience:

• M.A. in Mathematics, Wake Forest University, 2012

Thesis: *Statistical self-similarity in time series from financial data & dynamical systems*

• B.S. in Mathematics, Donghua University, China, 2009

Thesis: *Initial boundary value problem of dissipative non-linear elasticity equations*

Professional experience:

• Teaching assistant, Wake Forest University, Aug. 2011 - Jan. 2012

• Tutor, Maths Center, Wake Forest University, Sept. 2010 - May 2012

Honors and Awards:

• Pi Mu Epsilon, Wake Forest University