Memory indicators and their incorporation into dynamic models

Wen Li
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Statistics and Probability Commons

Recommended Citation
Li, Wen, "Memory indicators and their incorporation into dynamic models" (2009). Graduate Theses and Dissertations. 10683. https://lib.dr.iastate.edu/etd/10683

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Memory indicators and their incorporation into dynamic models

by

Wen Li

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Alicia Carriquiry, Co-major Professor
Wolfgang Kliemann, Co-major Professor
Cindy Yu, Co-major Professor
Michael Larsen
Huaiqing Wu
Peng Liu

Iowa State University
Ames, Iowa
2009

Copyright © Wen Li, 2009. All rights reserved.
I would like to dedicate this thesis to my family without whose support I would not have been able to complete this work. I would also like to thank my supervisors and committees for their loving guidance and financial assistance during my graduate study and during the writing of this work.
TABLE OF CONTENTS

LIST OF TABLES .......................................................... vi

LIST OF FIGURES ......................................................... vii

CHAPTER 1. INTRODUCTION ........................................... 1

  1.1 Time Memory of a Process .......................................... 1
  1.2 Memory Indicators .................................................. 2
  1.3 Dynamic Models .................................................... 3

CHAPTER 2. ASYMPTOTIC BEHAVIORS ............................... 7

  2.1 Introduction of the R/S Statistic and Hurst Exponent .......... 7
  2.2 The Self-Similarity Index, Fractional Brownian Motion and Fractional Gaussian Noise .............................................. 8
  2.3 Hurst Exponent and Self-Similarity Index for Fractional Brownian Motion .......... 13
  2.4 Other Estimators for the Self-Similarity Index for Fractional Brownian Motion . 21

CHAPTER 3. ESTIMATIONS OF TIME MEMORY PARAMETERS ........ 25

  3.1 Introduction .......................................................... 25
    3.1.1 Estimation methods and their properties .................... 25
    3.1.2 Analysis of S&P500 ............................................ 26
  3.2 Comparison of Estimators of the Self-Similarity Index ........ 27
    3.2.1 Simulation and summary statistics .......................... 27
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.2 Estimators of the self-similarity index $H$</td>
<td>28</td>
</tr>
<tr>
<td>3.2.3 Results</td>
<td>35</td>
</tr>
<tr>
<td>3.3 Estimating the Self-Similarity Index $H$ in the S&amp;P500 Series</td>
<td>40</td>
</tr>
<tr>
<td>3.3.1 The effect of maximum block size on the estimate of $H$ for fractional Gaussian noise and for S&amp;P500 returns for the time period Jan. 3, 1950 - Nov. 27, 2006</td>
<td>44</td>
</tr>
<tr>
<td>3.3.2 Time varying memory structure of S&amp;P500 returns for time segments during the period Jan. 3, 1950 - Nov. 27, 2006</td>
<td>49</td>
</tr>
<tr>
<td>3.3.3 The combined effect of maximum block size and decade on the estimated value of $H$</td>
<td>52</td>
</tr>
<tr>
<td>3.3.4 Validation of the estimation of the self-similarity index</td>
<td>56</td>
</tr>
<tr>
<td><strong>CHAPTER 4. FRACTIONAL STOCHASTIC MODELS</strong></td>
<td>61</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>61</td>
</tr>
<tr>
<td>4.2 Black-Scholes Model and its Extensions</td>
<td>62</td>
</tr>
<tr>
<td>4.2.1 The Black-Scholes model</td>
<td>62</td>
</tr>
<tr>
<td>4.2.2 General stochastic volatility model</td>
<td>63</td>
</tr>
<tr>
<td>4.2.3 Log Ornstein-Uhlenbeck (LogOU) model</td>
<td>63</td>
</tr>
<tr>
<td>4.2.4 Cox-Ingersoll-Ross (CIR) model</td>
<td>64</td>
</tr>
<tr>
<td>4.3 Memory Structure and the Hurst Exponent</td>
<td>65</td>
</tr>
<tr>
<td>4.4 Stochastic Integrals and Fractional Brownian Motion</td>
<td>67</td>
</tr>
<tr>
<td>4.5 Fractional Log Ornstein-Uhlenbeck (LogOU) Model</td>
<td>68</td>
</tr>
<tr>
<td>4.5.1 The well-definedness of the fractional LogOU model system</td>
<td>68</td>
</tr>
<tr>
<td>4.5.2 The joint behavior of $Y_t^H$ and $V_t^H$</td>
<td>72</td>
</tr>
<tr>
<td>4.6 Simulation Study of Fractional LogOU</td>
<td>76</td>
</tr>
<tr>
<td>4.6.1 Simulation study of $V_t^H$</td>
<td>80</td>
</tr>
</tbody>
</table>
4.6.2 Simulation study of $Y_t^H$ increments .......................... 83

4.7 Fractional Cox-Ingersoll-Ross (CIR) Model ....................... 94
  4.7.1 The well-definedness of the fractional CIR model system .... 94
  4.7.2 The joint behavior of $Y_t^H$ and $V_t$ .......................... 97

4.8 Future Work ......................................................... 101
  4.8.1 Time-varying structure ....................................... 101
  4.8.2 Dependent $W_t^{H^v}$ and $W_t^{H^w}$ or $W_t$ .................. 103

CHAPTER 5. SUMMARY AND CONCLUSIONS ............................. 104

BIBLIOGRAPHY ............................................................... 107
TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Statistical assessment of estimators for $H$</td>
<td>37</td>
</tr>
<tr>
<td>3.2</td>
<td>Number of estimates of $H$ outside of the (0, 1) range for each nominal $H$ value and each method that had at least one out-of-range event</td>
<td>39</td>
</tr>
<tr>
<td>3.3</td>
<td>Statistical assessment of GPH with $l = 1,000$ and $g(N) = 5,000$</td>
<td>40</td>
</tr>
<tr>
<td>3.4</td>
<td>Monotonicity intervals of $\hat{H}$ for maximum block size 4.5 years</td>
<td>52</td>
</tr>
<tr>
<td>4.1</td>
<td>Estimates of the Hurst exponent for the entire series</td>
<td>81</td>
</tr>
<tr>
<td>4.2</td>
<td>Estimates of the Hurst exponent for the $Y_t^H$ incremental process</td>
<td>94</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 3.1 Period method and modified period method for a certain fGn, and

\[ H=0.3, 0.5, 0.7 \] .................................................. 36

Figure 3.2 Bias and MSE plots for 13 estimators. ........................................ 38

Figure 3.3 Estimate \( \hat{H} \) and one-standard deviation confidence bands as a function of maximum block size \( L \) and using three different estimation methods. From (a) to (e) the fGn sequences had nominal \( H \) values equal to 0.2, 0.4, 0.5, 0.6 and 0.8, respectively. ............................... 46

Figure 3.4 \( \hat{H} \) as a function of time lag. ..................................................... 47

Figure 3.5 \( \hat{H} \) over time for three different maximum block sizes. .............. 51

Figure 3.6 Mean and one-standard deviation bands (over replicates) of the maximum value of the estimate of \( H \) obtained using three estimation methods and a moving time segment. Data are simulated fGn sequences of length 10,000. In each time segment, maximum block size was allowed to vary between 50 and 250 periods, with a step of 5. Panels (a)-(e) correspond to nominal values of \( H \) equal to 0.2, 0.4, 0.5, 0.6 and 0.8. ....................................................... 54
Figure 3.7  Panel (a): Maximum value of the estimate of $H$ obtained using three estimation methods and a moving 10-year time segment. In each time segment, maximum block size was allowed to vary between 48 and 240 weeks, with a step of 5 weeks. Panel (b) shows the maximum block size in each time segment for which the maximal $\hat{H}(L)$ was obtained.  

Figure 3.8  Scrambled test for simulated fGn with $H=0.3, 0.5, 0.7$ by using REG estimator.  

Figure 3.9  Histograms of REG estimations for unscrambled and scrambled weekly data.  

Figure 4.1  Hurst exponent estimates for different window sizes for daily S&P 500 data.  

Figure 4.2  Check on the random drivers.  

Figure 4.3  ACVF for simulations.  

Figure 4.4  Block effect on the estimates of the Hurst exponent.  

Figure 4.5  The estimates of the Hurst exponent in 10 segments.  

Figure 4.6  Theoretical curves.  

Figure 4.7  ACVF on the real time interval $(0, 1)$ for $H^v = 0.2$.  

Figure 4.8  ACVF on the real time interval $(0, 1)$ for $H^v = 0.5$.  

Figure 4.9  ACVF on the real time interval $(0, 1)$ for $H^v = 0.8$.  

Figure 4.10  ACVF on the real time interval $(0, 10)$ for $H^v = 0.2$.  

Figure 4.11  ACVF on the real time interval $(0, 10)$ for $H^v = 0.5$.  

Figure 4.12  ACVF on the real time interval $(0, 10)$ for $H^v = 0.8$.  

Figure 4.13  Block effect for the $Y_t^H$ increments.
ix

Figure 4.14 Time effect for the $Y_t^H$ increments . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 93
CHAPTER 1. INTRODUCTION

1.1 Time Memory of a Process

A time series is a sequence \( Y_1, Y_2, \ldots \) of random variables with \( Y_t \) recorded at discrete times \( t \in \{0, 1, \ldots \} = N \). Time series often incorporate a memory structure, such as short term or long term memory. Intuitively, a time series displays long memory when the correlation between current and lagged observations decays slowly. More formally (e.g. Beran, 1994), let \( \{Y_t, t \in N\} \) be a (strictly) stationary process with autocorrelation function \( \rho(h) \), where \( h \) denotes the time lag. If \( \sum_{h \in N} |\rho(h)| = \infty \), then \( \{Y_t\} \) is called a long memory process; if \( \sum_{h \in N} |\rho(h)| < \infty \), then \( \{Y_t\} \) is called a short memory process; and if \( \rho(h) = 0 \) for \( h \neq 0 \), then \( \{Y_t\} \) has no memory structure and follows a random walk. More specifically, one can define stationary processes with long memory (or long-range dependence, or with slowly decaying or long-range correlations), if there exists a real number \( \alpha \in (0, 1) \) and a constant \( c_\rho > 0 \) such that

\[
\lim_{h \to \infty} \rho(h)/[c_\rho h^{-\alpha}] = 1.
\]

Another way, equivalent to the definition in the time domain, is to define long term memory using the spectral density \( f(\lambda) \) of a stationary process: If there exists a real number \( \beta \in (0, 1) \) and a constant \( c_f > 0 \) such that

\[
\lim_{\lambda \to 0} f(\lambda)/[c_f |\lambda|^{-\beta}] = 1,
\]

then, \( \{Y_t\} \) is a stationary process with long-term memory.
In the frequency domain, such time series can therefore be thought of as having power at low frequencies, so that Granger (1966) considered it “the typical spectral shape of an economic variable”. Mandelbrot and Wallis (1968) referred to this particular feature of the data as the “Joseph Effect”. Mandelbrot (1972) connected the long memory feature of data to R/S (rescaled range) analysis (Hurst, 1951), and later to the self-similarity index and the fractal dimension. Palma (2007) provides an overview of the theory and methods developed to deal with long-memory structured data. Examples of the data sets that often contain short and/or long-term memory features include many economic and financial time series, such as stock prices.

1.2 Memory Indicators

Several indicators have been proposed over the last years to describe different memory features. Besides the classic way to measure process memory by autocovariance function, the R/S statistic (analysis) that defines the Hurst exponent is also a tool to determine long-range or short-range dependence. The statistic has several desirable properties relative to more conventional methods for detecting long-range dependence (e.g. analyzing autocorrelations, variance ratios, and spectral decompositions). Mandelbrot and Wallis (1969) show using Monte Carlo simulations that the R/S statistic is able to diagnose long-range dependence in highly non-Gaussian time series. Mandelbrot (1975) reports the almost sure convergence of the R/S statistic for stochastic processes with infinite variances, while autocorrelations and variance ratios need not to be well-defined for such processes. Mandelbrot and Taqqu (1979) derive a robustness property of the R/S statistic. Finally, Mandelbrot (1972) argues that, unlike spectral analysis detecting periodic cycles, R/S analysis can detect nonperiodic cycles with periods equal to or greater than the sample period (Lo and MacKinlay, 1999). Lo and MacKinlay (1999) argue that this classical rescaled range
statistic may be sensitive to short-range dependence. They propose a modification of the standard deviation by introducing a (maximum) time lag that is chosen depending on the given data set for the resulting short-term and long-term memory asymptotics.

In 1968, based on the idea of the Hurst exponent (Hurst, 1951), Mandelbrot defined fractional Brownian motion (fBm) and its increment process fractional Gaussian noise (fGn) with self similarity index \( H \), and introduced the word “fractal” to describe a self-similar structure in time series processes. Note that \( H = \frac{1}{2} \) corresponds to the classical situation of Brownian motion. Belly and Decreusefond (1997) extended the idea to the multidimensional case, and Penttinen and Virtamo (2004) studied the two-dimensional case with pertinent simulation methods. For these processes the parameter \( H \in (0, 1) \) is constant over time. Ayache et al. (2000) extended the idea to processes for which the self-similarity index can be time-varying, resulting in multifractional Brownian motion (mBm). This class of processes does not have stationary increments, an example of which is piecewise fBm (Perrin et al., 2005). Øksendal and Zhang (2001) discussed multiparameter fBm and Biagini and Øksendal (2003) defined multivariate fBm and extended the Wick-Itô integral to this case.

### 1.3 Dynamic Models

The dynamic models in this thesis arise from the finance area. Therefore, we briefly review some finance terms here. (Hull, 2002.)

- A call option gives the holder the right to buy the underlying asset by a certain date for a certain price. A put option gives the holder the right to sell the underlying asset by a certain date for a certain price. Often, they are simply labeled as a ”call” and a ”put”, respectively.
• The European option is an option that can be exercised only at the end of its life.

• The strike price is the price at which the asset may be bought or sold in an option contract. It is also called the exercise price.

• Volatility is a measure of the uncertainty of the return realized on an asset.

• The implied volatility is the volatility implied from an option price using the Black-Scholes approach for a similar price model.

In 1973, Black and Scholes proposed the first successful options pricing approach (the Black-Scholes option pricing model), and described a general framework for pricing other financial derivative instruments. The approach they proposed was appropriate to price what is known as European put or call options on a stock. This type of options does not pay a dividend or make other disbursements. The underlying stock price is assumed to follow a geometric Brownian motion with a constant volatility.

Further, given market prices for put or call options with different strike prices but the same expiration date, the Black-Scholes model can be applied to recursively back-compute the implied volatilities. Since the Black-Scholes model assumes a constant volatility, we expect the implied volatilities to be identical. However, in the equity options market, price data for post-1987 crash equity index options show that lower strike prices for put options have higher implied volatilities. This was first noticed by Rubinstein (1994). Later, Derman (2004) argued that this phenomenon is not limited to equity options. The phenomenon is known as volatility smile, volatility smirk or volatility skew. As a consequence of these observations it was discussed that the volatilities should be modeled by a stochastic process and then combined with the original Black-Scholes model to obtain a more realistic representation of volatility (Merton, 1976; Geske, 1979; Johnson, 1979; Johnson and Shanno, 1985). Among the many possible modifications to the original Black-Scholes model that
have been proposed, the Hull-White stochastic volatility model (Hull and White, 1987),
the Cox-Ingersoll-Ross (CIR) stochastic volatility model (Cox et al., 1985a, 1985b) and Log
Ornstein-Uhlenbeck (LogOU) stochastic volatility model (Uhlenbeck and Ornstein, 1930)
are the three best-known. The CIR model and the LogOU model are used extensively in
financial applications. Later, other models, like affine jump diffusion models (Duffie et al.,
2000) and Lévy models (Li et al., 2008), were also proposed. For all of these models,
their multivariate solutions are known to follow a Markovian-type process. Whether this
type of assumed memory structure is sufficient to capture the structure underlying the
corresponding dataset is unclear.

Finally, some research addressing financial modeling using long-term memory processes
or fractional processes has been carried out in recent years. For example, in discrete time
models, a general form of the classic ARMA model was introduced by Granger and Joyeux
(1980) and Hosking (1981), called an ARFIMA model. The solution to this model, is
(approximately) a fractional processes, and allows a short or a long memory structure in
itself. In continuous time models, Hu (2002) used a LogOU model with fBm as the random
driver to describe the price behavior of a security. Cheridito et al. (2003) had a detailed
discussion on the memory structure of the fractional Ornstein-Uhlenbeck process. Mishura
(2004) discussed fractional Black-Scholes equation with stochastic volatility processes,
under the Wick-Itô definition of the stochastic integral with respect to fBm. Mendes and
Oliveira (2004) dealt with the option pricing problem with fractional volatility for some
specific form of volatility process. Øksendal (2004) talked about the arbitrage problem for
Wick-Itô integral with respect to fBm in one of his study paper. Besides, in biophysics
field, Kou and Xie (2004) used the pathwise integral with respect to fBm for $H \in [0.5, 1)$
to extend Langevin equation for the protein study.
Continuous time models have been applied extensively in many areas, e.g. finance and biophysics. Here we not only review and study the memory indicators and the estimators of these indicators, but also explore continuous time models, mainly stochastic volatility models, extend the two continuous time models – LogOU and CIR model (popular models in the finance area) into fractional forms, argue carefully about the support area of the fractional parameters that define the pathwise integral with respect to fBm, and explore the memory structures underlying these two modified models through simulation and analytical approaches.
CHAPTER 2. ASYMPTOTIC BEHAVIORS

2.1 Introduction of the R/S Statistic and Hurst Exponent

The Hurst exponent, denoted by $H$, was proposed by Hurst (1951) to determine the design of an ideal water reservoir based upon observed discharges from a lake. To recall the original definition given by Hurst, let \{ $X_k$, $k=1, \ldots, N$ \} be an observed time series and denote by $\bar{X}_n$ the average of $X_k$ over $n$ periods. For $k = 1, \ldots, n$ one computes the running sum of the accumulated deviations from the mean as

$$ Y_{k,n} = \sum_{u=1}^{k} (X_u - \bar{X}_n), $$

where $\bar{X}_n = n^{-1} \sum_u X_u$. The range over the time period $n$ is defined as

$$ R(n) = \max_{k=1,\ldots,n} (Y_{k,n}) - \min_{k=1,\ldots,n} (Y_{k,n}), $$

and the rescaled range is $R/S(n) = \frac{R(n)}{S(n)}$, where $S(n)$ is the standard deviation of $X_k$, $k=1,\ldots,n$. This leads to the Hurst exponent of the observed time series on the time interval $k = 1, \ldots, n$ as

$$ H(n) = \frac{\log \left\{ \frac{R(n)}{S(n)} \right\}}{\log(an)}, $$

(2.1)

where the constant $a$ is often set to $a = 1/2$. Note that $H(n) \in [0, 1]$.

In practice, to avoid using an arbitrary value of the unknown constant $a$, the Hurst exponent is estimated by averaging the rescaled range $R/S(n)$ over several, non-overlapping
periods of different length $n$. More precisely, one partitions the time interval $[1, N]$ into non-overlapping subintervals of length $n$ for $n = \frac{N}{2}, \frac{N}{4}, \ldots$ and regresses $\log \{ R/S(n) \}$ on $\log n$ for several values of $n$. An estimate of the slope of this linear regression is taken as $\hat{H}_{R/S}$. Hall et al. (2000) discuss the asymptotic distribution of $\hat{H}_{R/S}$: for $\frac{3}{4} < H < 1$ the asymptotic distribution of $\hat{H}_{R/S}$ is the Rosenblatt distribution (Rosenblatt, 1961), while for $0 < H \leq \frac{3}{4}$ one obtains the normal distribution.

Intuitively, the Hurst exponent measures the smoothness of a time series based on the asymptotic behavior of the rescaled range of the process. If $H = \frac{1}{2}$, the behavior of the time series will be similar to that of a random walk; if $0 < H < 0.5$, the time-series will exhibit short-term memory; if $0.5 < H < 1$, the time-series will be characterized by long memory effects.

This chapter mainly addresses almost sure convergence and convergence in the first moment of the properly scaled R/S statistics of fractional Gaussian noise. Such results have been mentioned in the literature, compare, e.g., Taqqu et al. (1995), without proof. It is organized as follows. In Section 2.2 we review some of the properties of the classical R/S statistic. The main results on the R/S statistic are discussed in Section 2.3, and Section 2.4 presents convergence results for two other estimators of $H$ that have been proposed in the literature.

### 2.2 The Self-Similarity Index, Fractional Brownian Motion and Fractional Gaussian Noise

**Definition 2.2.1** A real-valued stochastic process $Y = \{ Y_t \}_{t \in \mathbb{R}}$ is self-similar with index $H > 0$ (H-ss) if, for any $a > 0$ and any $t \in \mathbb{R}$, $Y_{at} \overset{d}{=} a^H Y_t$, where $\overset{d}{=}$ denotes equality of the distributions. Self-similarity for processes $\{ Y_t \}_{t \geq 0}$ and $\{ Y_t \}_{t > 0}$ is defined in the same way as for $Y = \{ Y_t \}_{t \in \mathbb{R}}$. 
The self-similarity index describes invariance under time and space scaling of a process and therefore the index $H$ is also called the scaling exponent. It can be used to detect memory structures in stochastic processes. Note that a H-ss process $\{Y_t\}$ with first and second moments cannot be stationary, unless it is degenerate, compare Beran (1994), Section 2.3. (A process $\{Y_t\}$ is degenerate if $Y_t \equiv 0$.) While there are many different self-similar processes, the interest in time series analysis is usually on those processes that have stationary increments. Recall that a real-valued process $\{Y_t\}_{t \in \mathbb{R}}$ is said to have (strictly) stationary increments if all finite-dimensional distributions are shift invariant, i.e. for all $h \in \mathbb{R}$ and all finite number of time points $t_1, \ldots, t_k$ it holds that $D(Y_{t_1+h} - Y_{t_1+h-1}, \ldots, Y_{t_k+h} - Y_{t_k+h-1}) = D(Y_{t_1} - Y_{t_1-1}, \ldots, Y_{t_k} - Y_{t_k-1})$, where $D(\cdot)$ denotes the distribution of a random variable.

**Definition 2.2.2** A process $\{Y_t\}_{t \in \mathbb{R}}$ is called H-sssi if it is self-similar with index $H$ and has strictly stationary increments.

Self-similar processes with stationary increments are of great interest in applications to time series analysis. For future reference, we list some properties of H-sssi processes $\{Y_t\}_{t \in \mathbb{R}}$ with finite first and second moments. Corresponding results hold for processes defined on the time set $\{t \geq 0\}$, compare, e.g., Taqqu (2003). The underlying probability space is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$.

1. $Y_0 = 0$ a.s..
   
   This follows from $Y_0 = Y_{a0} \overset{d}{=} a^HY_0$, for any $a > 0$.

2. If $H \neq 1$, then $\mathbb{E}(Y_t) = 0$, for all $t \in \mathbb{R}$.
   
   This follows from two simple observations. By self-similarity $\mathbb{E}(Y_{2t}) = 2^H\mathbb{E}(Y_t)$, strict stationarity of the increments implies that $\mathbb{E}(Y_{2t}) = \mathbb{E}(Y_{2t} - Y_t) + \mathbb{E}(Y_t) = \mathbb{E}(Y_t)$. 

2\mathbb{E}(Y_t).

3. \( Y_{-t} = Y_{-t} - Y_0 \overset{d}{=} Y_0 - Y_t = -Y_t. \)

The proof is Property 2. together with strict stationarity of the increments.

4. \( \mathbb{E}(Y^2_t) = \mathbb{E}\left\{Y^2_{\lfloor t \text{sign}(t) \rfloor}\right\} = |t|^{2H}\mathbb{E}\left\{Y^2_{\text{sign}(t)}\right\} = |t|^{2H}\mathbb{E}(Y^2_1) = : |t|^{2H}\sigma^2. \)

The result follows by Property 3. and self-similarity. If \( \sigma^2 = \mathbb{E}(Y^2_1) = 1 \), we will say that the process \( \{Y_t\}_{t \in \mathbb{R}} \) is standard.

5. The covariance function \( \gamma_Y(s, t) = \mathbb{E}\left\{\{Y_s - \mathbb{E}(Y_s)\} \{Y_t - \mathbb{E}(Y_t)\}\right\} = \mathbb{E}(Y_sY_t) \) for \( s, t \in \mathbb{R} \), is given by \( \gamma_Y(s, t) = \frac{1}{2}\left\{\mathbb{E}(Y^2_s) + \mathbb{E}(Y^2_t) - \mathbb{E}(Y_s - Y_t)^2\right\} \), which follows from Property 4. and strict stationarity of the increments. Hence it holds that \( \gamma_Y(s, t) = \frac{\sigma^2}{2}(|s|^{2H} + |t|^{2H} - |s - t|^{2H}) \). Note that for \( 0 < H \leq 1 \) the function \( \gamma_Y(s, t) \) is non-negative definite.

6. The self-similarity parameter \( H \) satisfies \( H \leq 1. \)

Since \( \mathbb{E}|Y_2| = \mathbb{E}|Y_2 - Y_1 + Y_1| \leq \mathbb{E}|Y_2 - Y_1| + \mathbb{E}|Y_1| = 2\mathbb{E}|Y_1| \) and \( \mathbb{E}|Y_2| = 2^H\mathbb{E}|Y_1| \), we obtain that \( 2^H \leq 2 \) or \( H \leq 1. \)

7. If \( H = 1 \) then we have for all \( t \in \mathbb{R} \) that \( Y_t = tY(1) \) a.s.

If \( H = 1 \), it follows from Property 5. that

\[ \mathbb{E}(Y_tY_s) = st\mathbb{E}(Y^2_1) \]

and hence

\[ \mathbb{E}(Y_t - tY_1)^2 = \mathbb{E}\left\{Y^2_t - 2t\mathbb{E}(Y_tY_1) + t^2\mathbb{E}(Y^2_1)\right\} = (t^2 - 2tt + t^2)\mathbb{E}(Y^2_1) = 0, \]

which implies the statement. Note that for processes with discrete time set \( t \in \mathbb{Z} \) this implies that the trajectories of \( \{Y_t\}_{t \in \mathbb{Z}} \) and of \( \{tY_1\}_{t \in \mathbb{Z}} \) agree a.s.
Next, we turn to some properties of the increments of an H-sssi process \( \{ Y_t \}_{t \in \mathbb{R}} \). Let \( k \in \mathbb{Z} \) and denote by \( X_k := Y_k - Y_{k-1} \) the increment process \( \{ X_k \}_{k \in \mathbb{Z}} \).

8. \( \{ X_k \}_{k \in \mathbb{Z}} \) is strictly stationary with \( \mathbb{E} (X_k) = 0 \) and \( \mathbb{E} (X_k^2) = \sigma^2 = \mathbb{E} (Y_1^2) \).

This follows directly from the definitions.

9. The autocovariance function of \( \{ X_k \}_{k \in \mathbb{Z}} \) is given by

\[
\gamma_X (h) = \mathbb{E} (X_k X_{k+h}) = \frac{\sigma^2}{2} (|h + 1|^{2H} - 2|h|^{2H} + |h - 1|^{2H}).
\]

This result follows from Property 5. above.

10. Let \( h \neq 0 \). Then we have

\[
\gamma_X (h) = \begin{cases} 
0 & \text{if } H = \frac{1}{2} \\
< 0 & \text{if } 0 < H < \frac{1}{2} \\
> 0 & \text{if } \frac{1}{2} < H < 1.
\end{cases}
\]

Note that if \( \frac{1}{2} < H < 1 \), then \( f(x) = x^{2H} \) is a strictly convex function. Hence we have for \( h \geq 1 \)

\[
\frac{(h + 1)^{2H} + (h - 1)^{2H}}{2} = f \left( \frac{h + 1 + (h - 1)}{2} \right) > f \left( \frac{h + 1 + h - 1}{2} \right) = f(h) = h^{2H},
\]

which means that \( \gamma_X (h) > 0 \). The case \( 0 < H < \frac{1}{2} \) is similar and \( H = \frac{1}{2} \) is obvious.

11. If \( H \neq \frac{1}{2} \), then \( \gamma_X (h) \sim \sigma^2 H (2H - 1) |h|^{2H-2} \) as \( h \to \infty \).

Since \( \gamma_X (h) = \gamma_X (-h) \), it is enough to consider \( h > 0 \). By Property 9. we have for \( h \geq 1 \)

\[
\gamma_X (h) = \frac{\sigma^2}{2} \left\{ (h + 1)^{2H} - h^{2H} + (h - 1)^{2H} \right\} \\
= \frac{\sigma^2}{2} \ h^{2H-2} \times h^2 \left\{ \left( 1 + \frac{1}{h} \right)^{2H} - 2 + \left( 1 - \frac{1}{h} \right)^{2H} \right\}.
\]
We have \( \lim_{h \to \infty} h^2 \left\{ \left( 1 + \frac{1}{h} \right)^{2H} - 2 + \left( 1 - \frac{1}{h} \right)^{2H} \right\} = 2H (2H - 1) \) (by L’Hôpital’s rule) and hence the result.

Note that, according to Property 9., the covariance for the stationary increment sequence of a self-similar process with index \( H \) follows a power law structure. Further, if \( 0 < H < \frac{1}{2} \), the autocovariance of \( \{X_k\}_{k \in \mathbb{Z}} \) is negative and \( \sum_{h=1}^{\infty} |\gamma_X(h)| < \infty \), so the increment sequence \( \{X_k\}_{k \in \mathbb{Z}} \) of a H-sssi process \( \{Y_t\} \) is mean-reverting and anti-persistent, i.e. it has a short-term memory structure. If \( \frac{1}{2} < H < 1 \), the covariance of \( \{X_k\}_{k \in \mathbb{Z}} \) is positive and \( \sum_{h=1}^{\infty} |\gamma_X(h)| = \infty \), hence in this case the sequence \( \{X_k\}_{k \in \mathbb{Z}} \) is positively correlated and has a long-term memory structure.

The standard example for a self-similar process with stationary increments is fractional Brownian motion: any Gaussian H-sssi process \( \{B_H(t)\}_{t \in \mathbb{R}} \) with \( 0 < H < 1 \) is called a fractional Brownian motion (fBm). An fBm process is called standard, if \( \text{Var} \{B_H(1)\} = 1 \). The (stationary) increment process \( \{X_k\}_{k \in \mathbb{Z}} := \{B_H(k) - B_H(k-1)\}_{k \in \mathbb{Z}} \) is called fractional Gaussian noise (fGn). When the self-similarity index \( H \) of the process is fixed at \( H = 0.5 \), the fractional Brownian motion and the fractional Gaussian noise become standard Brownian motion and standard Gaussian (white) noise. We refer the reader to Taqqu (2003) and the references therein for details on fractional Brownian motion.

Stationary solutions to some classes of stochastic difference equations are also H-sssi processes. This is true, in particular, for fractional ARIMA (FARIMA) models, compare e.g. Granger and Joyeux (1980), Hosking (1981) and Section 2.5 in Beran (1994), and for linear ARCH (LARCH) models, see Levine et al. (2006).
2.3 Hurst Exponent and Self-Similarity Index for Fractional Brownian Motion

In 1995, Taqqu et al. mention that for fractional Gaussian noise (or fractional ARIMA) processes, one has the following asymptotic result: \( \mathbb{E} \{ R/S (n) \} \sim C_H n^H \), as \( n \to \infty \), where \( R/S (n) \) is the rescaled range (R/S) statistic (2.1), \( H \) is the self-similarity index defined in Definition 2.2.1, and \( C_H \) is a positive, finite constant not dependent on \( n \). In this section we provide a proof for this and related results for fractional Gaussian noise.

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be the underlying probability space on which a continuous time fractional Brownian motion \( \{ B_H(t) \}_{t\geq 0} \) is defined with \( B_H(0) = 0 \) a.s. For a given \( h > 0 \), the increment process (fractional Gaussian noise) is defined as \( X_0 = 0, X_k := B_H(kh) - B_H((k-1)h) \) for \( k \in \mathbb{N} \), with the variance \( (\sigma h^H)^2 \), where \( \sigma^2 = \text{Var} \{ B_H(1) \} \). Note that \( \sum_{i=1}^{n} X_i = B_H(nh) \), and we set \( S^2 (n) = \frac{1}{n} \sum_{i=1}^{n} X_i^2 = \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 \). Letting \( Y_{k,n} = B_H(kh) - \frac{k}{n} B_H(nh) \), the R/S statistic of \( \{ B_H(kh) \}_{k \in \{1,\ldots,n\}} \) according to (2.1) reads

\[
R/S (n) = \frac{1}{S (n)} \left[ \max_{k \in \{1,\ldots,n\}} \left\{ B_H(kh) - \frac{k}{n} B_H(nh) \right\} \right. \\
\left. - \min_{k \in \{1,\ldots,n\}} \left\{ B_H(kh) - \frac{k}{n} B_H(nh) \right\} \right].
\]

(2.2)

By the definition of self-similar processes we have that

- \( B_H(h) \overset{d}{=} (nh)^H B_H \left( \frac{1}{n} \right) \), ..., \( B_H(nh) \overset{d}{=} (nh)^H B_H(1) \),

- \( \mathbb{E} \{ B_H(h) \} = (nh)^H \mathbb{E} \{ B_H \left( \frac{1}{n} \right) \} = 0 \), ..., \( \mathbb{E} \{ B_H(nh) \} = (nh)^H \mathbb{E} \{ B_H(1) \} = 0 \), and
by Property 5. in Section 2.2 for any $k, m \in \mathbb{N}$ it holds that

$$
\text{Cov} \{B_H (kh), B_H (mh)\} = \mathbb{E} \{B_H (kh) B_H (mh)\}
= \frac{\sigma^2}{2} (|kh|^{2H} + |mh|^{2H} - |kh - mh|^{2H})
= \frac{\sigma^2 (nh)^{2H}}{2} \left( \left| \frac{k}{n} \right|^{2H} + \left| \frac{m}{n} \right|^{2H} - \left| \frac{k}{n} - \frac{m}{n} \right|^{2H} \right)
= (nh)^{2H} \mathbb{E} \left\{ B_H \left( \frac{k}{n} \right) B_H \left( \frac{m}{n} \right) \right\}
= \text{Cov} \left\{ (nh)^H B_H \left( \frac{k}{n} \right), (nh)^H B_H \left( \frac{m}{n} \right) \right\}.
$$

Therefore, we obtain for fractional Brownian motion

$$
\{B_H (h), \cdots, B_H (nh)\} \overset{d}{=} (nh)^H \left\{ B_H \left( \frac{1}{n} \right), \cdots, B_H (1) \right\}.
$$

We define

$$
\tilde{S}^2 (n) = \frac{(nh)^{2H}}{n} \sum_{i=1}^{n} \left\{ B_H \left( \frac{i}{n} \right) - B_H \left( \frac{i-1}{n} \right) \right\}^2
- \left[ \frac{(nh)^{H}}{n} \sum_{i=1}^{n} \left\{ B_H \left( \frac{i}{n} \right) - B_H \left( \frac{i-1}{n} \right) \right\} \right]^2
$$

and

$$
\tilde{R}/S (n) = \frac{(nh)^H}{\tilde{S} (n)} \left[ \max_{k \in \{1, \cdots, n\}} \left\{ B_H \left( \frac{k}{n} \right) - \frac{k}{n} B_H (1) \right\} \right.
- \min_{k \in \{1, \cdots, n\}} \left\{ B_H \left( \frac{k}{n} \right) - \frac{k}{n} B_H (1) \right\},
$$

which results in

$$
\mathbb{E} \{ \tilde{R}/S (n) \} = \mathbb{E} \left\{ \tilde{R}/S (n) \right\}.
$$

The following theorem on the asymptotic behavior of the $R/S$ statistics for fractional Gaussian noise is the main result of this section.
Theorem 2.3.1 The $\tilde{R}/S(n)$ statistics of a fractional Brownian motion $\{B_H(n)\}_{n \in \mathbb{N}}$ satisfies

(i) as $n \to \infty$,

$$\frac{1}{n^H} \tilde{R}/S(n) \xrightarrow{w.p.} \frac{1}{\sigma} \left[ \max_{0 \leq s \leq 1} \{B_H(s) - sB_H(1)\} - \min_{0 \leq s \leq 1} \{B_H(s) - sB_H(1)\} \right],$$

(ii) as $n \to \infty$,

$$\frac{1}{n^H} \mathbb{E} \left\{ \tilde{R}/S(n) \right\} \rightarrow \frac{1}{\sigma} \mathbb{E} \left[ \max_{0 \leq s \leq 1} \{B_H(s) - sB_H(1)\} - \min_{0 \leq s \leq 1} \{B_H(s) - sB_H(1)\} \right],$$

where $\sigma^2 = \text{Var} \{B_H(1)\}$.

The following (deterministic) lemma is used in the proof.

Lemma 2.3.2 Let $f(x)$ be a continuous function on the interval $[0, 1]$. Then,

$$\lim_{n \to \infty} \max_{x \in \{1, \ldots, n\}} f \left( \frac{x}{n} \right) = \max_{0 \leq x \leq 1} f(x), \quad \text{and} \quad \lim_{n \to \infty} \min_{x \in \{1, \ldots, n\}} f \left( \frac{x}{n} \right) = \min_{0 \leq x \leq 1} f(x).$$

Proof. It is sufficient to prove the maximum case. Let $x_0 \in [0, 1]$ with $f_0 := f(x_0) = \max_{0 \leq x \leq 1} f(x)$. Take $\varepsilon > 0$ arbitrary, then there exists $\delta > 0$ with $f_0 - f(y) < \varepsilon$ for all $y \in [0, 1]$ with $|x_0 - y| < \delta$. Let $N_0 \in \mathbb{N}$ with $1/N_0 < \delta$, then for all $n \geq N_0$ there exists $m_n \in \mathbb{N}$ with $|x_0 - m_n/n| < \delta$ and hence $f_0 - f \left( \frac{m_n}{n} \right) < \varepsilon$. Therefore,

$$\max_{x \in \{1, \ldots, n\}} f \left( \frac{x}{n} \right) > f_0 - \varepsilon \quad \text{for all } n \geq N_0.$$

Proof. (of Theorem 2.3.1)
(i) First we show, using the ergodic theorem, that \( \frac{\tilde{S}(n)}{\sigma n^H} \to 1 \) with probability 1 as \( n \to \infty \) where \( \tilde{S}(n) \) is defined in (2.3). Recall that a sufficient condition for a stationary Gaussian time series \( \{X_k\}_{k \in \mathbb{Z}} \) to be ergodic is that \( \text{Cov}(X_k, X_{k+h}) \to 0 \), as \( h \to \infty \) (compare, e.g. Sinai, 1976, page 111). Properties 9. and 11. from Section 2.2 then show that fractional Gaussian noise is ergodic. Applying the ergodic theorem to \( X_*^i := (nh)^H \{B_H \left( \frac{i}{n} \right) - B_H \left( \frac{i-1}{n} \right) \} \) and to \( (X_*^i)^2 \), for \( i = 1, \cdots, n \), yields
\[
\frac{1}{n} \sum_{i=1}^{n} X_*^i = \frac{(nh)^H B_H (1)}{n} \overset{w.p.1}{\to} 0, \quad \frac{1}{n} \sum_{i=1}^{n} (X_*^i)^2 \overset{w.p.1}{\to} \sigma^2 h^{2H}, n \to \infty.
\]

Therefore, \( \tilde{S}^2 (n) \to \sigma^2 h^{2H} \), i.e. \( \frac{1}{\tilde{S}(n)} \to \frac{1}{\sigma n^H} \) with probability 1 as \( n \to \infty \).

This means that there exists a measurable set \( B_0 \subset \Omega \) with \( \Pr(B_0) = 1 \), such that for all \( \omega \in B_0 \) we have \( \lim_{n \to \infty} \frac{1}{\tilde{S}(n)} = \frac{1}{\sigma} \). Define \( A(s) = B_H (s) - sB_H (1) \), then the continuity of the paths of \( B_H (t) \) and Lemma 2.3.2 imply
\[
\max_{k \in \{1, \cdots, n\}} A \left( \frac{k}{n} \right) \overset{w.p.1}{\to} \max_{0 \leq s \leq 1} A (s), \text{ as } n \to \infty.
\]

This means that there exists a measurable set \( B_1 \subset \Omega \) with \( \Pr(B_1) = 1 \), such that for all \( \omega \in B_1 \) it holds that \( \lim_{n \to \infty} \max_{k \in \{1, \cdots, n\}} A \left( \frac{k}{n} \right) = \max_{0 \leq s \leq 1} A (s) \). And similarly there exists a measurable set \( B_2 \subset \Omega \) with \( \Pr(B_2) = 1 \), such that for all \( \omega \in B_2 \) we have \( \lim_{n \to \infty} \min_{k \in \{1, \cdots, n\}} A \left( \frac{k}{n} \right) = \min_{0 \leq s \leq 1} A (s) \). Note that \( \Pr(B_0 \cap B_1 \cap B_2) = 1 \), and hence we obtain
\[
\frac{\max_{k \in \{1, \cdots, n\}} A \left( \frac{k}{n} \right) - \min_{k \in \{1, \cdots, n\}} A \left( \frac{k}{n} \right)}{\tilde{S}(n)} \overset{w.p.1}{\to} \frac{\max_{0 \leq s \leq 1} A (s) - \min_{0 \leq s \leq 1} A (s)}{\sigma h^H}, \text{ as } n \to \infty.
\]

In other words,
\[
\frac{1}{n^H R/S(n)} \overset{w.p.1}{\to} \frac{\max_{0 \leq s \leq 1} A (s) - \min_{0 \leq s \leq 1} A (s)}{\sigma}, \text{ as } n \to \infty.
\]

(ii) Using Part (i), we only need to show that \( \frac{1}{n^H R/S(n)} \) is uniformly integrable, i.e., it is enough to show that \( \frac{1}{n^H R/S(n)} \mathbb{E} \left\{ R/S(n) \right\}^{1+\eta} < \infty \), for a some \( \eta > 0 \). By Hölder’s
inequality,

$$\frac{1}{nH} \mathbb{E}\left\{ \tilde{R}/\tilde{S}(n) \right\}^{1+\eta} \leq \left( \mathbb{E}\left[ 1/ \left\{ \tilde{S}^2(n) \right\} \right] \right)^{\frac{1+\eta}{2}} \times \left[ \mathbb{E}\left\{ \tilde{R}(n) \right\}^{\frac{2(1+\eta)}{1-\eta}} \right]^\frac{1-\eta}{2} < \infty.$$ 

Let $\eta = \frac{1}{3}$, and then, it is sufficient to prove, for $n$ large, that $\mathbb{E}\left[ 1/ \left\{ \tilde{S}^2(n) \right\} \right] < \infty$ and \( \mathbb{E}\left\{ \tilde{R}^4(n) \right\} < \infty \), where $\tilde{S}$ was defined in $(2.3)$ and $\tilde{R} = \max_{k \in \{1, \ldots, n\}} A \left( \frac{k}{n} \right) - \min_{k \in \{1, \ldots, n\}} A \left( \frac{k}{n} \right)$.

Let $X_{\ast \ast}^i := B_H \left( \frac{i}{n} \right) - B_H \left( \frac{i-1}{n} \right)$, for $i = 1, \ldots, n$, then $X_{\ast \ast}^{\ast \ast}(n) := (X_{\ast \ast}^1, \ldots, X_{\ast \ast}^n)^T \sim N_n \left( 0^{(n)} , \Sigma \right)$ with $\Sigma$ is positive definite, where $0^{(n)}$ denotes an $n$-vector of 0’s. By Imhoff (1961) it holds for $A = \mathbb{I}_n - \frac{1}{n} n^{(n)} 1^{(n)} T$, where $\mathbb{I}_n$ is the identity matrix that

$$Q = \frac{1}{n} X_{\ast \ast}^{\ast \ast}(n)^T AX_{\ast \ast}^{\ast \ast}(n) = \frac{1}{n} \sum_{r=1}^{m} \lambda_r \chi^2_{h_r}, \tag{2.6}$$

where the $\lambda_r$ are the distinct non-zero roots of $A\Sigma$ (which are all positive in this case, see Lemma 2.3.4 below), the $h_r$ are their respective orders of multiplicity, the $m$ is the number of these distinct non-zero roots, and then $\sum_{r=1}^{m} h_r = (n-1)$ (see Lemma 2.3.4 below). In (2.6) the $\chi^2_{h_r}$ are independent central $\chi^2$-variables with $h_r$ degrees of freedom.

The following claim is needed in the proof.

- (C.1) $\int_0^\infty P(X > y) dy = \mathbb{E} X$ for any positive variable $X$.

Defining $\lambda_{(1)} := \min \{ \lambda_1, \ldots, \lambda_m \}$, we have:

$$\mathbb{E}\left[ 1/ \left\{ \tilde{S}^2(n) \right\} \right] = \mathbb{E}\left\{ (nh)^{2H} Q \right\}^{-1} \overset{(C.1)}{=} n (nh)^{-2H} \int_0^\infty \mathbb{P} \left( \frac{1}{nQ} > y \right) dy$$

$$= n (nh)^{-2H} \int_0^\infty \mathbb{P} \left( nQ < \frac{1}{y} \right) dy = n (nh)^{-2H} \int_0^\infty \mathbb{P} \left( \sum_{r=1}^{m} \lambda_r \chi^2_{h_r} < \frac{1}{y} \right) dy$$

$$\leq n (nh)^{-2H} \int_0^\infty \mathbb{P} \left( \sum_{r=1}^{m} \lambda_{(1)} \sum_{r=1}^{m} \chi^2_{h_r} < \frac{1}{y} \right) dy = n (nh)^{-2H} \int_0^\infty \mathbb{P} \left( \frac{1}{\lambda_{(1)} \chi^2_{n-1}} > y \right) dy$$

$$= \frac{n}{(nh)^{2H} \lambda_{(1)}} \mathbb{E} \left( \frac{1}{\chi^2_{n-1}} \right) = \frac{n}{(n-1) (nh)^{2H} \lambda_{(1)}} < \infty,$$
which proves the statement for \( \mathbb{E} \left[ 1 / \left\{ \tilde{S}^2 (n) \right\} \right] \).

To see the corresponding result for \( \mathbb{E} \left\{ \tilde{R}^4 (n) \right\} \), let \( V_1 = \max_{0 \leq s \leq 1} B_H (s) / \sigma^2 \) with the cumulative probability function \( F_{V_1} (v_1) \), \( V_2 = \min_{0 \leq s \leq 1} B_H (s) / \sigma^2 \), and denote by \( F_Z (z) \) the cumulative probability function of a random variable with \( N (0, 1) \) distribution. By Adler (1990), Theorem 5.5 and Corollary 5.6, we then have \( \lim_{x \to \infty} \frac{1 - F_{V_1} (x)}{1 - F_Z (x)} = 1 \), which means that for all \( \epsilon > 0 \) there exists \( x_0 \in \mathbb{R}^+ \) such that for all \( x \geq x_0 \) it holds that \( (1 - \epsilon) \{ 1 - F_Z (x) \} < 1 - F_{V_1} (x) < (1 + \epsilon) \{ 1 - F_Z (x) \} \). Therefore

\[
0 \leq \mathbb{E} V_1 \left( V_1^4 \right) \overset{(C.1)}{=} \int_0^\infty \mathbb{P} (V_1^4 > y) \, dy = \int_0^{x_0^4} \mathbb{P} (V_1^4 > y) \, dy + \int_{x_0^4}^{\infty} \mathbb{P} (V_1^4 > y) \, dy \\
\leq x_0^4 + \int_{x_0^4}^{\infty} \mathbb{P} (V_1 > y^{\frac{1}{4}}) \, dy \overset{(C.2)}{=} \int_{x_0^4}^{\infty} \mathbb{P} (Z > y^{\frac{1}{2}}) \, dy + \int_{x_0^4}^{\infty} \mathbb{P} \left\{ \left( B_H (0) / \sigma^2 \right) < -y^{\frac{1}{2}} \right\} \, dy \\
\leq x_0^4 + (1 + \epsilon) \int_{x_0^4}^{\infty} \mathbb{P} (Z > y^{\frac{1}{2}}) \, dy + \int_{x_0^4}^{\infty} \mathbb{P} \left\{ \left( B_H (0) / \sigma^2 \right) > y^{\frac{1}{2}} \right\} \, dy \\
= x_0^4 + (2 + \epsilon) \int_{x_0^4}^{\infty} \mathbb{P} (Z > y^{\frac{1}{2}}) \, dy = x_0^4 + (2 + \epsilon) \int_{x_0^4}^{\infty} \mathbb{P} (Z^4 > y) \, dy \\
\leq x_0^4 + (2 + \epsilon) \int_0^{\infty} \mathbb{P} (Z^4 > y) \, dy \overset{(C.1)}{=} x_0^4 + (2 + \epsilon) \mathbb{E} (Z^4) < \infty.
\]

Thus, we have \( 0 \leq \mathbb{E} \left\{ \max_{0 \leq s \leq 1} B_H (s) \right\}^4 = \sigma^2 \mathbb{E} (V_1^4) < \infty \). To prove \( \mathbb{E} \left\{ \tilde{R}^4 (n) \right\} < \infty \), we need the following four claims:

- (C.2) For any \( a, b \in \mathbb{R} \), \( (a \pm b)^4 \leq 8(a^4 + b^4) \).

- (C.3) \( \max_{0 \leq s \leq 1} B_H (s) \overset{d}{=} - \min_{0 \leq s \leq 1} B_H (s) \), i.e. \( V_1 \overset{d}{=} -V_2 \).

- (C.4) \( 0 \leq \max_{0 \leq s \leq 1} \{ B_H (s) - sB_H (1) \} \leq \max_{0 \leq s \leq 1} B_H (s) + |B_H (1)| \).

- (C.5) \( 0 \leq \min_{0 \leq s \leq 1} \{ B_H (s) - sB_H (1) \} \leq - \min_{0 \leq s \leq 1} B_H (s) + |B_H (1)| \).
Following notations in (i), we have results in

\[
\mathbb{E} \left\{ \tilde{R}^4(n) \right\} = \mathbb{E} \left\{ \max_{k \in \{1, \ldots, n\}} A \left( \frac{k}{n} \right) - \min_{k \in \{1, \ldots, n\}} A \left( \frac{k}{n} \right) \right\}^4
\]

\[
\leq \mathbb{E} \left\{ \max_{0 \leq s \leq 1} A(s) - \min_{0 \leq s \leq 1} A(s) \right\}^4 \leq 8 \mathbb{E} \left\{ \max_{0 \leq s \leq 1} A(s) \right\}^4 + 8 \mathbb{E} \left\{ \min_{0 \leq s \leq 1} A(s) \right\}^4
\]

\[
= 8 \mathbb{E} \left[ \max_{0 \leq s \leq 1} \{ B_H(s) - sB_H(1) \} \right]^4 + 8 \mathbb{E} \left[ \min_{0 \leq s \leq 1} \{ B_H(s) - sB_H(1) \} \right]^4 \text{, by (C.4) & (C.5)},
\]

\[
\leq 8 \mathbb{E} \left[ \max_{0 \leq s \leq 1} B_H(s) + |B_H(1)| \right]^4 + 8 \mathbb{E} \left[ - \min_{0 \leq s \leq 1} B_H(s) + |B_H(1)| \right]^4 \text{, then by (C.2)},
\]

\[
\leq 64 \mathbb{E} \left\{ \max_{0 \leq s \leq 1} B_H(s) \right\}^4 + 64 \mathbb{E} \{ B_H(1) \}^4 + 64 \mathbb{E} \left\{ - \min_{0 \leq s \leq 1} B_H(s) \right\}^4 + 64 \mathbb{E} \{ B_H(1) \}^4
\]

\[
= 64 \mathbb{E} \left\{ \max_{0 \leq s \leq 1} B_H(s) \right\}^4 + 64 \mathbb{E} \left\{ - \min_{0 \leq s \leq 1} B_H(s) \right\}^4 + 128 \mathbb{E} \{ B_H(1) \}^4 \text{, then by (C.3)},
\]

\[
= 128 \mathbb{E} \left\{ \max_{0 \leq s \leq 1} B_H(s) \right\}^4 + 128 \mathbb{E} \{ B_H(1) \}^4 < \infty.
\]

\[\blacksquare\]

**Corollary 2.3.3** The \( R/S(n) \) statistics of a fractional Brownian motion \( \{ B_H(n) \}_{n \in \mathbb{N}} \) satisfies, as \( n \to \infty \),

\[
\frac{1}{n^H} \mathbb{E} \{ R/S(n) \} \to \frac{1}{\sigma} \mathbb{E} \left[ \max_{0 \leq s \leq 1} \{ B_H(s) - sB_H(1) \} - \min_{0 \leq s \leq 1} \{ B_H(s) - sB_H(1) \} \right].
\]

**Proof.** The proof follows from \( \mathbb{E} \{ R/S(n) \} = \mathbb{E} \{ \tilde{R}/S(n) \} \) (compare (2.5)), and the theorem above. \( \blacksquare \)

**Lemma 2.3.4** Under the assumptions of Theorem 2.3.1 the matrix \( A \Sigma \) of (2.6) has \( n \) non-negative real eigenvalues, with \( (n - 1) \) positive ones.

**Proof.** The matrix \( A \) is a (symmetric) idempotent real matrix with \( \text{rank}(A) = n - 1 \), and so there exists an orthogonal matrix \( P \), such that \( P'P = I_n \) and

\[
P'AP = D_A = \begin{pmatrix}
I_{n-1} & 0 \\
0 & 0
\end{pmatrix}.
\]
Since $P$ is an orthogonal matrix, the eigenvalues of $A\Sigma$ are same as the eigenvalues of

$$P' A\Sigma P = P' A P' P' \Sigma = D_A P' \Sigma P = \begin{pmatrix} M_{11} & M_{12} \\ 0 & 0 \end{pmatrix}, \text{ for } P' \Sigma P := \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}.$$ 

Therefore, if $\lambda$ is the eigenvalue of $D_A P' \Sigma P$, then

$$0 = \det (\lambda I_n - D_A P' \Sigma P) = \det \begin{pmatrix} \lambda I_{n-1} - M_{11} & -M_{12} \\ 0 & \lambda \end{pmatrix} = \lambda \det (\lambda I_{n-1} - M_{11}).$$

Note that $M_{11}$ is also positive definite, since it is a $n-1$ dimensional matrix on the diagonal of the positive definite matrix $P' \Sigma P$. Therefore, the eigenvalues of $D_A P' \Sigma P$ (or $A\Sigma$) are $0$ (with the multiplicity 1), or they are the eigenvalues of the matrix $M_{11}$, and therefore there are $n$ non-negative real eigenvalues with $(n-1)$ positive ones.

**Remark 2.3.5**

1. Note that

$$C_H = \frac{1}{\sigma} \mathbb{E} \left[ \max_{0 \leq s \leq 1} \{B_H(s) - sB_H(1)\} - \min_{0 \leq s \leq 1} \{B_H(s) - sB_H(1)\} \right]$$

is always positive.

2. For $H = \frac{1}{2}$, i.e. for regular Brownian motion $B(t)$, $t \in [0,1]$, the difference $B(t) - tB(1)$ is the Brownian bridge on the unit interval and $\frac{1}{\sigma} \{B(t) - tB(1)\}$ is a process with variance 1. Similarly, for $H \in (0,1)$, and $t \in [0,1]$, $B_H(t) - tB_H(1)$ may be called “fractional Brownian bridge” on the unit interval, but this is not the most natural definition despite the fact that it is “tied down”, compare Jonas (1983), Chapter 3.3 for a discussion. Our results in Theorem 2.3.1 are consistent with the long-term memory asymptotics of the R/S statistic described in Section 2.1.

3. Corollary 2.3.3 implies that for fractional Gaussian noise time series, the R/S statistic is an estimator of the $H$-ss index.
4. For a discussion of convergence in distribution (and in the weak sense) of more
general processes see Mandelbrot (1975) (Lemma 5 and Theorem 5 on pp. 276) and
Mandelbrot and Taqqu (1979, Section 3 on pp. 78-83).

2.4 Other Estimators for the Self-Similarity Index for Fractional
Brownian Motion

In this section we briefly discuss two other estimators for the self-similarity index of frac-
tional Brownian motion, the aggregated variance method (AVM) and an approach based
on the absolute values of the aggregated time series (AVA), and show their consistency.
More details of these two estimators and more estimators will be provided in Chapter 3.

Aggregated Variance Method - AVM

Consider a time series \( \{Y_k, k \geq 0\} \) with increments \( \{X_k := Y_k - Y_{k-1}, k \geq 1\} \). The AVM
approach divides the increment time series \( \{X_k, k \geq 1\} \) into blocks of size \( m \) and within each
block, computes the sample mean and variance. This procedure is repeated for different
values of \( m \) and a plot of the logarithm of the sample variance versus \( \log m \) is obtained.
The slope of the regression line is an estimator for \( 2H - 2 \). More precisely, consider the
aggregated series \( X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \ k = 1, 2, 3, \ldots \) for successive values of \( m \),
with index \( k \) labeling the block. The sample variance of \( X^{(m)}(k) \) for sample size \( N \) is
\[
\hat{\text{Var}}(X^{(m)}) = \frac{1}{N/m} \sum_{k=1}^{N/m} \left\{ X^{(m)}(k) \right\}^2 - \left\{ \frac{1}{N/m} \sum_{k=1}^{N/m} X^{(m)}(k) \right\}^2,
\]
which is an estimator of \( \text{Var}(X^{(m)}) \).

Since we have for fractional Gaussian noise with \( \beta := 2H - 2 < 0 \) that \( \text{Var}(X^{(m)}) = \sigma^2 m^\beta \), as \( m \to \infty \), the slope of the straight line \( -\log \{\hat{\text{Var}}(X^{(m)})\} \) versus \( \log m \) will
(approximately) be \( 2H - 2 \), i.e. an estimator of \( H \) is \( \hat{H}_{AVM} = \frac{1}{2}\beta + 1 \), where \( \beta \) is...
the estimated slope of the regression line. Usually values of the block size \( m \) are chosen equidistant on a log– scale, so that \( m_{i+1}/m_i = C \) for successive blocks, where \( C \) is a constant which depends on the time series. The following proposition shows that as sample size \( N \) goes to \( \infty \), the AVM estimator of \( H \) for fractional Gaussian noise is consistent.

**Proposition 2.4.1** Let \( \{X_k, k \geq 1\} \) be a fractional Gaussian noise sequence. We fix a set of block sizes \( \{m_i\}_{i=1}^M \). Denote by \( \hat{H}_{AVM}(N) \) the estimator \( \hat{H}_{AVM} \) obtained as above for the sequence \( \{X_k, 1 \leq k \leq N\} \). Then \( \hat{H}_{AVM}(N) \) converges to \( H \) with probability 1 as \( N \to \infty \).

**Proof.** Given a set of block sizes \( \{m_i\}_{i=1}^M \), define \( y_{m_i} := \log \{ \text{Var}(X^{(m_i)}) \} \) and \( \hat{y}_{m_i} := \log \{ \hat{\text{Var}}(X^{(m_i)}) \} \), and let \( \bar{y}_M \) and \( \hat{\bar{y}}_M \) be their respective sample means. Then, by the ergodic theorem and the continuous mapping theorem, it is known that, for the set of block sizes \( \{m_i\}_{i=1}^M \), we have \( |y_{m_i} - \hat{y}_{m_i}| \overset{w.p.}{\to} 0 \) and \( |\bar{y}_M - \hat{\bar{y}}_M| \overset{w.p.}{\to} 0 \) as \( N/m_i \to \infty \), for any \( i \in \{1, \cdots, M\} \), i.e., as \( N \to \infty \).

Furthermore, with the definitions

\[
\beta := \frac{\sum_{i=1}^M (y_{m_i} - \bar{y}_{N_{m_i}}) (\log m_i - \log m_i)}{\sum_{i=1}^M (\log m_i - \log m_i)^2}, \quad \text{and} \quad \hat{\beta}_M := \frac{\sum_{i=1}^M (\hat{y}_{m_i} - \hat{\bar{y}}_{N_{m_i}}) (\log m_i - \log m_i)}{\sum_{i=1}^M (\log m_i - \log m_i)^2},
\]
we obtain as \( N \to \infty \),

\[
\left| \hat{\beta}_M - \beta \right| = \left| \frac{\sum_{i=1}^{M} (\hat{y}_{m_i} - \bar{y}_M - y_{m_i} + \bar{y}_M) (\log m_i - \log m_i)}{\sum_{i=1}^{M} (\log m_i - \log m_i)^2} \right|
\]

\[
\leq \left\{ \sum_{i=1}^{M} \left( \frac{\hat{y}_{m_i} - \bar{y}_M - y_{m_i} + \bar{y}_M}{\sqrt{\sum_{i=1}^{M} (\log m_i - \log m_i)^2}} \right)^2 \right\}^{\frac{1}{2}} \leq \left\{ \sum_{i=1}^{M} \frac{(\log m_i - \log m_i)^2}{\sum_{i=1}^{M} (\log m_i - \log m_i)^2} \right\}^{\frac{1}{2}} \xrightarrow{w.p.} 0.
\]

Therefore, as \( N \to \infty \), \( \hat{H}_{AVM} = \frac{1}{2} \hat{\beta}_{Nm} + 1 \xrightarrow{w.p.} H \), for \( H = \frac{1}{2} \beta_{Nm} + 1 \). \( \blacksquare \)

Note that since \( \hat{H}_{AVM} \) converges to \( H \) with probability 1, \( \hat{H}_{AVM} \) converges (for fixed block sizes) to \( H \) in probability as \( N \to \infty \), i.e. \( \hat{H}_{AVM} \) is a consistent estimator of \( H \).

**Absolute Values of the Aggregated Series - AVA**

This method builds on the aggregated variance method, but uses the sum of the absolute values of the aggregated series, i.e. \( \frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k)| \), instead of \( \text{Var} \left( X^{(m)} \right) \). Hence the slope \( \delta \) of the logarithm of this statistic versus \( \log m \) is \( H - 1 \) and the estimator \( \hat{H}_{AVA} \) is given by \( \hat{\delta} + 1 = \hat{H}_{AVA} \).

**Proposition 2.4.2** For fractional Brownian motion the estimator \( \hat{H}_{AVA} \) converges to \( H \) with probability 1 as \( N \to \infty \).
Proof. The proof is similar to the one for $\hat{H}_{AVM}$. ■

Note that since $\hat{H}_{AVA}$ converges to $H$ with probability 1, $\hat{H}_{AVA}$ converges (for fixed block sizes) to $H$ in probability as $N \to \infty$, and therefore $\hat{H}_{AVA}$ is a consistent estimator of $H$. 
CHAPTER 3. ESTIMATIONS OF TIME MEMORY PARAMETERS

3.1 Introduction

3.1.1 Estimation methods and their properties

Simulations and empirical studies of the self-similarity index depend on the stationarity assumptions on the time series \( \{ Y_t, t \in \mathbb{Z} \} \). Estimators based on versions of the R/S statistic and on a wavelet approach do not require any stationarity assumptions, while estimators based on different spectral aspects of \( \{ Y_t \} \) (Geweke and Porter-Hudak, 1983) require weak stationarity of \( \{ Y_t \} \). A third class of estimators uses (weak) stationarity of the increment process \( \{ X_k := Y_k - Y_{k-1}, k \in \mathbb{Z} \} \), e.g. by relying on spectral properties of \( \{ X_k, k \in \mathbb{Z} \} \). This chapter analyzes systematically the statistical properties of 13 estimators for \( H \) that have been proposed in the literature, using simulations of fractional Brownian motion (with strictly stationary increment process, fractional Gaussian noise) for \( H \)-values ranging from 0.1 to 0.9. We study bias, mean squared error, and out-of-range properties. As it turns out, few of the proposed estimators have acceptable statistical properties.

Other authors have studied the properties of estimators of the Hurst exponent via simulation. Taqqu et al. (1995) simulated sequences of fractional Gaussian noise and fractional ARIMA(0,d,0) for nominal \( H \) values between 0.5 and 0.9. They applied nine different methods to these sequences in order to estimate \( H \) and computed, using Monte Carlo methods, the variance and the MSE of the estimators. They found that for most
nominal values of $H$ (or $d$), the $R/S$ estimator had the worst performance in terms of bias and while the Whittle estimator had the best. The estimator proposed by Whittle was also the one with smallest mean squared error for any value of $H$. Gorg (2007) investigated the behavior of a small set of estimators of $H$ on simulated ARIMA$(0,d,0)$ sequences. For nominal $d \in [0,0.5]$ he found that the $R/S$ estimator can exhibit significant biases and that the GPH estimator (Geweke and Porter-Hudak, 1983) while unbiased, has large mean squared error.

3.1.2 Analysis of S&P500

The memory structure of actual financial data, such as the S&P500 time series, has been analyzed using the R/S statistic (Peters, 1996, Fig 5.1 on pp. 47, pp. 75-77, pp. 83-88, and pp. 112-113), wavelet analysis (Bayraktar et al., 2004) and many other approaches. The goal of these analyses is to determine the time-varying structure of $H$ in order to find time lag intervals, in which the data show more or less memory dependence. Here we investigate two different time effects: the effect of the block size used to compute estimators of $H$ and the actual variability of $H$ over time, for a fixed block size. The results are confirmed using a scrambling test that enables empirical hypothesis testing about the true, underlying $H$.

The chapter is organized as follows. The statistical properties of 13 estimators of $H$ are analyzed in Section 3.2 using simulations of fractional Brownian motion and fractional Gaussian noise for $H$—values ranging from 0.1 to 0.9. In Section 3.3 we illustrate the implementation of some estimators for processes with time-varying $H$ by using a subset of the S&P500 series, and we summarize the financial interpretation of our findings for the S&P500 series for the time period January 1950 through November 2006. A companion chapter that is after this one, will study additional properties of the memory structure of financial time series using a model based approach that includes a stochastic market and
a volatility equation.

3.2 Comparison of Estimators of the Self-Similarity Index

A large number of estimators for the self-similarity index $H$ in time series (with time-invariant $H$) have been proposed in the literature (see e.g. Taqqu et al., 1995 and Gorg, 2007 for a partial list). In this section we analyze the statistical properties of 13 of these estimators using numerical simulations of fractional Brownian motion and fractional Gaussian noise for $H$–values ranging from 0.1 to 0.9. We study summary statistics including bias, mean squared error, and out-of-range properties of the estimators over the simulated replicate sequences.

3.2.1 Simulation and summary statistics

We generated $R = 100$ fractional Gaussian noise sequences of length $N = 10,000$ for each value of $H$, $H = 0.1, 0.2, \ldots, 0.9$, except for the case of the wavelet estimator where we used $N^* = 2^{18}$. Our approach to generating fractional Gaussian noise trajectories is based on the method introduced by Davies and Harte (1987), which relies on a fast Fourier transformation. We refer the reader to Dieker (2002, pp. 13-29) for a discussion of common exact and approximate simulation methods for fractional Brownian motion and fractional Gaussian noise.

For each estimation method described below and for each nominal value of $H$, we calculate the estimators $\hat{H}_r$ for $r = 1, \ldots, R = 100$ and their mean, the sample variance $\hat{\sigma}^2$,
average bias $\hat{b}$ and mean squared error (MSE), where

$$\hat{\sigma}^2 = \frac{1}{R-1} \left( \sum_{r=1}^{R} \hat{H}_r^2 - \frac{1}{R} \left( \sum_{r=1}^{R} \hat{H}_r \right)^2 \right),$$

$$\hat{b} = \frac{1}{R} \sum_{r=1}^{R} \hat{H}_r - H,$$

$$MSE = \frac{1}{R} \sum_{r=1}^{R} (\hat{H}_r - H)^2.$$  

We also keep track of the number of estimates outside of the parameter space (i.e. $\hat{H}_r > 1$ or $\hat{H}_r < 0$, for $1 \leq r \leq R$).

### 3.2.2 Estimators of the self-similarity index $H$

We evaluated the performance of 13 estimators of the self-similarity index $H$ of fractional Brownian motion. Each estimator (identified by a three-letter code) is described below. We order the estimation methods according to their stationarity requirements.

#### I. Estimators that do not require stationarity assumptions

**R/S statistic - RRS**

The R/S statistic is (Hurst, 1951) is a consistent estimator of the self-similarity index for fractional Brownian motion (Theorem 2.3.1) and is calculated via (2.1) by regressing $\log(R/S(n))$ on $\log n$ for several values of $n \leq N$ as explained in Section 2.1.

**Empirical R/S statistic - RSE**

The empirical R/S statistic is computed from $H(N) = \frac{\log(R/S(N))}{\log(aN)}$. In our simulations we used $a = \frac{1}{2}$.

**Higuchi’s method - HGC**

Let $Y_k = \sum_{i=1}^{k} X_i$ be fBm where $\{X_k : k = 1, ..., N\}$ is the corresponding fGn process. The estimator of $H$ is obtained as a function of the fractal dimension of the series $\{Y_k, k = \ldots\}$.  

Consider the normalized length of the curve $Y_k$, i.e. for block size $n$ we define

$$L(n) = \frac{N-1}{n^3} \sum_{j=1}^{n} \left\lfloor \frac{N-j}{n} \right\rfloor^{-1} \sum_{i=1}^{\left\lfloor (N-j)/n \right\rfloor} |Y(j + in) - Y(j + (i-1)n)|,$$

where $\lfloor \cdot \rfloor$ denotes the greatest integer function. Then, $\mathbb{E}L(n) \sim C_H n^{-D}$ for $n \to \infty$, where $D = 2 - H$ is the fractal dimension of these data. Hence the slope of a log-log plot $(L(n)$ vs. $n)$ will be $D = 2 - H$, and an estimator for $H$ is $\hat{H}_{HGC} = 2 - \hat{D}$ (Higuchi, 1988). To implement HGC we set $C = e$, and $n_i = \left\lfloor e^{i+2} \right\rfloor$ for $i = 1, \ldots, 4$, where $\lfloor \cdot \rfloor$ denotes rounding to the nearest integer.

**Estimation using wavelets - WAV**

For a stochastic process $\{Y_t\}_{t \in \mathbb{Z}}$ the wavelet detail coefficient $d_j(i), (j, i) \in \mathbb{Z}^2$ at scale $j$ and shift $i$ is given by

$$d_j(i) = 2^{-j/2} \int_{-\infty}^{+\infty} \psi(2^{-j}t - i)Y(t)dt,$$

where $\psi$ is a function satisfying the vanishing moments condition. The scale spectrum of the scale parameter $j$ is defined as

$$S_j = \frac{1}{K/2^j} \sum_{i=1}^{K/2^j} [d_j(i)]^2, \text{ for } j \leq \log_2(K),$$

where $K$ is the number of initial approximation coefficients (for $j = 0$). If $\{Y_t\}_{t \in \mathbb{Z}}$ is (discrete time) fractional Brownian motion then

$$\mathbb{E}S_j = K(H)\sigma^2 2^{(2H+1)j},$$

where $K(H) = \frac{1-2^{-2H}}{(2H+1)(2H+2)}$, and $\sigma^2$ is the variance of the fractional Gaussian noise corresponding to the fBm (Bayraktar et al., 2004). For a given series of observations we therefore divide the data into segments, average the value of $S_j$ over each segment, and perform linear regression on the log $S_i$ scale. The slope of the regression line yields an estimator $\hat{H}_{WAV}$ of the self-similarity index $H$. 
In our implementation we used Daubechies’ wavelets with $p = 2$. We set $N^* = 2^{18} = 262,144$ (instead of $N = 100,000$ that was used for all the other estimators). As segment length we used $N^*/2^j$ for $j = 0, 1, \ldots, 13$.

II. Estimators requiring stationary increments

Covariance relation method - COR

A simple estimator for the self-similarity index $H$ can be developed from one of the properties of the increments process $\{X_k\}_{k \in \mathbb{Z}}$: If $H \neq \frac{1}{2}$ then the autocovariance function satisfies $\gamma_X(h) \sim \sigma^2 H(2H - 1)|h|^{2H-2}$, as $h \to \infty$, i.e. $\gamma_X(h) \sim c|h|^{2H-2}$ (see Property 11. in Section 2.2). Hence we can estimate $\gamma_X(h)$ as $\hat{\gamma}(h) = \frac{1}{N} \sum_{i=1}^{N-|h|} (X_{i+|h|} - \overline{X})(X_i - \overline{X})$, where $X_k = Y_k - Y_{k-1}$ for a given set of observations $\{Y_k, i = 1, \ldots, N\}$. This leads to an estimator of $H$ via $\hat{\gamma}(h) \sim c|h|^{2H-2}$, as long as $h$ is large enough and $H \neq \frac{1}{2}$. By aggregating data into segments, the behavior of this estimator can be improved, as we discuss in Section 3.2.3.

In our implementation, we used $h = \lfloor \frac{4N}{7} \rfloor, \lfloor \frac{5N}{7} \rfloor, \lfloor \frac{6N}{7} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the greatest integer function.

Aggregated variance method - AVM

We divide the increment time series $\{X_k, k \geq 1\}$ of the observed data into blocks of size $n$. Consider the aggregated series $X^{(n)}(j) = \frac{1}{n} \sum_{i=(j-1)n+1}^{jn} X_i$, $j = 1, 2, 3, \ldots$, for successive values of $n$, with index $j$ labeling the block. The sample variance of $X^{(n)}(j)$, is

$$\hat{\text{Var}}X^{(n)} = \frac{1}{N/n} \sum_{j=1}^{N/n} \left( X^{(n)}(j) \right)^2 - \left( \frac{1}{N/n} \sum_{j=1}^{N/n} X^{(n)}(j) \right)^2,$$

which is an estimator of $\text{Var}X^{(n)}$. Since we have for fractional Gaussian noise with $\beta := 2H - 2 < 0$ that $\text{Var}X^{(n)} \sim \sigma^2 n^\beta$, as $n \to \infty$, the slope of the straight line $-\log(\hat{\text{Var}}X^{(n)})$ versus $\log(n)$ will be $2H - 2$, i.e. an estimator of $H$ is $\hat{H}_{AVM} = \frac{1}{2} \hat{\beta} + 1$, where $\hat{\beta}$ is the estimated slope of the regression line. Usually values of $n$ are chosen to be equidistant
on a log scale, so that $n_{i+1}/n_i = C$ for successive blocks, where $C$ is a constant which depends on the time series.

Note that since $\hat{H}_{AVM}$ converges to $H$ with probability 1 for fractional Brownian motion (compare Proposition 4.1 in Section 2.2), $\hat{H}_{AVM}$ converges to $H$ in probability, as $n \to \infty$, i.e. $\hat{H}_{AVM}$ is a consistent estimator of $H$.

In our implementation of AVM we have chosen $C = e$, and $n_i = \lceil e^{i+2} \rceil$ for $i = 1, ..., 4$, where $\lceil \cdot \rceil$ denotes rounding to the nearest integer.

**Variance differencing method - DVM**

This method builds on the aggregated variance method and is less sensitive to discontinuities of the mean and to slowly decaying trends. The estimate $\hat{H}_{DVM}$ is defined as the difference of the variances, i.e. $\sqrt{\text{Var}X^{(n_{i+1})}} - \sqrt{\text{Var}X^{(n_i)}}$. Since $\hat{H}_{DVM}$ is based on aggregated variance of data, it should also converge to $H$ with probability 1 as $n \to \infty$.

In our implementation of DVM we have chosen $C = e$, and $n_i = \lceil e^{i+2} \rceil$ for $i = 1, ..., 4$, where $\lceil \cdot \rceil$ denotes rounding to the nearest integer.

**Absolute values of the aggregated series - AVA**

This method again builds on the aggregated variance method, but uses the sum of the absolute values of the aggregated series, i.e. $\frac{1}{N/n} \sum_{j=1}^{N/n} |X^{(n)}(j)|$, instead of $\sqrt{\text{Var}X^{(n)}}$. Hence the slope $\delta$ of the logarithm of this statistic versus $\log(n)$ is $H - 1$ and the estimator $\hat{H}_{AVA}$ is given by $\hat{\delta} + 1 = \hat{H}_{AVA}$.

Note that for fractional Brownian motion $\hat{H}_{AVA}$ converges to $H$ with probability 1 as $n \to \infty$, (compare Proposition 4.2 in Section 2.2) and hence $\hat{H}_{AVA}$ converges to $H$ in probability, as $n \to \infty$, i.e. $\hat{H}_{AVA}$ is a consistent estimator of $H$.

In our implementation of AVA we have chosen $C = e$, and $n_i = \lceil e^{i+2} \rceil$ for $i = 1, ..., 4$, where $\lceil \cdot \rceil$ denotes rounding to the nearest integer.

**Residuals of regression method - REG**
For a given time series \( \{Y_k, k = 1, \ldots, N\} \) with increment process \( \{X_k, k = 1, \ldots, N\} \) we choose blocks of size \( n \). Within the \( j \)th block we compute the partial sum \( Y^{(j)}(i) \) of the increment process, i.e.

\[
Y^{(j)}_i = \sum_{u=1}^{i} X_{(j-1)n+u} = Y_{(j-1)n+i} - Y_{(j-1)n}.
\]

For each \( j \) we regress \( Y^{(j)}(i) \) on its index \( i \), and compute the sample variance of the residuals. We repeat this procedure for each block \( j \), and average the sample variances. Then the expectation of this averaged sample variance is proportional to \( n^{2H} \) for fractional Brownian motion \( \{Y_k\} \) as \( n \to \infty \) (Taqqu et al., 1995).

In our implementation of REG we have chosen values of the block size \( n \) to be equidistant on a log-scale, so that \( n_{i+1}/n_i = C \) for successive blocks. Specifically we use \( C = e \), and \( m_i = \lfloor e^{i+2} \rfloor \) for \( i = 1, \ldots, 4 \), where \( \lfloor \cdot \rfloor \) denotes rounding to the nearest integer.

**Periodogram method - PER**

This and the following two methods are based on the periodogram of the (stationary) incremental time series \( X_k = Y_k - Y_{k-1} \) for \( k = 1, \ldots, N \). If \( \{X_k, k = 1, \ldots, N\} \) is the incremental time series, then

\[
I(\lambda_{u,N}) := \frac{1}{2\pi N} \left| \sum_{j=1}^{N} X_j e^{j\lambda_{u,N}} \right|^2
\]

for \( \lambda_{u,N} = 2\pi u/N \) and (integer) \( u \in [-N/2, N/2] \) is an estimator of the spectral density of \( X_k \). Here, \( \lambda \) denotes frequency and \( i \) denotes the complex unit. For fractional Gaussian noise we have, close to the origin, that \( I(\lambda_{u,N}) \) is proportional to \( |\lambda_{u,N}|^{1-2H} \). Hence a regression of the logarithm of the periodogram on the logarithm of the frequency \( \lambda \) should result in a coefficient of \( 1 - 2H \) for the slope \( \beta \) of the regression line. An estimator of \( H \) is therefore \( \hat{H}_{PER} = \frac{1}{2}(1 - \hat{\beta}) \). In our implementation we consider the lowest 10% of the \( N/2 = 5,000 \) frequencies.
Modified periodogram method - MPR

In the periodogram method, the log – log data often have most of the low frequencies in the range close to −1, exerting a strong influence on the least-squares fitted line. Thus, in the modified periodogram method, the frequency axis is divided into logarithmically equally spaced boxes, and the periodogram values corresponding to the frequencies inside each box are averaged. In practice, several values for very low frequencies may be left unchanged, since there are often few of them to begin with. Taqqu et al. (1995) use a robustified least-squares approach (least-trimmed squares regression) to deal with very scattered modified periodograms.

For the modified periodogram method, 1% of the data at the beginning were left unchanged, the rest were divided into 60 boxes, and the first 80% of the resulting points were used to fit the data. From Figure 3.1 in Section 3.2.3 below we see that for the periodogram method many frequencies indeed fall on the far right part of the log – log plot, and for the modified periodogram method the situation is improved.

Whittle estimator - WHI

The method proposed by Whittle (1951, Chapter 4) is also based on the periodogram. For the incremental process \( \{X_k\} \), \( k = 1,...N \) of an observed time series define

\[
Q(\eta) = \int_{-\pi}^{\pi} \frac{I(\lambda_{u,N})}{f_X(\lambda_{u,N};\eta)} d\lambda,
\]

where \( I(\lambda_{u,N}) \) is the periodogram, \( f_X(\lambda_{u,N};\eta) \) is the spectral density at a frequency \( \lambda_{u,N} \), \( \lambda_{u,N} \) is as defined above and \( \eta \) is the vector of unknown parameters for the time series, i.e. \( \eta = H \) in case of fractional Brownian motion. The Whittle estimator \( \hat{H}_{WHI} \) is the value of \( \eta \) which minimizes the function \( Q \). The asymptotic behavior of the Whittle estimator was discussed by Beran (1994, Section 5.5) and by Fox and Taqqu (1986).

III. Estimators requiring (weak) stationarity
As an example of a model based estimator we include the method proposed by Geweke and Porter-Hudak (1983) in our study. Let \( \{v_k\}_{k \in \mathbb{Z}} \) be a linear stationary process (i.e. the solution of a Gaussian ARMA model) with spectral density function \( f_v(\lambda_{u,N}) \) which is assumed to be bounded, bounded away from zero, and continuous on the interval \([-\pi, \pi]\). Here, \( \lambda_{u,N} \) is frequency as defined earlier. Then the spectral density function of a process \( \{X_k\}_{k \in \mathbb{Z}} \) with representation \((1 - B)^d X_k = v_k\) is \( f_X(\lambda_{u,N}) = \frac{\sigma^2}{2\pi} [4 \sin^2 \frac{\lambda_{u,N}}{2}]^{-d} f_v(\lambda_{u,N}), \) where \( d \in (-\frac{1}{2}, \frac{1}{2}) \) and \( B \) is the backshift operator. Geweke and Porter-Hudak (1983) show that \( \{X_k\}_{k \in \mathbb{Z}} \) has the representation \((1 - B)^d X_k = v_k\) iff \( \{X_k\} \) is fractional Gaussian noise with parameter \( H = d + \frac{1}{2} \) (see e.g. Beran, 1994 Section 2.5 for a detailed discussion).

For an observed (stationary) time series \( \{X_k\}, k = 1, ..., N \) let \( \lambda_{u,N} \) be as defined earlier and denote by \( I(\lambda_{u,N}) \) the periodogram at these coordinates. Taking logarithms results in

\[
\log [I(\lambda_{u,N})] = \log \left[ \frac{\sigma^2 f_v(0)}{2\pi} \right] - d \log \left[ 4 \sin^2 \left( \frac{\lambda_{u,N}}{2} \right) \right] + \log \left[ \frac{f_v(\lambda_{u,N})}{f_v(0)} \right] + \log \left[ \frac{I(\lambda_{u,N})}{I(\lambda_{u,n})} \right].
\]

In this expression \( \log \left[ \frac{I(\lambda_{u,N})}{I(\lambda_{u,n})} \right] \) is negligible as attention is focused on harmonic frequencies close to zero. For \( u = l, ..., g(N) \), we can estimate \( 2d \) by regression, resulting in the estimator \( \hat{H}_{GPH} = \hat{d} + \frac{1}{2} \). Geweke and Porter-Hudak (1983) show that this estimator is consistent for \( d < 0 \), and they conjecture that this result also holds for \( d \in (-\frac{1}{2}, \frac{1}{2}) \).

For actual applications Geweke and Porter-Hudak propose \( l = 1 \) and \( g(N) = \sqrt{N} \). Robinson (1995) suggested parameter values with the properties \( l \to \infty \) and \( g(N) = m \), where \( \frac{m}{N} \to 0 \) and \( \frac{m}{t} \to \infty \). Hurvich et al. (1998) and Moulines and Soulier (1999) advocate \( l = 1 \) and \( g(N) = m \), where \( \frac{m \log m}{N} \to 0 \). In our implementation for fractional Gaussian noise we used the original Geweke and Porter-Hudak proposal with \( l = 1 \) and \( g(N) = \sqrt{N} = 100. \)
3.2.3 Results

Before presenting our results on the statistical behavior of the estimators listed in Section 3.2.2, we comment briefly on the issues mentioned in the introduction of the periodogram (PER) and the modified periodogram (MPR) methods. Figure 3.1 shows the log−log plots for both estimators. As expected, most data points for the periodogram estimator fall in the range $[-3, -1]$ for each of the simulated $H$−values 0.3, 0.5, and 0.7. The modified periodogram method shows a more uniform distribution of the data points, resulting in an estimator that weighs the different frequencies more evenly.

Tables 3.1 and 3.2, and Figure 3.2 present the statistical evaluation of the different estimators via simulated fractional Brownian motion with $H \in [0.1, 0.9]$. For each estimation method described in Section 3.2.2 above and for each nominal value of $H$, we calculate the estimators $\hat{H}_r$ for $r = 1, \ldots, R = 100$ replicates. Table 3.1 lists the means of $\hat{H}_r$, the standard deviations $\hat{\sigma}$, and the root mean square error $\sqrt{\text{MSE}}$ calculated over the 100 replications. Recall that the covariance relation estimator (COR) is only defined for $H \neq \frac{1}{2}$.

Bias of the estimators can be inferred from $\text{Mean}(\hat{H}_r) − \text{nominal } H$. The (simulated) bias values are shown in Figure 3.2, left panel. From top to bottom, the left panel of Figure 3.2 includes estimators with lower, moderate, and higher bias. The right panel of Figure 3.2 shows the MSE for the three groups of estimators.

Two groups of estimators show the lowest bias: The “sophisticated” block-based estimators AVA and REG, and the estimators PER, MPR and WHI that are based directly on the periodogram. The R/S-based estimators RRS and RSE, the covariance relation estimator COR, and the Geweke / Porter-Hudak estimator show the largest bias. For these five estimators the bias results depend strongly on $H$, which is also the case for the moderate
Figure 3.1  Period method and modified period method for a certain fGn, and $H=0.3$, 0.5, 0.7.
### Table 3.1 Statistical assessment of estimators for $H$

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>Nominal H</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>RRS Mean ($H_r$)</td>
<td>0.228</td>
<td>0.309</td>
<td>0.389</td>
<td>0.468</td>
<td>0.55</td>
<td>0.629</td>
<td>0.708</td>
<td>0.782</td>
<td>0.842</td>
<td></td>
</tr>
<tr>
<td>RSE Mean ($H_r$)</td>
<td>0.279</td>
<td>0.347</td>
<td>0.417</td>
<td>0.49</td>
<td>0.56</td>
<td>0.638</td>
<td>0.705</td>
<td>0.776</td>
<td>0.831</td>
<td></td>
</tr>
<tr>
<td>HGC Mean ($H_r$)</td>
<td>0.088</td>
<td>0.187</td>
<td>0.288</td>
<td>0.388</td>
<td>0.491</td>
<td>0.588</td>
<td>0.684</td>
<td>0.79</td>
<td>0.877</td>
<td></td>
</tr>
<tr>
<td>WAV Mean ($H_r$)</td>
<td>0.037</td>
<td>0.161</td>
<td>0.273</td>
<td>0.381</td>
<td>0.486</td>
<td>0.591</td>
<td>0.702</td>
<td>0.813</td>
<td>0.933</td>
<td></td>
</tr>
<tr>
<td>COR Mean ($H_r$)</td>
<td>0.359</td>
<td>0.341</td>
<td>0.335</td>
<td>0.335</td>
<td>-</td>
<td>0.348</td>
<td>0.355</td>
<td>0.389</td>
<td>0.435</td>
<td></td>
</tr>
<tr>
<td>AVM Mean ($H_r$)</td>
<td>0.089</td>
<td>0.183</td>
<td>0.292</td>
<td>0.386</td>
<td>0.488</td>
<td>0.585</td>
<td>0.668</td>
<td>0.752</td>
<td>0.819</td>
<td></td>
</tr>
<tr>
<td>DVM Mean ($H_r$)</td>
<td>0.107</td>
<td>0.193</td>
<td>0.292</td>
<td>0.387</td>
<td>0.488</td>
<td>0.584</td>
<td>0.694</td>
<td>0.8</td>
<td>0.886</td>
<td></td>
</tr>
<tr>
<td>AVA Mean ($H_r$)</td>
<td>0.092</td>
<td>0.088</td>
<td>0.097</td>
<td>0.086</td>
<td>-</td>
<td>0.113</td>
<td>0.195</td>
<td>0.27</td>
<td>0.353</td>
<td></td>
</tr>
<tr>
<td>REG Mean ($H_r$)</td>
<td>0.109</td>
<td>0.260</td>
<td>0.303</td>
<td>0.403</td>
<td>0.502</td>
<td>0.599</td>
<td>0.7</td>
<td>0.8</td>
<td>0.988</td>
<td></td>
</tr>
<tr>
<td>PER Mean ($H_r$)</td>
<td>0.069</td>
<td>0.192</td>
<td>0.299</td>
<td>0.4</td>
<td>0.505</td>
<td>0.602</td>
<td>0.698</td>
<td>0.801</td>
<td>0.901</td>
<td></td>
</tr>
<tr>
<td>MPR Mean ($H_r$)</td>
<td>0.111</td>
<td>0.202</td>
<td>0.298</td>
<td>0.401</td>
<td>0.5</td>
<td>0.601</td>
<td>0.702</td>
<td>0.8</td>
<td>0.898</td>
<td></td>
</tr>
<tr>
<td>WHI Mean ($H_r$)</td>
<td>0.076</td>
<td>0.193</td>
<td>0.299</td>
<td>0.4</td>
<td>0.499</td>
<td>0.6</td>
<td>0.701</td>
<td>0.799</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>GPH Mean ($H_r$)</td>
<td>0.305</td>
<td>0.352</td>
<td>0.397</td>
<td>0.449</td>
<td>0.5</td>
<td>0.548</td>
<td>0.6</td>
<td>0.65</td>
<td>0.701</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.016</td>
<td>0.029</td>
<td>0.041</td>
<td>0.038</td>
<td>0.039</td>
<td>0.046</td>
<td>0.034</td>
<td>0.042</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.02</td>
<td>0.032</td>
<td>0.043</td>
<td>0.039</td>
<td>0.044</td>
<td>0.051</td>
<td>0.063</td>
<td>0.055</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.008</td>
<td>0.009</td>
<td>0.008</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.012</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.063</td>
<td>0.028</td>
<td>0.021</td>
<td>0.018</td>
<td>0.015</td>
<td>0.012</td>
<td>0.019</td>
<td>0.041</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.092</td>
<td>0.088</td>
<td>0.097</td>
<td>0.086</td>
<td>-</td>
<td>0.113</td>
<td>0.195</td>
<td>0.27</td>
<td>0.353</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.275</td>
<td>0.166</td>
<td>0.103</td>
<td>0.107</td>
<td>-</td>
<td>0.276</td>
<td>0.396</td>
<td>0.474</td>
<td>0.527</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.03</td>
<td>0.034</td>
<td>0.054</td>
<td>0.036</td>
<td>0.049</td>
<td>0.045</td>
<td>0.047</td>
<td>0.044</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.054</td>
<td>0.057</td>
<td>0.057</td>
<td>0.057</td>
<td>0.05</td>
<td>0.047</td>
<td>0.056</td>
<td>0.065</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.052</td>
<td>0.059</td>
<td>0.062</td>
<td>0.062</td>
<td>0.071</td>
<td>0.068</td>
<td>0.055</td>
<td>0.063</td>
<td>0.056</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.053</td>
<td>0.059</td>
<td>0.062</td>
<td>0.063</td>
<td>0.071</td>
<td>0.07</td>
<td>0.035</td>
<td>0.063</td>
<td>0.058</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.056</td>
<td>0.058</td>
<td>0.057</td>
<td>0.057</td>
<td>0.051</td>
<td>0.046</td>
<td>0.054</td>
<td>0.065</td>
<td>0.055</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.056</td>
<td>0.057</td>
<td>0.057</td>
<td>0.056</td>
<td>0.051</td>
<td>0.046</td>
<td>0.054</td>
<td>0.065</td>
<td>0.055</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.006</td>
<td>0.011</td>
<td>0.013</td>
<td>0.019</td>
<td>0.021</td>
<td>0.024</td>
<td>0.021</td>
<td>0.03</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.011</td>
<td>0.012</td>
<td>0.013</td>
<td>0.019</td>
<td>0.021</td>
<td>0.024</td>
<td>0.021</td>
<td>0.03</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.03</td>
<td>0.031</td>
<td>0.027</td>
<td>0.033</td>
<td>0.033</td>
<td>0.029</td>
<td>0.028</td>
<td>0.028</td>
<td>0.031</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.043</td>
<td>0.032</td>
<td>0.027</td>
<td>0.033</td>
<td>0.033</td>
<td>0.029</td>
<td>0.028</td>
<td>0.028</td>
<td>0.031</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.051</td>
<td>0.057</td>
<td>0.059</td>
<td>0.056</td>
<td>0.061</td>
<td>0.058</td>
<td>0.055</td>
<td>0.054</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.051</td>
<td>0.057</td>
<td>0.059</td>
<td>0.056</td>
<td>0.06</td>
<td>0.058</td>
<td>0.055</td>
<td>0.054</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.025</td>
<td>0.009</td>
<td>0.005</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.007</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.025</td>
<td>0.009</td>
<td>0.005</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.007</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.035</td>
<td>0.037</td>
<td>0.044</td>
<td>0.036</td>
<td>0.035</td>
<td>0.034</td>
<td>0.034</td>
<td>0.034</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>M SE σ</td>
<td>0.208</td>
<td>0.157</td>
<td>0.103</td>
<td>0.061</td>
<td>0.035</td>
<td>0.062</td>
<td>0.106</td>
<td>0.154</td>
<td>0.202</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.2  Bias and MSE plots for 13 estimators.
Table 3.2 Number of estimates of $H$ outside of the $(0, 1)$ range for each nominal $H$ value and each method that had at least one out-of-range event

<table>
<thead>
<tr>
<th>Nominal H</th>
<th>WAV</th>
<th>COR</th>
<th>AVM</th>
<th>DVM</th>
<th>AVA</th>
<th>PER</th>
<th>MPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H=0.1$</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>$H=0.2$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$H=0.7$</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$H=0.8$</td>
<td>0</td>
<td>44</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$H=0.9$</td>
<td>1</td>
<td>64</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>115</td>
<td>5</td>
<td>11</td>
<td>4</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

bias estimators AVM, DVM, HGC and WAV. Because of their uneven bias for different $H$-values, one has to interpret carefully any results that are obtained for time series with time-varying $H$ using these estimators, as we discuss in Section 3.3. In the group of low bias estimators, the residuals of regression method (REG) and the periodogram method (PER) show consistently low mean square error. So does the Whittle estimator (WHI). We also note that the R/S-based estimators RRS and RSE exhibit larger MSE that is also strongly $H$-dependent. The wavelet estimator (WAV) shows moderate bias and moderate MSE, both of them depend strongly on the value of $H$. This estimator performs well for $H \in [0.5, 0.8]$, but part of the apparent good performance may be due to the larger length of the simulated data series used for WAV.

Table 3.2 shows the number of trajectories (out of $R = 100$) that resulted in an estimate of $H$ outside of the parameter space $(0, 1)$. We list only those estimators and $H$-values that actually resulted in out-of-range values. Of the five low bias estimators, AVA, PER, and MPR show some out-of-range events for $H = 0.1$ and/or $H = 0.9$. This leaves the residuals of regression method (REG) as the low bias, low MSE estimator with no out-of-range events. This estimator will be used in Section 3.3 to analyze the S&P500 data series.
Table 3.3  Statistical assessment of GPH with \( l = 1,000 \) and \( g(N) = 5,000 \)

<table>
<thead>
<tr>
<th>Estimator Method</th>
<th>Nominal H</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPH Mean ( \hat{H}(m) )</td>
<td>0.125</td>
<td>0.23</td>
<td>0.328</td>
<td>0.416</td>
<td>0.499</td>
<td>0.579</td>
<td>0.661</td>
<td>0.736</td>
<td>0.813</td>
<td></td>
</tr>
<tr>
<td>( \hat{\sigma} )</td>
<td>0.015</td>
<td>0.014</td>
<td>0.016</td>
<td>0.017</td>
<td>0.016</td>
<td>0.017</td>
<td>0.016</td>
<td>0.016</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>( \sqrt{MSE} )</td>
<td>0.029</td>
<td>0.033</td>
<td>0.032</td>
<td>0.023</td>
<td>0.016</td>
<td>0.027</td>
<td>0.042</td>
<td>0.066</td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td>ERR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Several of the estimation methods discussed above require the choice of several parameters, and these choices may greatly influence the results. Consider, e.g., the Geweke / Porter-Hudak estimator (GPH). If one changes settings for GPH, from \( l = 1 \) and \( g(N) = m = \sqrt{N} = 100 \) to \( l = N^{0.75} = 1000 \) and \( g(N) = m = N^{0.92475} \approx 5000 \) with \( \frac{m}{N} \rightarrow 0 \) and \( \frac{n}{\tau} \rightarrow \infty \), then the results are improved as can be seen in Table 3.3.

With these parameter settings, the GPH estimator now shows moderate bias and greatly improved MSE that is still high for larger \( H \)-values. Similar improvements for specific data series may be obtained for estimators like RRS, RSE, and COR, but the non-robustness of these estimators makes them, in general, unsuitable for actual observed (self-similar) data series, specifically if one expects the data to exhibit a time-varying \( H \)-structure.

### 3.3 Estimating the Self-Similarity Index \( H \) in the S&P500 Series

In Section 3.2 we evaluated the performance of several estimators of the self-similarity index \( H \) in time series for which \( H \) is constant over time. In observed data one often suspects a memory structure that varies in time, i.e. different time lags of the series display different short or long-term memory. This is true, in particular, in financial time series for which one expects short-term momentum effects and long-term averaging effects. As explained in Section 2.2, the self-similarity index \( H \) can be used to measure such memory effects.
The Standard and Poor 500 (S&P500) series has been analyzed for its memory structure, notably by Peters (1996, for example, pages 47, 77, 83, 88, 112 and 113 as well as Figures 7.6, 7.7, 8.1, 8.2, 8.3, 9.1 and 9.2) and by Bayraktar et al. (2004, Section 3, pp. 16-20). Peters bases his analysis on the R/S statistic and Bayraktar et al. on a wavelet analysis. According to Section 3.2, the R/S method shows substantial bias (for fractional Brownian motion) that also depends on the given value of $H$ and may therefore mask the actual memory structure in the S&P500 series.

In this section, we explore two issues that may impact inferences about the memory structure in the S&P500 series:

- The effect of maximum block size used in the estimation of $H$ when implementing methods for estimating $H$ that require blocking methods, and
- The effect of time on $H$, that is, whether the memory structure in the S&P500 changes over time.

The first of these two issues refers to a potential nuisance effect creeping into the estimation process. Is it possible that different choices of maximum block size might lead to different inferences about memory? The issue in the second bullet may be of greater interest: If financial markets have different memory structures when looked at over different periods then it might be possible to identify economic or other factors that are associated with those structural memory changes.

At the end of this discussion in Section 3.3.4 we also study how to empirically test hypotheses about the value of the underlying true $H$ in a given series or in a subset of a given series. To do so we revisit the scrambling method discussed by Peters (1996) that permits testing whether estimates of $H$ that differ from the null value of 0.5, can be
attributed to memory structure in the series (or in the subset of the series) or are due to noise in the estimate resulting from, for example, inadequate data.

In exploring the different effects of time on the estimation of $H$ we first consider simulated fGn sequences and then we analyze the S&P500 data series so that we can compare results. In both cases, we implement the residuals of regression method (REG) and the method due to Whittle (WHI) that showed low bias and mean square error in our simulations in Section 3.2. For comparison with the literature we also show results for the R/S statistic. We now introduce notation that will be used in the remainder of this section.

Denote by $\{P_k, k = 1, ..., N\}$ the observed weekly closing values of the S&P500 index series between Jan. 3, 1950 and Nov. 27, 2006. To account for the compounding effect of financial data, we consider the time series $\{Q_k := \log(\frac{P_k}{P_1}), k = 1, ..., N\}$. But the series $\{Q_k\}$ cannot be self-similar unless we expect $E(P_k) = E(P_1)$ for all $k$. It is therefore typical to focus on the log of returns

$$S_k := \log\left(\frac{P_k}{P_{k-1}}\right)$$

for $k = 2, ..., N$ to assess memory structures based on self-similarity within the S&P500 time series, i.e. we assume that $\{S_k, k = 1, ..., N\}$ follows a self-similar process with stationary increments, which we denote $\{X_k, k = 1, ..., N\}$.

We can explore the S&P500 series at different levels of resolution. These include (but are not limited to) daily, weekly and monthly observations, which correspond to the value of the index at closing of each time period. Regardless of the data we use for analysis, we can obtain estimates of $H$ using the different methods described in the previous section. Here, we focus on the estimators RSS, RSE and REG, all of which require that we divide the series into a sequence of non-overlapping blocks as described below, and on the Whittle estimator (WHI), which does not require blocking. We describe the approaches using the example of weekly data.
Let \( \{X_k, k = 1, \ldots, N\} \) be a given time series (e.g., a series of increments) with values in \( \mathbb{R} \), and denote by \( 0 < L_1 < L_2 < \ldots < L_U \) a sequence of maximum block sizes where \( L_u = L_1 + (u - 1) \cdot \text{step} \). For weekly data, we choose \( L_1 = 48, \text{step} = 4 \) weeks and consequently, \( L_U = 960 \) for \( U = 228 \).

To implement the REG and the RRS methods for estimating \( H \), given a maximum block size \( L^* \), we divide the observed series into a sequence of non-overlapping blocks of sizes \( n_{j+1} = n_{\text{min}} + j; \ j = 0, 1, \ldots L^* - n_{\text{min}} \). We chose \( n_{\text{min}} = 24 \). For any given block size \( n_j < L^* \), the number of blocks into which we divide the data is therefore given by \( b_j = N/n_j \). To compute an estimator \( \hat{H}(L^*) \) we proceed as follows:

1. In each non-overlapping block of size \( n_j \) we obtain the statistic \( \eta(n_j, L^*) \), which we will use to compute the estimator \( \hat{H}(L^*) \) according to one of the estimation methods presented in Section 3.2.2. Thus, for a block size equal to \( n_j \), we obtain \( b_j \) values of the statistic.

2. Averaging these \( b_j \) values results in a statistic denoted by \( \bar{\eta}(n_j, L^*) \). We obtain \( L^* - n_{\text{min}} + 1 \) such averages for each \( L^* \).

3. The estimator \( \hat{H}(L^*) \) of the self-similarity index is computed from the averages \( \bar{\eta}(n_j, L^*) \) according to one of the methods outlined in Section 3.2.2. For example, in the case of the RRS method, the averages \( \bar{\eta} \) correspond to averages of the \( R/S \) statistic and to obtain \( \hat{H}(L^*) \) we regress those \( \bar{\eta}(n_j, L^*) \) on the log of \( n_j \leq L^* \).

In the case of the WHI estimator, blocks of size \( L_u \) are constructed but no further blocking is required. The series is then divided into blocks of size \( L_u, u = 1, \ldots, U \) to obtain \( b_u = N/L_u \) blocks. The WHI estimator is computed using observations in each block. For a given maximum block size \( L^* \), \( \hat{H}(L^*) \) is obtained as the average, over the \( b_u \) blocks, of
each of the block-level WHI estimators. Note that repeating this procedure for several maximum block sizes \( L_1, L_2, \ldots, L_U \) leads to an assessment of the effect of maximum block size on the estimator \( \hat{H}(L_u) \) (as stated in the first bullet above). We carried out the procedure described above for weekly S&P500 observations and results are presented in Section 3.3.1 below.

For a given maximum block size, the potential time-varying nature of the memory index (estimated as the function \( \hat{H}(L_u) \)) can be investigated by dividing the series into segments and then obtaining the estimator of \( H \) separately in each segment. The segments can be non-overlapping or can be constructed using a moving-window type of approach. For example, for the S&P500 series we could divide the period Jan 3., 1950 - Nov. 27, 2006 into a set of overlapping time periods (e.g., from 1950 to 1960, 1952 to 1962, etc.) and then obtain an estimate of \( H \) using data from each of the segments exclusively. In that way, we can investigate whether inferences about the memory structure in the S&P500 depend on the time span under study. We carried out an analysis of this type and results are presented in Section 3.3.2.

### 3.3.1 The effect of maximum block size on the estimate of \( H \) for fractional Gaussian noise and for S&P500 returns for the time period Jan. 3, 1950 - Nov. 27, 2006

Several of the estimators of \( H \) that have been proposed in the literature require that the data first be divided into non-overlapping blocks of appropriate size. For example, the RRS, RSE and REG approaches rely on blocking (or windowing) approach. Since the number of blocks and the maximum block size are, to a large extent, arbitrary, the question arises as to whether the choice of maximum block size (and consequently, of number of blocks) affects the value of the estimate of \( H \) computed from the data.
We first implemented the REG, RRS and WHI methods for estimating $H$ on $R = 10$ simulated fGn sequences of length $N = 10,000$ with nominal $H$ values equal to 0.2, 0.4, 0.5, 0.6 and 0.8. (We ignore RSE, since it requires less assumptions and behaves worse than the other three estimators.) As described earlier, maximum block sizes were fixed at $L_u = L_1 + (u - 1) \cdot \text{step}$, where $L_1 = 48$ and $\text{step} = 4$ weeks. The minimum block size in every case was 24. Thus, for $L_1 = 48$, the possible block sizes were 24, 25, ..., 48.

Figures 3.3 (a)-(e) show results. In the figures, we plot, for each of the three estimation methods, the mean (over the $R$ replicates) estimate of $H$ for each maximum block size $L$, as well as the ± one standard deviation bands for $\hat{H}$ (where the standard deviations are also computed over the $R$ replicates).

We note that in the case of fGn sequences, the estimate of $H$ is essentially independent of maximum block size and that this holds for any of the three estimation methods.

To explore whether the choice of maximum block size $L$ affects the estimate of $H$ computed from weekly S&P500 observations, we proceeded in a similar way. For weekly data (closing index values at the end of each week of Jan. 3, 1950 - Nov. 27, 2006), we choose the same sequence of maximum block sizes $L = 48, 52, ..., 960$ that was used to explore the fGn sequences and that for the S&P500 series cover about 20 years in 4 week intervals. We then proceeded with the estimation of $H$ using a subset of the methods described in Section 3.2.2. The results of the computations are shown in Figure 3.4, where we depict $\hat{H}(L)$ for each of the four estimators obtained from the weekly time series.

As expected, the estimated $\hat{H}(L)$ values for the R/S based estimators RRS and RSE are considerably larger than the values obtained via REG or via WHI, most likely a consequence of the systematic positive bias of these two methods that we found in Section 3.2.3. Indeed, the $H$-dependent bias of the RSE estimator results in a $\hat{H}_{RSE}(L)$ curve that shows little structure for weekly data. The residuals of regression method (REG) shows some roughness.
Figure 3.3  Estimate $\hat{H}$ and one-standard deviation confidence bands as a function of maximum block size $L$ and using three different estimation methods. From (a) to (e) the fGn sequences had nominal $H$ values equal to 0.2, 0.4, 0.5, 0.6 and 0.8, respectively.
Weekly price data, 48–960 Ws every 4

Figure 3.4  $\hat{H}$ as a function of time lag.
for the weekly data graphs for small window sizes.

For the REG estimator, we obtain the following results for S&P500 weekly returns over the time period Jan. 3, 1950 to Nov. 27, 2006:

- The estimated memory parameter $\hat{H}(L)$ increases from about $\hat{H}(48) = 0.47$ to $\hat{H}(236) = 0.55$ and then decreases to $\hat{H}(960) = 0.43$.
- Within the interval of increase [48, 236] and the interval of decrease [236, 960] the behavior of $\hat{H}(L)$ is basically monotone.
- The curve $\hat{H}(L)$ intersects the $H = 0.5$ line at maximum block sizes of 92 and of 480.

Hence the S&P500 data shows anti-persistent and mean-reverting behavior for blocks of maximum size $L \leq 24$ months and $L \geq 120$ months, while for maximum block sizes $L \in [24, 120]$ months the series shows persistent and long-term effects, with a maximum at about 80 months. The daily and monthly data (not shown) confirm these results.

Comparing these findings to the ones obtained in the literature by Peters (1996) we notice, on the one hand, that the estimates obtained here result in substantially lower values for $\hat{H}$ that the ones reported by Peters in Figure 9.1 for daily data. The main reason for this difference might be the overestimation of $H$ that results from using the R/S statistic. Another reason could be that the time period used by Peters is 1928-1990 (for daily data), while our analysis is based on the period 1950-2006, compare also the discussion in Section 3.3.2 below about the change of $H$ over the years. The systematic overestimation of $H$ obtained when using the R/S statistic calls into question the relationship to the Pareto distribution discussed by Peters on pp. 107-112. On the other hand, our analysis confirms the observation that there is a specific maximum block size for which the long-range memory parameter $H$ achieves its maximum. We estimate this maximum block size to be $L = 236$. 

trading weeks (or about 58 months), as compared to Peters’ estimate of 1,000 trading days, or roughly 50 months (see Peters, 1996, Figures 8.2 and 9.1).

One other finding arises from the very different behavior of the estimators of $H$ when applied to the simulated fGn sequences and the real S&P500 data. We note that for fGn sequences, the effect of maximum block size on $\hat{H}$ for any method is negligible for any of the estimators used here. This is not true, however, in the case of the actual S&P500 data, were we observe a pronounced effect of $L$ on $\hat{H}$ except in the case of the empirical R/S estimator and the Whittle estimator. This seems to suggest that the S&P500 sequence does not behave like a fractional Gaussian noise sequence.

### 3.3.2 Time varying memory structure of S&P500 returns for time segments during the period Jan. 3, 1950 - Nov. 27, 2006

The previous section discussed the effect of the choice of maximum block size on the estimate of the memory parameter in fGn simulated sequences and in the S&P500 returns over the entire period Jan. 3, 1950 - Nov. 27, 2006. Of potentially greater interest is the question whether the memory structure of a time series changes over time. If financial markets have different memory structures when looked at over different periods then it might be possible to identify economic or other factors that are associated with those structural memory changes.

Using daily log –returns for the period Jan. 2, 1928 - July 5, 1990 Peters found (via the R/S estimator) that the memory parameter $H$ was constant over these decades (Peters, 1996, pp. 113). Bayraktar et al. (2004) analyzed the period Jan. 1989 - May 2000 using a wavelet estimator for the time series given by the log –prices, i.e. they considered $Y_k := \log \left( \frac{P_k}{P_1} \right)$. Their findings include a drop in $\hat{H}_{WAV}$ around 1997 to a level close to 0.5.

Similar to our approach in Section 3.3.1, we first simulated fGn sequences for different
values of $H$ and estimated the memory parameter in different time segments using the estimators REG, RRS, and WHI. (We ignore RSE, since it requires less assumptions and behaves worse than the other three estimators.) None of the three methods showed any time-dependence of the index $H$ in fGn sequences.

Using the REG estimator, we studied the segment-dependent memory structure of S&P500 returns, based on weekly closing data for the time period Jan. 3, 1950 - Nov. 27, 2006. Figure 3.5 shows the time development of $\hat{H}_{REG}(L)$ for blocks of maximum length (a) $L = 139$ weeks (about 32 months), (b) $L = 236$ (about 54 months), and (c) $L = 332$ (about 77 months). These maximum block sizes were chosen to reflect the maximal estimated $H$-value ($\hat{H}_{REG} = 0.55$ at $L = 236$) and the crossings of the $\hat{H}_{REG} = 0.53$ level ($L = 139$ and $L = 332$). For each maximum block size the figure shows the estimates $\hat{H}_{REG}(L)$, the linear regression line and a non-parametric regression line using local polynomials with optimal bandwidth which is implemented as function “lopoly” in the statistical software package R (e.g., Wand and Jones, 1995).

All three graphs show a linear negative trend over time for $\hat{H}_{REG}(L)$. The non-parametric regressions also suggest negative trends for all three block sizes, with short time periods of increasing $H$ in the 1950’s and 2000’s for the shortest window size $L = 139$.

In order to interpret these trends and to put the estimates for the time behavior of the self-similarity index $H$ of S&P500 logarithmic returns in context, we take a closer look at Figure 3.5 (b), the variation of $\hat{H}_{REG}$ over time for block size $L = 236$ (about 54 months, or 4.5 years). For each time segment $i = 1, \ldots, 12$ the value $\hat{H}_{REG}(i)$ estimates $H$ over the previous period of 236 trading weeks, e.g. $\hat{H}_{REG}^{(6)}(L = 236)$ estimates $H$ for the period July 1972 - Jan. 1977 and $\hat{H}_{REG}^{(12)}(L = 236)$ estimates $H$ for the period July 1999 - Jan. 2004. According to the monotonicity intervals of Figure 3.5 (b), one can aggregate the 12 time segments into seven segments, as shown in Table 3.4. To account for the block size, we list
Figure 3.5 $\hat{H}$ over time for three different maximum block sizes.
Table 3.4  Monotonicity intervals of $\hat{H}$ for maximum block size 4.5 years

<table>
<thead>
<tr>
<th>Interval</th>
<th>Period</th>
<th>Length</th>
<th>$\hat{H}$ tendency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/50-10/56</td>
<td>6y 9m</td>
<td>↑</td>
</tr>
<tr>
<td>2</td>
<td>10/56-4/61</td>
<td>4y 6m</td>
<td>↓</td>
</tr>
<tr>
<td>3</td>
<td>4/61-10/74</td>
<td>13y 6m</td>
<td>↑</td>
</tr>
<tr>
<td>4</td>
<td>10/74-4/79</td>
<td>4y 6m</td>
<td>↓</td>
</tr>
<tr>
<td>5</td>
<td>4/79-10/83</td>
<td>4y 6m</td>
<td>↑</td>
</tr>
<tr>
<td>6</td>
<td>10/83-4/97</td>
<td>13y 6m</td>
<td>↓</td>
</tr>
<tr>
<td>7</td>
<td>4/97-1/2004</td>
<td>6y 9m</td>
<td>↑</td>
</tr>
</tbody>
</table>

the intervals with breaks at 1/2 the maximum block size, i.e. at 2 years and 3 months.

A decrease of $\hat{H}$ during period $i$ indicates stronger short term memory dependence within this period, i.e. the time series shows less persistence and a stronger mean-reverting tendency. It appears that the memory index $H$ of the S&P500 weekly data has shown values below 0.5 since around 1982, meaning that the market has been dominated for this time period by increasingly short memory effects. This trend may have been reversed during the last few years, at least for shorter block sizes.

3.3.3  The combined effect of maximum block size and decade on the estimated value of $H$

We now investigate the combined effect of choice of maximum block size and time segment on the estimated $H$. As we did earlier, we first focus on the estimates computed from simulated fGn sequences and then explore the results obtained from the S&P500 series.

To investigate the combined effect of maximum block size and time segment we proceeded as follows for the simulated fGn sequences:

1. We considered moving time segments of size 500 observations.
2. For a 100-observation step, each segment had an overlap of 400 observations with the preceding segment.

3. Within each segment, we computed $\hat{H}(L_u)$, for $L_u = 50, 55, ..., 250$ observations.

   That is, we considered $b = 250 - 50 + 1 = 201$ possible maximum block sizes.

   In the case of the observed S&P500 series, the process was similar except that the moving window segments were of length 10 years and the moving step was two years. Further, for the S&P500, $L_u = 48, 52, 56, ..., 240$ for a total of $b = 193$ possible values of $L$.

   Figure 3.6 (a)-(e) correspond to results obtained from the fGn simulated sequences with nominal $H$ values of 0.2, 0.4, 0.5, 0.6 and 0.8. Figure 3.7 (a) and (b) correspond to results obtained from the S&P500 series.

   In Figure 3.6 (a)-(e) and in the (a) panel of Figure 3.7 we present the maximum (over the $b$ maximum block sizes) value of $\hat{H}(L_u)$ in each time segment. The (b) panel in Figure 3.7 shows the maximum block size $L_u$ within each time segment at which the highest value of $\hat{H}(L_u)$ was obtained. As before, we see that for fGn sequences, the effect of maximum block size is negligible for all estimation methods and all nominal $H$ values. (Clearly, we do not expect to see a time segment effect in the simulated series.) In the case of the S&P500, however, we observe that the maximum value of $\hat{H}(L_u)$ varies between a low of about 0.4 and a high of about 0.6 (for the Whittle estimator) and that the maximum block sizes that produce those maximal estimates are also highly variable.

   Panel (a) of Figure 3.7 is consistent with the results discussed in Section 3.3.2. The variability in panel (b) of this figure calls into question the significance of the block size when analyzing memory structures in the S&P500 return data: None of the three estimators shows a consistent block size over the time period from 1950 - 2006, indicating that the observations in Section 3.3.1 refer to averages over time segments within this period. In
Figure 3.6 Mean and one-standard deviation bands (over replicates) of the maximum value of the estimate of $H$ obtained using three estimation methods and a moving time segment. Data are simulated fGn sequences of length 10,000. In each time segment, maximum block size was allowed to vary between 50 and 250 periods, with a step of 5. Panels (a)-(e) correspond to nominal values of $H$ equal to 0.2, 0.4, 0.5, 0.6 and 0.8.
Figure 3.7 Panel (a): Maximum value of the estimate of $H$ obtained using three estimation methods and a moving 10-year time segment. In each time segment, maximum block size was allowed to vary between 48 and 240 weeks, with a step of 5 weeks. Panel (b) shows the maximum block size in each time segment for which the maximal $\hat{H}(L)$ was obtained.
particular, interpretations of notable memory effects in the S&P500 time series based on
maximal block sizes (compare, e.g. Peters, 1996) have to be taken with some care, at least
for the second half of the 20th century.

3.3.4 Validation of the estimation of the self-similarity index

Regardless of the method that is used to obtain an estimate \( \hat{H}(L) \), it is often of interest
to test hypotheses about the underlying, unobservable \( H \). One hypothesis of interest is
whether \( H = 0.5 \) or in other words, whether the series on which the estimate \( \hat{H} \) is based has
no memory structure. In principle, to carry out a test of the null hypothesis \( H_0 : H = 0.5 \)
against the alternative \( H_a : H \neq 0.5 \) we need to derive the distribution of \( \hat{H}(L) \) under the
null hypothesis.

An alternative approach consists in approximating the distribution using an empiri-
cal non-parametric approach. The resulting test is similar to the usual permutation test
originally proposed by Pitman (1938, pp. 323-324) and more recently discussed by, e.g.,
Welch (1990, Section 2). In the context of an empirical non-parametric test, Scheinkman
and LeBaron (1989, pp. 317-318) proposed a test called the scrambling test to investigate
whether a series has the property of long memory.

The idea behind the scrambling test is the following: If a series exhibits long memory,
then the residuals of the regression of the current observation on the previous observation
would be correlated. Suppose now that the observation at time \( k \) is substituted by the value
obtained by adding a randomly chosen estimated regression residual to the observation at
time \( k - 1 \). The new sequence is called a scrambled sequence and in principle would have
less of a memory structure than the original sequence. The test therefore proceeds as
follows:

1. Obtain the estimated residuals from the regression of \( Y_k \) on \( Y_{k-1} \), \( k = 2, \ldots, N \).
2. Using the observed $Y_1$ as the “anchor”, construct a scrambled series of length $N$ by replacing $Y_2, ..., Y_N$ by values obtained as the sum of the previous value and a residual randomly chosen from among the $N$ residuals computed in Step 1.

3. Repeat the second step $M$ times to obtain $M$ scrambled series.

We then obtain $M + 1$ estimates of $H$, one from the original series and one from each of the $M$ scrambled series, which we denote $\hat{H}(L)$ and $\hat{H}^{(i)}(L)$, $i = 1, ..., M$, respectively. The empirical distribution of the $\hat{H}^{(i)}(L)$ can be viewed as the distribution of the estimator under the null hypothesis (i.e. under the random walk hypothesis for $H^* = \frac{1}{2}$), although this is not strictly the case since only residuals and not observations themselves are permuted. This pseudo-null distribution nonetheless provides a reference against which to compare $\hat{H}(L)$. If $\hat{H}(L)$ is likely under the pseudo-null distribution, then this suggests that the original series does not exhibit a memory structure.

In order to test the validity of our estimates of the memory parameter $H$ we implemented the scrambling test as described in above. To illustrate the procedure, we first consider three simulated fractional Brownian motion sequences of length 10,000 with different $H$—values $H = 0.3$, $0.5$, and $0.7$. For each of the original sequences we implemented the regression method to obtain an estimate $\hat{H}_{REG}$ of $H$, which in all cases was close to the nominal value. We then scrambled each sequence $M = 100$ times and re-estimated $H$ in each of the scrambled sequences. Figure 3.8 shows, for each nominal value of $H$, the histogram of the estimates computed from each of the scrambled sequences. The vertical bar on each of the plots indicates the value of $\hat{H}_{REG}$ obtained from each of the three un-scrambled series. From the figure we see that for small or large values of $H$, the estimate $\hat{H}_{REG}$ is very unlikely under the random walk assumption. In both cases we would tend to conclude that the original sequence had short (for $H = 0.3$) or long (for $H = 0.7$) memory.
When the nominal $H$ was 0.5 (i.e., the random walk case) the estimate $\hat{H}_{REG}$ falls in the middle of the reference distribution, as would be expected.

Note that only for $H = 0.5$ the reference distribution is symmetrically distributed around 0.5 as would be expected from a true random walk sequence. For $H = 0.3$ and $H = 0.7$ the distributions are not symmetric and in the case of $H = 0.7$ is not even centered around 0.5. This may be because the scrambled sequences preserve some of the memory structure in the original sequence as only the residuals from the regression of $B_H(k)$ on $B_H(k-1)$ are permuted. Thus the pseudo-null or reference distributions are shifted either downward or upward depending on the true value of $H$ in the unscrambled sequence.

We implemented the scrambling test with $M = 100$ for the S&P500 index sequence, using weekly closing data and considering weeks 48 to 960 under the regression estimator (REG). The results are shown in Figure 3.9.

This figure shows the pseudo-null distributions and as a vertical line the estimate of $H$ obtained from the unscrambled data for lags of 100 to 800 weeks. The estimate $\hat{H}_{REG}$ is highest at the 200 week lag and decreases as lag time increases, as has already been seen from Figure 3.4. Furthermore, we observe the following:

1. The distance between the center of the histogram and the vertical line (representing
Figure 3.9  Histograms of REG estimations for unscrambled and scrambled weekly data.
the unscrambled data) decreases, as the block size grows from 100 weeks to 800 weeks. This means that the estimated value of $H$ becomes less dependent on the order of the observations as the block size grows.

2. For intermediate block sizes, i.e. when the estimator $\hat{H}_{REG}$ of the original, unscrambled data series is above 0.53, the center of the histogram is closer to 0.5 than the vertical line, i.e. the value of $\hat{H}_{REG}$ for the unscrambled data. This indicates that S&P500 index behaves like a long memory time series for blocks of size $L \in [100, 350]$ weeks.

3. For blocks of size $L < 100$ weeks and $L > 350$ weeks, the center of the histogram is below 0.5, which would imply that the scrambled data has short memory. Thus when the estimated $H$ from the unscrambled data is around 0.5 the implemented scrambling test fails. We comment on several reasons for this behavior below.

The main reason for the failure of the scrambling test in our context for maximum block sizes $L < 100$ weeks and $L > 350$ weeks, seems to be the number of available observations in the blocks. The maximum block size for weekly data includes 1000 observations that can be used for the estimation of $H$, which may be insufficient for $H < 0.53$ as shown above for fractional Brownian motion. Furthermore, the scrambled sequences preserve some of the memory structure in the original sequence as only the residuals from the regression of $S_n$ on $S_{n-1}$ are permuted. Finally, the actual S&P500 log-return data series may not satisfy the assumption on stationary increments that is implicit in the use of the regression estimator (REG).
CHAPTER 4. FRACTIONAL STOCHASTIC MODELS

4.1 Introduction

In order to make these existing financial models capable to obtain a long or short memory structure, a heuristic method is to incorporate fractional processes, e.g. fractional Brownian motion, in them. Following this idea, our research deals with the stochastic volatility models, extend the LogOU and CIR model (two of the popular existing finance models), argued carefully about the support area of the fractional parameters that well define the pathwise integral with respect to fBm, and explore the memory structures underlying these two modified models through simulation and analytical approaches.

The chapter is organized as follows. We first revisit the Black-Scholes model and its extensions (Section 4.2). We then summarize existing results on the Hurst exponent and show that the current LogOU and CIR model fail to capture the memory structure in the data (Section 4.3). In Section 4.4, we introduce fractional Brownian motion processes. Later, in Section 4.5, we propose the modified LogOU stochastic volatility models, and discuss the well-definedness of its solution in the pathwise sense. The joint behavior of the price and volatility processes under the modified LogOU is also studied in this section. The related simulations are conducted in Section 4.6. The well-definedness problem and the joint behavior of the price and volatility of modified CIR models are explored similarly in Section 4.7. Finally in Section 4.8 we propose future avenues for research.
4.2 Black-Scholes Model and its Extensions

Linear models that can be used to describe the evolution of prices and the variability in the prices of financial instruments over time have been discussed extensively in the literature (Hull, 2002). One of the earlier efforts to jointly model prices and volatility was the work by Black and Scholes (1973) which is described below. The Black-Scholes approach involved several assumptions that cannot be justified in the case of financial time series and thus several more general versions of the Black-Scholes model have been proposed recently. Below we first introduce the simple Black-Scholes pricing model. We then formulate a general version of the linear stochastic volatility model and then describe several special cases of the general linear model that have appeared in the literature.

We use $S_t$ to denote the stock price at time $t$ and let $Y_t = \log(S_t)$. Both $Y_t$ and $S_t$ are described by a probability space $(\Omega, F, P)$ with the usual properties. In the models we discuss below, $Y_t$ is the response process. The more general models also involve a process $V_t$, the volatility of prices at time $t$. Note that for the general model, the $V_t$ process is a Markov process, while $Y_t$ itself is not. Jointly, however, $V_t$ and $Y_t$ form a two-dimension Markov process.

4.2.1 The Black-Scholes model

Under the assumptions of European options the Black-Scholes model for $Y_t$, depends on a constant drift $\mu$ and a constant volatility $\sigma$:

$$dY_t = \mu dt + \sigma dW_t,$$  \hspace{1cm} (4.1)

where $W_t$ is Brownian motion defined on the probability space, and $t \in R^+$. 
4.2.2 General stochastic volatility model

If the constant volatility assumption in the simple Black-Scholes model is relaxed then a more general approach to model $Y_t$ is

$$\sigma_t = f(V_t),$$

$$dY_t = \mu dt + \sigma_t dW^y_t,$$

$$dV_t = (a + bV_t)dt + c(V_t)^\gamma dW^\nu_t,$$

$$dW^\nu_t = \rho dW^y_t + \sqrt{1 - \rho^2} dW_t,$$

(4.2)

for $\mu, a, b, c, \gamma, \rho$ unknown constant over time, and for $W^y_t$ and $W_t$ independent Brownian motion processes. The model above is general in that it lets both prices and their volatility change over time and allows for the presence of a feedback mechanism between $Y_t$ and $V_t$. Because the two random drivers of the $Y_t$ and $V_t$ processes are correlated with correlation coefficient $\rho$, volatility can affect price and vice versa. However, instead of the correlation between $Y_t$ and $V_t$, the focus of the manuscript is on the memory structure of the stochastic volatility models, we therefore set $\rho = 0$ in the general model form, and later their extensions do not consider the correlated random drivers of the $Y_t$ and $V_t$ processes.

4.2.3 Log Ornstein-Uhlenbeck (LogOU) model

The Log Ornstein-Uhlenbeck model (Uhlenbeck and Ornstein, 1930) is one stochastic volatility model. Here, the two random drivers are assumed to be independent and the dynamics of $V_t$ allow for a constant drift:

$$\sigma_t = f(V_t) = e^{V_t},$$

$$dY_t = \mu dt + \sigma_t dW^y_t,$$

$$dV_t = (a + bV_t)dt + c dW^\nu_t,$$

(4.3)
for $\mu, a, b, c$ constant over time and for $W_t^y$ and $W_t^v$ independent Brownian motion processes.

Further, to have a stationary solution for the volatility equation, we set $a = 0$, $b < 0$ and $c > 0$, and then have $V_t$ is normally distributed with expectation and covariance

$$E(V_t) = 0, \text{Cov}(V_s, V_t) = -\frac{c^2}{2b} e^{b|s-t|},$$

and the conditional distribution of $V_t$ given $V_s$ ($t > s$) is also a normal distribution with expectation and variance

$$E(V_t|V_s) = V_s e^{b(t-s)}, \text{Var}(V_t|V_s) = -\frac{c^2}{2b} \left(1 - e^{b(t-s)} \right),$$

and the conditional distribution of $V_t$ given $V_0 = v_0$ is a normal with expectation $v_0 e^{bt}$ and covariance $\text{Cov}(V_t, V_s|V_0 = v_0) = -\frac{c^2}{2b} \left(e^{b(t-s)} - e^{b(t+s)} \right)$. (When $t \to \infty$, the limiting distribution for $V_t$ is a normal distribution with expectation 0 and variance $-\frac{c^2}{2b}$.)

4.2.4 Cox-Ingersoll-Ross (CIR) model

The Cox-Ingersoll-Ross model (Cox et al., 1985a, 1985b) is another commonly used stochastic volatility model in financial applications. The CIR model is similar to the LogOU model except the function used to define the volatility:

$$\sigma_t = f(V_t) = \sqrt{V_t}$$
$$dY_t = \mu dt + \sigma_t dW_t^y,$$
$$dV_t = (a + bV_t)dt + c V_t dW_t^v,$$

for $\mu, a, b, c$ constant over time and for $W_t^y$ and $W_t^v$ independent Brownian motion processes. In application, the CIR model is often written in a different form:
\[
\sigma_t = f(V_t) = \sqrt{V_t}
\]
\[
dY_t = \mu dt + \sigma_t dW^y_t,
\]
\[
dV_t = \kappa(\theta - V_t) dt + \sigma_v \sqrt{V_t} dW^v_t,
\]
where \( \kappa = -b, \ \theta = a/\kappa \) and \( \sigma_v = c \). Usually, we assume that \( \kappa > 0, \ \theta > 0 \) and \( \sigma_v > 0 \). (I.e., assume \( a > 0, \ b < 0 \) and \( c > 0 \) in the first expression of CIR model.) Under these assumptions the \( V_t \) process is mean-reverting. Under the assumption that both \( \kappa, \theta \) are positive, \( V_t \) can be shown to be a non-central \( \chi^2 \) random variable with the expectation equal to
\[
E(V_t|V_0 = v_0) = \theta + (v_0 - \theta) e^{-\kappa t}
\]
and variance equal to
\[
Var(V_t|V_0 = v_0) = \frac{\theta \sigma_v^2}{2 \kappa} + \frac{\sigma_v^2}{\kappa} (v_0 - \theta) e^{-\kappa t} + \frac{\sigma_v^2}{\kappa} \left( \frac{\theta}{2} - v_0 \right) e^{-2\kappa t}.
\]
The limiting distribution of \( V_t \) is a Gamma distribution with expectation \( \theta \) and variance \( \frac{\theta \sigma_v^2}{2 \kappa} \) (Cox et al., 1985a, 1985b).

### 4.3 Memory Structure and the Hurst Exponent

Peters (1996) used the Hurst exponent to study a real dataset – S&P 500 price data in a special way. Using the notation introduced earlier, the response variable for his analysis was the return – \( (Y_t - Y_{t-1}) \), for \( Y_t = log(S_t) \) and \( S_t = \) the stock price at time \( t \). Given a maximum block size \( N \) Peters computed the value of the \( R/S \) statistic in the following steps:

1. For a certain reasonable increment of time \( N_0, N_0 \leq N \), divide the data into several non-overlapping blocks, with sample size in each block equal to \( N_0 \).
2. Calculate the $R/S$ statistic in each of the blocks.

3. Average the $R/S$ values of the statistic over all the blocks.

After obtaining the averaged value of the statistic for all reasonable increments of time $N_0$, $N_0 \leq N$, a regression based on the relationship between the $R/S$ statistic and $H$ can be used to estimate $H$ for a given maximum block of size $N$. Further, Peters repeated the process for other choices of time window size $N$. This allows people to investigate the estimates of $H$ corresponds to different maximum block sizes. Below, we followed Peters (1996) and produced Figure 4.1 shows the estimated Hurst exponent for different time window sizes computed from the S&P 500 daily returns from January 2$^{nd}$, 1980 to December 29$^{th}$, 2000.

![Figure 4.1](image)

In Figure 4.1, the values of the estimated Hurst exponent over time are not always 0.5, but increase from 0.5 at the very beginning, peak to a value above 0.5 sometime around three years, and then decrease to be around 0.5 (after some time around 6 years). This figure unveils the complicated memory structure underlying a process. Other studies supportive that memory structure underlying a process can be different from a random walk or a short memory process include the study of precipitation in a local area (Mandelbrot and Wallis, 1968), and the study of nanoscale single-molecule biophysics experiments (Kou
and Xie, 2004). These studies motivated us to extend the LogOU and the CIR model into more general form in order to account for a short or long memory in a process.

4.4 Stochastic Integrals and Fractional Brownian Motion

Definition 4.4.1 (Øksendal, 2004): Let \( H \in (0, 1) \) be a constant. The (1-parameter) fractional Brownian motion (fBm) with Hurst parameter (the self-similar index of this process) \( H \) is the Gaussian process \( B_H(t) = B_H(t, \omega), \ t \in \mathbb{R}, \ \omega \in \Omega, \) satisfying \( B_H(0) = E[B_H(t)] = 0, \) for all \( t \in \mathbb{R}, \) and \( E[B_H(s)B_H(t)] = \frac{1}{2}\{|s|^{2H} + |t|^{2H} - |s - t|^{2H}\}, \) for all \( s, t \in \mathbb{R}. \) Here \( E(\cdot) \) denotes the expectation with respect to the probability law \( \mathbb{P} \) for \( \{B_H(t)\}_{t \in \mathbb{R}} = \{B_H(t, \omega) : t \in \mathbb{R}, \omega \in \Omega\}, \) where \( (\Omega, \mathcal{F}) \) is a measurable space.

For this Gaussian process, two types of stochastic integrals are defined by people and used a lot. One is the pathwise or forward integral, the other is the Skorohod (Wick-Itô) integral. To facilitate the simulation method for the stochastic integral with respect to a fractional Brownian motion, we choose the former way to define stochastic integrals with respect to fractional Brownian motions.

The pathwise integral is usually denoted by

\[
\int_0^T \phi(t, \omega) d^-W^H_t.
\]

Given the integrand \( \phi(t, \omega) \) is caglad (left-continuous with right sided limits), this integral is described by Riemann sums:

Let \( 0 = t_0 < t_1 < \cdots < t_N = T \) be a partition of \([0, T]\). Put \( \Delta t_k = t_{k+1} - t_k \) and define

\[
\int_0^T \phi(t, \omega) d^-W^H_t = \lim_{\Delta t_k \to 0} \sum_{k=0}^{N-1} \phi(t_k, \omega) \left( W^H_{t_{k+1}} - W^H_{t_k}\right),
\]
if the limit exists (e.g. in probability). Using a classical integration theory, one can prove that the pathwise integral exists if the p-variation of \( t \to \phi(t, \omega) \) is finite for all \( p > (1-H)^{-1} \) and a given value of \( H \in (0,1) \). (Øksendal, 2004.)

Notice that some forms of stochastic integrals with respect to fractional Brownian motions may not be able to define in the pathwise sense for all the \( H \) values on \((0,1)\). Therefore, the support area of the \( H \) value for the well-defined pathwise integrals should be carefully stated. Some sufficient conditions for the well-defined pathwise integrals are provided in literature, for example, by using q-variation which definition is mentioned in Øksendal (2003, Page 19). Since \( t \to W^H_t \) has finite q-variation if and only if \( q \geq \frac{1}{1-H} \), we see that if \( H < \frac{1}{2} \) then this theory does not even include integrals like (Øksendal, 2004)

\[
\int_0^T W^H_t \, dW^H_t.
\]

In other words, the support area of the \( H \) value this specific integral is \([0.5,1)\).

### 4.5 Fractional Log Ornstein-Uhlenbeck (LogOU) Model

#### 4.5.1 The well-definedness of the fractional LogOU model system

An extension to the Ornstein-Uhlenbeck model, is to have a fractional Brownian motion in the volatility equation to capture the possible long-range memory in the data. In other words, a fractional Ornstein-Uhlenbeck (volatility) model can be defined as:

\[
\begin{align*}
    dV^H_t &= b V^H_t \, dt + c dW^H_t,
\end{align*}
\]

for \( b \) and \( c \) constant over time and \( W^H_t \) being a fBm with the index \( H^\nu \in (0,1) \). The way to define \( dW^H_t \) is to follow the pathwise explanation. Since now, \( c \) is constant, this extended version of the Ornstein-Uhlenbeck (volatility) model is well defined automatically.
Further, the solution of the stochastic equation in above mentioned by Hu (2002) is in the following form:

\[ V_t^H = e^{bt}V_0^H + c \int_0^t e^{(t-s)b}dW_s^{H^\nu}. \]

Cheridito et al. (2003) discussed the behaviors of the fractional Ornstein-Uhlenbeck process in a pathwise sense, and mentioned that a stationary solution to this model is

\[ V_t^H = c \int_{-\infty}^t e^{b(t-s)}dW_s^{H^\nu}. \]

Further, they proved some interesting results for to this integral.

**Proposition 4.5.1** (Proposition A.1, Cheridito et al., 2003.) Let \( \{W_t^H\}_{t \in \mathbb{R}} \) be an fBm with \( H \in (0, 1] \). Let \(-\infty \leq a < \infty\), and \( \lambda > 0 \). Then, for almost all \( \omega \in \Omega \), we have the following:

1. For all \( t > a \),
   \[
   \int_a^t e^{\lambda s}dW_s^H(\omega)
   \]
   exists as a Riemann-Stieltjes integral and is equal to
   \[
   e^{\lambda t}W_t^H(\omega) - e^{\lambda a}W_a^H(\omega) - \lambda \int_a^t W_s^H(\omega)e^{\lambda s}ds.
   \]

2. The function
   \[
   \int_a^t e^{\lambda s}dW_s^H(\omega), t > a
   \]
   is continuous in \( t \).

3. Let \( H \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1] \), \( \lambda > 0 \) and \(-\infty \leq a < b \leq c < d < \infty\). Then
   \[
   E \left( \int_a^b e^{\lambda u}dW_u^H \int_c^d e^{\lambda \nu}dW_\nu^H \right) = H(2H - 1) \int_a^b \int_c^d e^{\lambda u}(\nu - u)^{2H-2}d\nu du.
   \]
Theorem 4.5.2 (Theorem 2.3, Cheridito et al., 2003.) Let $H \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$, and $N = 1, 2, \ldots$. Then, for fixed $t \in \mathbb{R}$ and $s \to \infty$,

$$
\text{Cov} \left( V_H^t, V_H^{t+s} \right) = \frac{1}{2} c^2 \sum_{n=1}^{N} |b_n|^{-2n} \left\{ \prod_{k=0}^{2n-1} (2H - k) \right\} s^{2H - 2n} + O(s^{2H - 2N - 2}).
$$

Theorem 4.5.3 (Remark 2.4, Cheridito et al., 2003.) For all $t, s \in \mathbb{R}$ and $H \in (0, 1)$, the autocovariance function (ACVF) of the solution from fractional Ornstein-Uhlenbeck model can be expressed explicitly as:

$$
\text{Cov} \left( V_H^t, V_H^{t+s} \right) = c^2 \Gamma(2H + 1) \sin \left( \frac{\pi H}{2} \right) \int_{-\infty}^{\infty} \frac{e^{isx}|x|^{1-2H}}{|b|^2 + x^2} dx, \tag{4.6}
$$

for $i = \sqrt{-1}$.

Below, we provide a theorem on the well-definedness of the whole fractional LogOU dynamic model system which includes both the fractional OU model as the fractional volatility and the price model in the fractional form.

Theorem 4.5.4 Define a fractional LogOU dynamic model system as

$$
dY_H^t = \mu dt + e^{V_H^t} dW^{H_Y}_t, \quad dV_H^t = bV_H^t dt + c dW^{H_V}_t,
$$

where $W^{H_Y}_t$ and $W^{H_V}_t$ are two independent fractional Brownian motions with their own self-similarity index being $H_Y$ and $H_V$ and with $\text{Var} \left( W^{H_Y}_1 \right) = \text{Var} \left( W^{H_V}_1 \right) = 1$, and $b < 0$, $c > 0$. Then, this dynamic system is well defined for all pairs of $(H_V, H_Y) \in (0, 1) \times (0, 1)$ that satisfy $H_V + H_Y \geq 1$.

Proof. For the theorem, it is enough to show that the following integral

$$
Y_H^T = Y_H^0 + \mu T + \int_{0}^{T} e^{V_H^r} dW^{H_Y}_r \tag{4.7}
$$
is well defined for all pairs of \((H^v, H^y) \in (0, 1) \times (0, 1)\) that satisfy \(H^v + H^y \geq 1\). It is then equivalent to show that

\[
\int_0^T e^{V^H_t} dW^H_t
\]

is well defined for all pairs of \((H^v, H^y) \in (0, 1) \times (0, 1)\) that satisfy \(H^v + H^y \geq 1\).

According to Øksendal (2003), the p-variation of a process \(X_t\) on a time interval \([0, T]\) with \(p > 0\) is defined as

\[
V_p(X_t, 0, T, \omega) = \lim_{|\Pi| \to 0} \sum_{\Pi} |X_{t_{k+1}}(\omega) - X_{t_k}(\omega)|^p = \lim_{|\Pi| \to 0} \sum_{\Pi} |X_{t_{k+1}} - X_{t_k}|^p,
\]

where \(\Pi\) is a partition on the time interval \([0, T]\).

Further, given \(\omega\) offering the solution \(V^H_t\) a continuous function of time \(t\), \(V^H_t\) would be bounded on the time interval \([0, T]\). We denote the lower and higher bound of \(V^H_t\) as \(V^H_t > 0\) and \(V^H_t > 0\). By the mean value theorem, we then have

\[
e^{V^H_{t_i}} V_p \left( V^H_t, 0, T, \omega \right) \leq V_p \left( e^{V^H_{t_i}}, 0, T, \omega \right) \leq e^{V^H_{t_h}} V_p \left( V^H_t, 0, T, \omega \right).
\]

Therefore, the feature of p-variation of \(e^{V^H_{t_i}}\) which can be thought as a function of \(t\), is same as \(V^H_t\). (In other words, the p-variation of \(e^{V^H_{t_i}}\) is finite or infinite, if and only if the p-variation of \(V^H_t\) is finite or infinite.)

Similarly, but using the mean value theorem for integration, we can have that the feature of p-variation of \(e^{-bt}V^H_t\) is same as \(W^H_t\) – the random driver of \(V^H_t\), since the p-variation of \(W^H_t\) finite if and only if \(q \geq \frac{1}{2p}\), we have a finite p-variation of \(e^{-bt}V^H_t\) if and only if \(p \geq \frac{1}{2q}\).

Further, if \(p \geq \frac{1}{2q} \geq 1\), we can show that
\[ V_p (V^H_t, 0, T, \omega) = \lim_{|\Pi| \to 0} \sum \prod \left| V^H_{t_{k+1}} - V^H_{t_k} \right|^p \]

\[ = \lim_{|\Pi| \to 0} \sum \prod \left| e^{bt_{k+1}} - e^{bt_k} \right| e^{-bt_{k+1}} V^H_{t_{k+1}} + e^{bt_k} (e^{-bt_{k+1}} V^H_{t_{k+1}} - e^{-bt_k} V^H_{t_k}) \right|^p \]

\[ \leq \text{a positive number} \times \left( 1 - e^{bT} \right) + \text{a positive number} \times V_p \left( e^{-bt} V^H_t, 0, T, \omega \right) < \infty. \]

Thereby, the support area of the equation (4.7) is determined by the well-definedness of

\[ \int_0^T V^H_t dW^H_t. \]

Following the p-variation condition for the well-definedness of the pathwise solution to a stochastic integral of fBm, we need the p-variation of \( V^H_t \) finite for all \( p > (1 - H^y)^{-1} \), and since the p-variation of \( V^H_t \) finite if \( p \geq \frac{1}{1 - H^y} \), we have

\[(1 - H^y)^{-1} \geq \frac{1}{H^v} \iff H^v + H^y \geq 1.\]

\section*{4.5.2 The joint behavior of \( Y^H_t \) and \( V^H_t \)}

Next, we discuss the memory structures of the volatility process described by the fractional OU equation and the corresponding price process.

**Proposition 4.5.5** For well-defined fractional LogOU dynamic model system

\[ dY^H_t = \mu dt + e^{V^H_t} dW^H_t, dV^H_t = bV^H_t dt + c dW^H_t, \]

where \( W^H_t \) and \( W^H_t \) are two independent fractional Brownian motions with their own self-similarity index being \( H^y \) and \( H^v \) and with \( \text{Var}(W^H_t) = \text{Var}(W^H_t) = 1 \), and \( b < 0 \), \( c > 0 \).
1. \( V_t^H \) has a short time memory if \( 0 < H^v \leq 0.5 \) and shows a long time memory if \( 0.5 < H^v < 1 \). (The incremental process of \( V_t^H \) always has a short time memory.)

2. The memory structure of \( \Delta Y_t^H := \mu \Delta + e^{V_t^H} \Delta W_t^{H^v} = \mu \Delta + e^{V_t^H} (W_{t+1}^{H^v} - W_t^{H^v}) \) depends on \( H^v \), i.e. a short time memory, a long time memory or no memory for \( 0 < H^v < 0.5, 0.5 < H^v < 1 \) or \( H^v = 0.5 \) respectively.

Proof.

1. For \( V_t^H \) process

   - if \( H^v = 0.5 \),
     \[
     \sum_{s=1}^{\infty} |\text{Cov}(V_t^H, V_{t+s}^H)| = \frac{e^2}{2b} \sum_{s=1}^{\infty} e^{2bs} < \infty.
     \]

   - if \( 0 < H^v < 0.5 \), by Theorem 4.5.2 (Theorem 2.3, Cheridito et al., 2003), there exists a \( \eta > 0 \) such that
     \[
     0 \leq \sum_{s=1}^{\infty} |\text{Cov}(V_t^H, V_{t+s}^H)| \leq \frac{1}{2} e^2 \sum_{s=1}^{\infty} \sum_{n=1}^{N} |b|^{-2n} \left| \prod_{k=0}^{2n-1} (2H^v - k) \right| s^{2H^v-2n} + \text{a finite term}
     \]
     \[
     \leq \frac{1}{2} e^2 \sum_{n=1}^{N} |b|^{-2n} \left| \prod_{k=0}^{2n-1} (2H^v - k) \right| \sum_{s=1}^{\infty} s^{-(1+n)} + \text{a finite term} < \infty.
     \]

   - if \( 0.5 < H^v < 1 \), by Theorem 4.5.2 (Theorem 2.3, Cheridito et al., 2003),
     \[
     \sum_{s=1}^{\infty} |\text{Cov}(V_t^H, V_{t+s}^H)| \geq \frac{1}{2} e^2 \sum_{s=1}^{\infty} |b|^{-2} |2H^v (2H^v - 1)| s^{2H^v-2} + \text{a finite term} = \infty.
     \]
As to the incremental process of $V_t^H$, we use $\delta$ for time lag and define $[\delta]$ as the smallest integer greater than $\delta$. Then, we have

$$
\sum_{s=1}^{\infty} |\text{Cov}(\Delta V_t^H, \Delta V_{t+s}^H)|
= \sum_{s=[\delta]+1}^{\infty} \left[ \frac{1}{2} c^2 \sum_{n=1}^{N} |b|^{2n} \prod_{k=0}^{2n-1} (2H^v - k) \right] \sum_{s=[\delta]+1}^{\infty} |2s^{2H^v-2n} - (s + \delta)^{2H^v-2n} - (s - \delta)^{2H^v-2n}| + \text{a finite term}
$$

Thus, the incremental process of $V_t^H$ always has short memory.

2. We then consider the memory structure of the price process in which the random
driver is a fBm and it is easy to check that

- By the equation (4.6),
  \[ \text{Var}(V_H^t) = \text{Var}(V_H^0) = c^2 \frac{\Gamma(2H^v + 1) \sin(\pi H^v)}{\pi} \int_0^\infty \frac{|x|^{1-2H^v}}{|b|^2 + x^2} dx < \infty. \]

- Still by the equation (4.6) and by the form of the moment generating function of a multivariate normal distribution,
  \[ E\left( e^{V_H^t + V_H^{t+s}} \right) = e^{Var(V_H^0) + \text{Cov}(V_H^t, V_H^{t+s})} \]
  \[ = e^{c^2 \frac{\Gamma(2H^v + 1) \sin(\pi H^v)}{\pi} \int_0^\infty \frac{|1 + \cos(sx)|}{|b|^2 + x^2} |x|^{1-2H^v} dx} < \infty. \]

- To approximate the incremental price process, we therefore define \( \Delta Y_H^t := \mu \delta + e^{V_H^t} \Delta W_H^y, \) \( = \mu \delta + e^{V_H^t} (W_{t+\delta}^H - W_t^H) \) and by the independence between \( \{W_t^H\}_{t \in \mathbb{R}^+} \) and \( \{W_t^{Hy}\}_{t \in \mathbb{R}^+} \),
  \[ \text{Cov}(\Delta Y_H^t, \Delta Y_H^{t+s}) = \text{Cov}(e^{V_H^t} \Delta W_H^y, e^{V_H^{t+s}} \Delta W_H^{y}) \]
  \[ = E(\Delta W_H^y \Delta W_H^{y}) E\left( e^{V_H^t + V_H^{t+s}} \right) \]
  \[ = e^{c^2 \frac{\Gamma(2H^v + 1) \sin(\pi H^v)}{\pi} \int_0^\infty \frac{|1 + \cos(sx)|}{|b|^2 + x^2} |x|^{1-2H^v} dx} E(\Delta W_H^y \Delta W_H^{y}). \] \( (4.8) \)

Consequently,
- if \( H^y = 0.5 \) and \( s \neq 0 \),
  \[ \text{Cov}(\Delta Y_H^t, \Delta Y_H^{t+s}) = 0. \]
- if \( 0.5 < H^y < 1 \), we have \( \cos(sx) \geq -1 \) and hence
  \[ |\text{Cov}(\Delta Y_H^t, \Delta Y_H^{t+s})| \geq |E(\Delta W_H^y \Delta W_H^{y})|. \]
• if $0 < H^y < 0.5$, we have $\cos(sx) \leq 1$ and hence

$$|Cov(\Delta Y_t^H, \Delta Y_{t+s}^H)| \leq e^{2\text{Var}(V_0)} |E(\Delta W_t^{H^y} \Delta W_{t+s}^{H^y})|. $$

We thereby obtain that the memory structure in the incremental process of $Y_t$ is determined by $H^y$.

\section{4.6 Simulation Study of Fractional LogOU}

Besides the autocovariance function (ACVF), when judging the memory structure underlying a process, scientists can use another method – the Hurst exponent. Moreover, if a process is a fractional Brownian motion which is a particular self-similar process, the Hurst exponent (the self-similar index of this process) can be estimated in many other ways. Below we use the R/S analysis (RRS) and the regression estimator (REG) invented by Taqqu et al. (1995). We use the R/S analysis (RRS) because this method is applicable to all the processes, and choose the regression estimator (REG) because for fractional Brownian motions, the method performs well in the sense of bias, mean squared errors and out of range estimates. However, when we use the regression estimator, we have two conjectures here.

• REG is applicable to fractional Brownian motions. We conjecture that it is also applicable for all the processes with stationary increments.

• In the simulation study of the estimator for the Hurst exponent for fractional Brownian motions, REG shows less bias than RRS. We conjecture that when REG is applicable to the process, it always shows less bias than RRS for the estimated Hurst exponent.
As to the stochastic differential equation (SDE) with respect to Brownian motion (BM), Kloeden et al. (1994) discussed many different simulation schemes. However, to extend these schemes to the fractional cases is not straightforward. Considering our problems, we should focus on the strong schemes which provide reliable trajectories for solutions from the corresponding SDE, since our concern is on the memory properties from process trajectories. Among the strong schemes, we decide to use the simple one – Euler scheme for the simulations, which is stated in details below and is easy to be extended. It turns out that Euler scheme serves our concern well for the fractional LogOU processes or the fractional OU process. Besides this simple scheme – Euler scheme, Milstein scheme is another strong one but having a higher convergent order than Euler, since it includes an additional term obtained from the Ito-Taylor expansion of SDE with respect to regular BM. By including more terms from the Ito-Taylor expansion, higher order strong schemes called as (explicit) strong Taylor scheme are then developed. One disadvantage of these Taylor type schemes is that the coefficients before both the deterministic and the random increments need to be derived for different orders. Therefore, strong schemes with deterministic settings are developed. Besides, when the solutions of SDE are stiff, implicit schemes which have a wide range of step sizes, would be suitable for the trajectory approximation, but these schemes need to solve additional algebraic equation at each time step, e.g. implicit Milstein scheme, implicit strong Taylor schemes, implicit strong Rung-Kutta schemes.

To generate data from fractional LogOU models approximately, we follow the first Euler scheme, and discretize the stochastic integrals as

\[
Y_{t+\delta}^H - Y_t^H = \mu \delta + e^{V_t^H} (W_{t+\delta}^{H^\nu} - W_t^{H^\nu}),
\]

\[
V_{t+\delta}^H - V_t^H = bV_t^H \delta + c(W_{t+\delta}^{H^\nu} - W_t^{H^\nu}), \text{ for some } \delta > 0.
\]
To obtain $W_{t+\delta}^{H_y} - W_t^{H_y}$ and $W_{t+\delta}^{H_v} - W_t^{H_v}$, we begin with a fractional Brownian motion with time lag as 1, and then by using the self-similar property of a fractional Brownian motion, for some $H \in (0,1)$, we argue that

$$
Cov\left(W_{s_{\delta}}^H - W_{(s-1)\delta}^H, W_{t_{\delta}}^H - W_{(t-1)\delta}^H\right)
= \delta^{2H} Cov\left(W_{s_{\delta}}^H - W_{s_{\delta}-1}^H, W_{t_{\delta}}^H - W_{t_{\delta}-1}^H\right)
= Cov\left(\delta^H (W_{s_{\delta}}^H - W_{s_{\delta}-1}^H), \delta^H (W_{t_{\delta}}^H - W_{t_{\delta}-1}^H)\right),
$$

for some $\delta > 0$. Therefore, based on this argument, we are able to obtain a Gaussian process with time lag $\delta$ by multiplying $\delta^H$ in front of the original Gaussian process. The incremental process of this new fractional Brownian motion process observed at time $\delta, 2\delta, \cdots$, should still be a standard normal distribution, if the time lag of this incremental process is 1. We check this in Figure 4.2, in the upper and down panel of which, the red dashed line is the density function curve of a standard normal distribution, and the histogram is for the incremental process with the time lag 1 of the fractional Brownian motion process observed at time $\delta, 2\delta, \cdots$. $\delta$ here is set to be 0.01.

In addition, when preparing data, we also notice that a trade-off exists between the length and the replicates of processes, given the similar amount of computing time and given that the concern is the accuracy of the autocovariance function (ACVF) of LogOU processes. By roughly examining the ACVF of fractional LogOU processes generated from several settings of the length and the replicates, we finally decide on five fractional LogOU processes with the length equal to 100,000, each with $\mu = 0.05$, $b = -1$ and $c = 1$ for $H_y = 0.2 & H_v = 0.2, H_y = 0.5 & H_v = 0.2, H_y = 0.8 & H_v = 0.2, H_y = 0.2 & H_v = 0.5, H_y = 0.5 & H_v = 0.5, H_y = 0.8 & H_v = 0.5, H_y = 0.2 & H_v = 0.8, H_y = 0.5 & H_v = 0.8$ and $H_y = 0.8 & H_v = 0.8$ respectively.
Figure 4.2  Check on the random drivers
4.6.1 Simulation study of $V_t^H$

When simulating $V_t^H$, $\delta/100$ was used for the discrete approximation. Later, based on the fractional OU processes simulated by Euler scheme, we calculate the sample ACVF and the exact ACVF given by Theorem 4.5.2 (Theorem 2.3, Cheridito et al., 2003), and compare them in Figure 4.3. In this figure, the solid and the dashed line represent respectively the sample ACVF and the exact/theoretical ACVF values, and different colors of the lines represent different $H^v$ levels in the random drivers of the fractional OU processes. We
then have confidence that the simulated processes exhibit the characteristics of ACVF in the fractional OU processes.

### 4.6.1.1 Estimates for the entire sequence

The regression estimator (REG) and R/S statistic (R/S) are used to estimate the Hurst exponent in each of the five replications for the entire series. For REG and R/S estimators, we follow the procedures below to produce the results in Table 4.1.

1. Divide one simulated sequence into blocks with the length equal to $m$. Use $j$ to index the blocks (with block size $m$), then $j = 1, \ldots, \lfloor L/m \rfloor$. For $j^{th}$ block, calculate the statistic (denote as $S_m$) corresponding to the estimator, and then, average the statistic over blocks (denote as $\bar{S}_m$).

2. Repeat the above procedure for different values of $m$. (In my computation, $m$ is set to be from 24 to 1000.)

3. Do the regression of the logarithm of the averaged statistic (i.e. $\log(\bar{S}_m)$) vs. $\log(m)$ to derive the estimate of the Hurst exponent for the entire simulated sequence.

<table>
<thead>
<tr>
<th>$H^\nu$</th>
<th>1st REP</th>
<th>2nd REP</th>
<th>3rd REP</th>
<th>4th REP</th>
<th>5th REP</th>
<th>Mean</th>
<th>Sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2, REG</td>
<td>0.436</td>
<td>0.428</td>
<td>0.446</td>
<td>0.438</td>
<td>0.429</td>
<td>0.435</td>
<td>0.008</td>
</tr>
<tr>
<td>0.2, RRS</td>
<td>0.478</td>
<td>0.472</td>
<td>0.484</td>
<td>0.474</td>
<td>0.470</td>
<td>0.476</td>
<td>0.005</td>
</tr>
<tr>
<td>0.5, REG</td>
<td>0.501</td>
<td>0.518</td>
<td>0.520</td>
<td>0.514</td>
<td>0.507</td>
<td>0.512</td>
<td>0.008</td>
</tr>
<tr>
<td>0.5, RRS</td>
<td>0.549</td>
<td>0.558</td>
<td>0.558</td>
<td>0.556</td>
<td>0.552</td>
<td>0.555</td>
<td>0.004</td>
</tr>
<tr>
<td>0.8, REG</td>
<td>0.793</td>
<td>0.817</td>
<td>0.806</td>
<td>0.818</td>
<td>0.802</td>
<td>0.807</td>
<td>0.010</td>
</tr>
<tr>
<td>0.8, RRS</td>
<td>0.778</td>
<td>0.795</td>
<td>0.785</td>
<td>0.793</td>
<td>0.781</td>
<td>0.786</td>
<td>0.007</td>
</tr>
</tbody>
</table>

For the fractional OU processes, Table 4.1 provides the estimates of the entire series for the Hurst exponent, and these numerical results are consistent to the ACVF memory
analysis in Proposition 4.1. In other words, when $H^v$ is no more than 0.5, the estimates tend to be around or below 0.5, and when $H^v$ is larger than 0.5, the estimates are greater than 0.5.

### 4.6.1.2 Block effect

We then follow Peters (1996) to study the effect of the maximum block size on the estimates of the Hurst exponent, and this effect is called as block effect. The results for the maximum block size values taken on 100, 200, ..., 1000, are shown in Figure 4.4, of which the x-axis is the maximum block size, and y-axis is the estimate of the Hurst exponent.

![Figure 4.4 Block effect on the estimates of the Hurst exponent.](image)

This figure implies that the Hurst exponent changes around a certain value (not necessarily to be the $H^v$ value), and shows no block effect in the fractional OU processes.
4.6.1.3 Time effect

We are also curious about the effect of different time pieces of a process on the estimates of the Hurst exponent. Therefore, we divide each simulated series into 10 “segments/blocks” (i.e. 10,000 obs in each segment/block), and within each segment/block, the Hurst exponent is estimated as in Section 4.6.1.1 with \( m \) from 24 to 1000, but \( \bar{S}_m \) is averaged over blocks and replications. The results are in Figure 4.5, of which the x-axis is the index of the segments/blocks (Blk), and y-axis is the estimate of the Hurst exponent, and they tell us that the simulated fractional OU processes have no time effect.

![Diagram](image)

Figure 4.5 The estimates of the Hurst exponent in 10 segments.

4.6.2 Simulation study of \( Y_t^H \) increments

According to (4.8), we calculate the theoretical ACVF curves for the incremental process of \( Y_t^H \) with the incremental length being \( \delta^* = 0.01 \) for \( b = -1 \) and \( c = 1 \), given \( H^y = 0.2 \& \)
\[ H^v = 0.2, \ H^y = 0.8 \ \& \ H^v = 0.2, \ H^y = 0.2 \ \& \ H^v = 0.5, \ H^y = 0.8 \ \& \ H^v = 0.5, \ H^y = 0.2 \ \& \n\]

\[ H^v = 0.8, \ \& \ H^y = 0.8 \ \& \ H^v = 0.8. \]

Figure 4.6 shows that these theoretical ACVF curves are consistent to the memory properties of the \( Y_t^H \) increments that we have proved. In other words, the memory properties of the \( Y_t^H \) increments are determined by the random driver in the \( Y_t^H \) equation.

![Theoretical ACVF of \( Y_t \) increments](image.png)

**Figure 4.6** Theoretical curves

We then study the ACVF of \( Y_t^H \) in two natural time scales: one from 0 to 1 and the other from 0 to 10. We simulated five pairs of fractional LogOU processes (\( Y_t^H \) and \( V_t^H \)) for a given pair of \( H^y \) and \( H^v \). Each \( Y_t^H \) or \( V_t^H \) has 100,000 observations, with \( \mu = 0.05, \ b = -1 \) and \( c = 1 \) for \( H^y = 0.2 \ \& \ H^v = 0.2, \ H^y = 0.5 \ \& \ H^v = 0.2, \ H^y = 0.8 \ \& \ H^v = 0.2, \ H^y = 0.2 \ \& \ H^v = 0.8, \ 

\[ H^v = 0.8, \ \& \ H^y = 0.8 \ \& \ H^v = 0.8. \]
\(H^y = 0.2 \& H^v = 0.5, H^y = 0.5 \& H^v = 0.5, H^y = 0.8 \& H^v = 0.5, H^y = 0.2 \& H^v = 0.8, H^y = 0.5 \& H^v = 0.8 \& H^y = 0.8 \& H^v = 0.8\) respectively. The recorded time lag (or step size), denoted by \(\delta\), is set to be 0.01 as well as the length of the increments \(\delta^*\). The lags of ACVF are then computed at 0.01, 0.02, \(\cdots\), 100 \times 0.01 = 1 in Figure 4.7 - 4.9, while the lags are at 0.1, 0.2, \(\cdots\), 1000 \times 0.01 = 10 in Figure 4.10 - 4.12. The upper panel in each figure includes ACVF curves for both the \(V_t^H\) and the \(Y_t^H\) increments for three pairs of \(H^y\) and \(H^v\) values, and the down panel shows ACVF for the \(Y_t^H\) increments only and with a smaller scale in the y-axis than the upper panel.

Figure 4.7 ACVF on the real time interval \((0, 1)\) for \(H^v = 0.2\)

### 4.6.2.1 Discussion on the simulation results for ACVF

Since in the simulation, \(Y_t^H\) has fixed time lag \(\delta\) and increment length \(\delta^*\) being a multiple of \(\delta\), we denote the ACVF of \(Y_t^H\) increments as \(\gamma(k\delta; \delta^* = k_0\delta, H^v, H^v)\) for the lag value \(k\delta\) with \(k, k_0 \in \mathbb{Z}^+\). We can show the following proposition for the theoretical
Figure 4.8 ACVF on the real time interval (0, 1) for $H^v = 0.5$

Figure 4.9 ACVF on the real time interval (0, 1) for $H^v = 0.8$
Figure 4.10  ACVF on the real time interval $(0, 10)$ for $H^v = 0.2$

Figure 4.11  ACVF on the real time interval $(0, 10)$ for $H^v = 0.5$
ACVF curves of the increments with a certain $\delta$.

**Proposition 4.6.1** Given $H^y$, $H^v$, $k_0$ and $k$, when $\delta$ increases, $|\gamma(k\delta; k_0\delta, H^y, H^v)|$ increases.

**Proof.** Recall that, for $k \in \mathbb{R}^+$,

\[
\gamma(k\delta; k_0\delta, H^y, H^v) = Cov\left( \Delta Y^H_0, \Delta Y^H_{k\delta} \right)
\]

\[
= Cov\left( e^{V^H_0 \Delta W^{H_y}_0}, e^{V^H_{k\delta} \Delta W^{H_y}_{k\delta}} \right)
\]

\[
= E\left( \Delta W^{H_y}_0 \Delta W^{H_y}_{k\delta} \right) E\left( e^{V^H_{k\delta} + V^H_0} \right)
\]

\[
= E\left( \Delta W^{H_y}_0 \Delta W^{H_y}_{k\delta} \right) e^{\text{Var}(V^H_0) + Cov(V^H_{k\delta}, V^H_0)}
\]

\[
= e^{\frac{\Gamma(2H^v + 1) \sin(H^v)}{\Gamma(2H^v + 2)}} \int_0^{\infty} (1 + \cos(k\delta x)) e^{\frac{|x|^{1-2H^v}}{|x|^2 + x^2}} dx
\]

\[
= 0.5e^{\frac{\Gamma(2H^v + 1) \sin(H^v)}{\Gamma(2H^v + 2)}} \int_0^{\infty} (1 + \cos(k\delta x)) e^{\frac{|x|^{1-2H^v}}{|x|^2 + x^2}} dx
\]

\[
\times \left( |k\delta + k_0\delta|^{2H^v} + |k\delta - k_0\delta|^{2H^v} - 2|k\delta|^{2H^v} \right).
\]
For $\delta_1 < \delta_2$,

$$0 \leq \frac{\gamma(k\delta_1; k_0\delta_1, H^y, H^v)}{\gamma(k\delta_2; k_0\delta_2, H^y, H^v)} = \frac{\delta_1^{H_y}}{\delta_2^{H_y}} < 1.$$

In Figure 4.6, we observed that

**Proposition 4.6.2** Given $\delta$, $H^y$, $k_0$ and $k$, when $H^v$ increases, $|\gamma(k\delta; k_0\delta, H^y, H^v)|$ increases.

**Proof.** Given the values of $\delta$, $H^y$, $k_0$ and $k$,

$$|\gamma(k\delta; k_0\delta, H^y, H^v)| = a \text{ positive constant} \times e^{Var(V_0^H) + Cov(V_{k\delta}^H, V_0^H)}$$

From Figure 4.3, we observe the fact that $|Var(V_0^H)|$ and $|Cov(V_{k\delta}^H, V_0^H)|$ increase, as $H^v$ increases, for $k \in \mathbb{Z}^+$, then it is obvious that $|\gamma(k\delta; k_0\delta, H^y, H^v)|$ increases as $H^v$ increases.

From Figure 4.7 to Figure 4.12, we observed the following things for ACVF of the $Y_t^H$ increments.

- When $H^y = 0.8$, the ACVF curve of the $Y_t^H$ increments takes on the positive values and converges to zero. When $H^y = 0.2$, the ACVF curve converges to zero from the negative values, while $H^y = 0.5$, the ACVF curve is almost on the x-axis all the time. The proposition below explains this phenomenon.

**Proposition 4.6.3** Given $\delta$ and $H^v$, when $H^y > 0.5$, $H^y < 0.5$ and $H^y = 0.5$, we have $\gamma(k\delta; k_0\delta, H^y, H^v) > 0$, $\gamma(k\delta; k_0\delta, H^y, H^v) < 0$, $\gamma(k\delta; k_0\delta, H^y, H^v) = 0$ respectively.
Proof.

\[ \gamma(k\delta; k_0\delta, H^y, H^v) \]

\[ = E(\Delta W_0^{H^y}\Delta W_{k\delta}^{H^v}) e^{V_{\text{var}}(V_0^{H^y})+\text{Cov}(V_{k\delta}^{H^y}, V_0^{H^v})} \]

\[ = e^{\frac{1}{2}(2H^v+1)} \int_0^{\infty} (1+\cos(k\delta x)) \frac{\psi^1-2H^v}{|x|^{2H^v+2}} dx \]

\[ = a \text{ positive function of } k\delta \times E(\Delta W_0^{H^y}\Delta W_{k\delta}^{H^v}) . \]

Therefore, when \( H^y > 0.5 \), \( H^y < 0.5 \) and \( H^y = 0.5 \), we have \( E(\Delta W_0^{H^y}\Delta W_{k\delta}^{H^v}) \) positive, negative and equal to 0 respectively, and then similar to \( \gamma(k\delta; k_0\delta, H^y, H^v) \).

Remark 4.6.4 Here, we also discuss the convergent speed for different \( H^y \) values. When \( k\delta > 0 \) is large, we know that \( E(\Delta W_0^{H^y}\Delta W_{k\delta}^{H^v}) \sim O\left((k\delta)^{2H^y-2}\right) \) and \( \text{Cov}(V_{k\delta}^{H^y}, V_0^{H^v}) \sim O\left((k\delta)^{2H^y-2}\right) \). Thereby, \( \gamma(k\delta; k_0\delta, H^y, H^v) \sim O\left((k\delta)^{2H^y-2}\right) \). In other words, the convergent speed is mainly determined by the \( Y_t^H \) random driver process, or \( H^y \), i.e. the larger the \( H^y \) is, the more slowly the process converges.

- The ACVF curve of the \( Y_t^H \) increments for \( H^v = 0.8 \) tends to have a larger magnitude in fluctuations than \( H^v = 0.2 \).

Argument: We recall that in the theoretical expression of ACVF for the \( Y_t^H \) increments, the part including \( H^v \) is the power of the exponential function, which is in the form of ACVF of \( V_t^H \). In the simulation, we also notice that the sample ACVF of \( V_t^H \) averaged over 5 samples, tends to have larger variance with \( H^v = 0.8 \) than with \( H^v = 0.5 \) or \( H^v = 0.2 \), which may cause the larger magnitude in fluctuations for the ACVF curve of the \( Y_t^H \) increments for \( H^v = 0.8 \). As to the reason of that the sample ACVF of \( V_t^H \) averaged over 5 samples has larger variance with \( H^v = 0.8 \)
than with $H^y = 0.5$ or $H^v = 0.2$, it could be a research interest but currently is not the focus of our study.

### 4.6.2.2 Discussion on $Y_t^H$ increments

Moreover, we simulate data more carefully to study $H$. The data are still five pairs of fractional LogOU processes ($Y_t^H$ and $V_t^H$) for a given pair of $H^y$ and $H^v$. The recorded time lag (or step size) is again set to be $\delta = 0.01$, but more considerately, $\delta/100$ is used for the discrete approximation. Therefore, to have similar amount of computational work, each $Y_t^H$ or $V_t^H$ currently has 100,000 observations with $\mu = 0.05$, $b = -1$ and $c = 1$ for $H^y = 0.2 \& H^v = 0.2$, $H^y = 0.5 \& H^v = 0.2$, $H^y = 0.8 \& H^v = 0.2$, $H^y = 0.2 \& H^v = 0.5$, $H^y = 0.5 \& H^v = 0.5$, $H^y = 0.8 \& H^v = 0.5$, $H^y = 0.2 \& H^v = 0.8$, $H^y = 0.5 \& H^v = 0.8$ and $H^y = 0.8 \& H^v = 0.8$ respectively. The incremental length ($\delta^*$) of $Y_t^H$ is still 0.01, and the actual time length for the simulated fractional OU processes is then equal to 1000.

In Table 4.2, we see that whether $H$ estimates for the $Y_t^H$ increments are greater than, around or smaller than 0.5, is determined by $H^y$ value in the random drivers of $Y_t^H$ processes. This notice is consistent to what we have proven for the $Y_t^H$ increments. Figure 4.13 and Figure 4.14 show that there are no block or time effects in the $Y_t^H$ increments. However, we realize that the $H$ estimates are neither the $H^y$ value nor the $H^v$ value. Further, the relationship between the $H$ estimates and the $H^y$ value or the $H^v$ value has not been studied yet and could be a research topic, but currently not the focus of this study.
Figure 4.13 Block effect for the $Y^H_\ell$ increments
Figure 4.14  Time effect for the $Y_t^H$ increments
Table 4.2  Estimates of the Hurst exponent for the $Y^H_t$ incremental process.

<table>
<thead>
<tr>
<th>$H^v$</th>
<th>EST</th>
<th>1st REP</th>
<th>2nd REP</th>
<th>3rd REP</th>
<th>4th REP</th>
<th>5th REP</th>
<th>MEAN</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>REG</td>
<td>0.423</td>
<td>0.428</td>
<td>0.437</td>
<td>0.430</td>
<td>0.426</td>
<td>0.429</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>RRS</td>
<td>0.457</td>
<td>0.461</td>
<td>0.457</td>
<td>0.454</td>
<td>0.462</td>
<td>0.458</td>
<td>0.003</td>
</tr>
<tr>
<td>0.5</td>
<td>REG</td>
<td>0.485</td>
<td>0.503</td>
<td>0.515</td>
<td>0.485</td>
<td>0.488</td>
<td>0.495</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>RRS</td>
<td>0.528</td>
<td>0.531</td>
<td>0.540</td>
<td>0.530</td>
<td>0.525</td>
<td>0.531</td>
<td>0.006</td>
</tr>
<tr>
<td>0.8</td>
<td>REG</td>
<td>0.748</td>
<td>0.750</td>
<td>0.764</td>
<td>0.780</td>
<td>0.781</td>
<td>0.765</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>RRS</td>
<td>0.754</td>
<td>0.752</td>
<td>0.759</td>
<td>0.759</td>
<td>0.772</td>
<td>0.759</td>
<td>0.008</td>
</tr>
<tr>
<td>0.2</td>
<td>REG</td>
<td>0.478</td>
<td>0.481</td>
<td>0.497</td>
<td>0.489</td>
<td>0.493</td>
<td>0.491</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>RRS</td>
<td>0.504</td>
<td>0.513</td>
<td>0.506</td>
<td>0.531</td>
<td>0.494</td>
<td>0.509</td>
<td>0.014</td>
</tr>
<tr>
<td>0.5</td>
<td>REG</td>
<td>0.497</td>
<td>0.504</td>
<td>0.506</td>
<td>0.506</td>
<td>0.500</td>
<td>0.502</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>RRS</td>
<td>0.540</td>
<td>0.544</td>
<td>0.553</td>
<td>0.550</td>
<td>0.545</td>
<td>0.542</td>
<td>0.006</td>
</tr>
<tr>
<td>0.8</td>
<td>REG</td>
<td>0.788</td>
<td>0.798</td>
<td>0.796</td>
<td>0.823</td>
<td>0.794</td>
<td>0.790</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>RRS</td>
<td>0.770</td>
<td>0.770</td>
<td>0.760</td>
<td>0.780</td>
<td>0.764</td>
<td>0.769</td>
<td>0.007</td>
</tr>
</tbody>
</table>

4.7  Fractional Cox-Ingersoll-Ross (CIR) Model

4.7.1  The well-definedness of the fractional CIR model system

The solution to the CIR model with the random driver in $V_t$ being a fBm, has not been discussed extensively in literature. Below, we follow the idea of solving the CIR model with a BM as the random driver, and see how further we can get for the CIR with a fBm. A pathwise solution of CIR model can be generated as follows (Shreve, 2004).

Define

$$y_t = \begin{pmatrix} X_{1,t} \\ \vdots \\ X_{d,t} \end{pmatrix}, y_0 = \begin{pmatrix} X_{1,0} \\ \vdots \\ X_{d,0} \end{pmatrix}, t_0 = 0, A = \frac{b}{2}, f(t) = \frac{c}{2}, W_t = \begin{pmatrix} dW_{1,t} \\ \vdots \\ dW_{d,t} \end{pmatrix},$$

for $b < 0, c > 0$, and $d \geq 1$, and then, the Ornstein-Uhlenbeck (OU) functions

$$dX_{j,t} = \frac{b}{2} X_{j,t} dt + \frac{c}{2} dW_{j,t}, j = 1, \cdots, d$$
can be expressed as
\[ dy_t = Ay_t dt + f(t) dW_t, \]
and the solution in the vector form should be
\[ y_t = e^{(t-t_0)A} y_0 + \int_{t_0}^{t} e^{(t-s)A} f(s) dW_s. \]

In other words, the solution is
\[
\begin{pmatrix}
X_{1,t} \\
\vdots \\
X_{d,t}
\end{pmatrix}
= e^{\frac{bt}{2}}
\begin{pmatrix}
X_{1,0} \\
\vdots \\
X_{d,0}
\end{pmatrix}
+ \frac{c}{2}
\begin{pmatrix}
\int_{0}^{t} e^{(t-u)b/2} dW_{1,u} \\
\vdots \\
\int_{0}^{t} e^{(t-u)b/2} dW_{d,u}
\end{pmatrix},
\]

based on which, people define
\[
\begin{align*}
\begin{cases}
\quad a = dc^2/4 \\
\quad V_t = \sum_{j=1}^{d} X_{j,t}^2 \\
\quad W_t = \sum_{j=1}^{d} \int_{0}^{t} \frac{X_{j,s}}{\sqrt{V_s}} dW_{j,s}
\end{cases}
\quad (4.9)
\end{align*}
\]

and, by the martingale property of BM and Lévy’s Theorem, check that \( W_t \) is still a Brownian motion and \( V_t \) is the pathwise solution for the CIR model with this BM as the random driver (\( dV_t = (a + bV_t) dt + c\sqrt{V_t} dW_t \)).

If the random driver is fBm in the volatility equation, to get the solution to \( d \)-dimensional fractional OU functions in the similar way, it would be
\[
\begin{pmatrix}
X_{1,t}^H \\
\vdots \\
X_{d,t}^H
\end{pmatrix}
= e^{\frac{bt}{2}}
\begin{pmatrix}
X_{1,0}^H \\
\vdots \\
X_{d,0}^H
\end{pmatrix}
+ \frac{c}{2}
\begin{pmatrix}
\int_{0}^{t} e^{(t-u)b/2} dW_{1,u}^H \\
\vdots \\
\int_{0}^{t} e^{(t-u)b/2} dW_{d,u}^H
\end{pmatrix}
\]
Now, if we still want to use $V_t^H = \sum_{j=1}^d \left( X_{j,t}^H \right)^2$ as the solution, we need to check that
\[
\sum_{j=1}^d \int_0^t \frac{X_{j,s}^H}{\sqrt{V_{j,s}^H}} dW_{j,s}^H
\]
is a fBm, which becomes difficult since fBm is usually not a martingale.

The form of the CIR solution with fBm as the random driver is then unclear, and later, we focus on the price equation with a fractional Brownian motion as the random driver and the volatility equation with a regular BM as the random driver by following the notations in the equation (4.5).

**Theorem 4.7.1** Define fractional CIR dynamic model system as

\[
dY_t^H = \mu dt + \sqrt{V_t} dW_t^{H^y}, dV_t = \kappa(\theta - V_t) dt + \sigma_v \sqrt{V_t} dW_t,
\]

where $W_t^{H^y}$ is a fBm with index $H^y$ and $\text{Var} \left( W_t^{H^y} \right) = 1$, $W_t$ is a standard BM independent of $W_t^{H^y}$, and $\kappa > 0$, $\theta < 0$, $\sigma_v > 0$. Then, this dynamic model system is well defined for $H^y \in [0.5, 1)$.

**Proof.** The crucial part of the proof for this theorem is still the price equation. Further, similar to Theorem 4.5.4, after reparameterizing $dV_t = \kappa(\theta - V_t) dt + \sigma_v \sqrt{V_t} dW_t$ into $dV_t = (a + bV_t) dt + c \sqrt{V_t} dW_t$ for $a = \kappa \theta$, $b = -\kappa$ and $c = \sigma_v$, and by the equation (4.9) and the mean value theorem in several variables, the pathwise solution of CIR equation ($V_t$) obtained in the corresponding way has the p-variation featured similarly to the p-variation of the solutions of the OU functions, and thus has a finite p-variation if $p \geq \frac{1}{0.5} = 2$.

Following the p-variation condition for well-definedness of pathwise solution of stochastic integral with respect to a fractional Brownian motion, we have $(1 - H^y)^{-1} \geq 2$ and thus obtain the theorem. ■
4.7.2 The joint behavior of $Y_t^H$ and $V_t$

To understand the joint behavior in the memory structure of the pricing and volatility processes in CIR model, we review some existing results (Cox et al., 1985a, 1985b), and then provide a proposition for the memory features of $Y_t^H$ and $V_t$.

Following the commonly used notations for CIR model, the fractional Cox-Ingersoll-Ross (CIR) model system can be written as

$$
\begin{align*}
    dY_t^H &= \mu dt + \sqrt{V_t} dW_t^H, \\
    dV_t &= \kappa(\theta - V_t) dt + \sigma_v \sqrt{V_t} dW_t,
\end{align*}
$$

where $W_t^H$ is a fBm with index $H$ and $\text{Var}(W_t^H) = 1$, $W_t$ is a standard BM independent of $W_t^H$, and $\kappa > 0, \theta < 0, \sigma_v > 0$.

**Fact I**: (Cox et al.,1985b) The distribution of $V_s$ given $V_t$ for $s > t$ is a non-central chi-square,

$$
2cV_s|V_t \sim \chi^2(2q + 2, 2u),
$$

for $c = 2\kappa/\sigma_v^2(1 - e^{-\kappa(s-t)})$, non-centrality parameter $u = cV_te^{-\kappa(s-t)}$, and degrees of freedom $q = \frac{2\theta}{\sigma_v^2} - 1$. As $s$ becomes large, $V_s|V_t$ goes to a Gamma distribution with parameters $(\frac{2\theta}{\sigma_v^2}, \frac{2\kappa}{\sigma_v^2})$. The stationary marginal distribution of the $V_s$ process is the same Gamma distribution with mean $\theta$ and variance $\frac{\theta\sigma_v^2}{2\kappa}$.

**Fact II**: (Cox et al.,1985b) Straightforward calculations to obtain the expected value and variance of $V_s$ given $V_t$, for $s \geq t$, show that:

$$
\begin{align*}
    E(V_s|V_t) &= V_te^{-\kappa(s-t)} + \theta(1 - e^{-\kappa(s-t)}), \\
    \text{Var}(V_s|V_t) &= V_t \frac{\sigma_v^2}{\kappa} \left( e^{-\kappa(s-t)} - e^{-2\kappa(s-t)} \right) + \theta \frac{\sigma_v^2}{2\kappa} \left( 1 - e^{-\kappa(s-t)} \right)^2.
\end{align*}
$$

(4.10)
If $t = 0$, then

\[
E(V_s|V_0) = V_0 e^{-\kappa s} + \theta(1 - e^{-\kappa s}),
\]
\[
\operatorname{Var}(V_s|V_0) = V_0 \frac{\sigma_v^2}{\kappa} (e^{-\kappa s} - e^{-2\kappa s}) + \theta \frac{\sigma_v^2}{2\kappa} (1 - e^{-\kappa s})^2
\]
\[
= \theta \frac{\sigma_v^2}{2\kappa} + \frac{\sigma_v^2}{\kappa} (V_0 - \theta) e^{-\kappa s} + \frac{\sigma_v^2}{\kappa} \left( \frac{\theta}{2} - V_0 \right) e^{-2\kappa s}.
\]

(4.11)

In the remainder, we still assume that $s > t$ and $s - t = h > 0$. Then, for the volatility process $(V_t)$ and the price process $(Y_t^H)$, we can derive the following proposition about their covariance structure and about their memory structure.

**Proposition 4.7.2** For well-defined fractional CIR dynamic model system,

\[
dY_t^H = \mu dt + \sqrt{V_t} dW_t^{H^y},
\]
\[
dV_t = \kappa(\theta - V_t) dt + \sigma_v \sqrt{V_t} dW_t,
\]

where $W_t^{H^y}$ is a fBm with index $H^y$ and $\operatorname{Var}(W_t^{H^y}) = 1$, $W_t$ is a standard BM independent of $W_t^{H^y}$, and $\kappa > 0$, $\theta < 0$, $\sigma_v > 0$, we have

1. $V_t$ process (and also its incremental process) has a short time memory.

2. the memory structure of $\Delta Y_t^H := \mu \Delta + e^{V_t} \Delta W_t^{H^y}$ depends on $H^y$, i.e. a short time memory, a long time memory or no memory for $0 < H^y < 0.5$, $0.5 < H^y < 1$ or $H^y = 0.5$ respectively.

**Proof.**

1. For the stationary process $V_t$ and $h = s - t > 0$, it is easy to check that

\[
E(V_sV_t) = E(V_t E(V_s|V_t)) = E \left( V_t \left( V_t e^{-\kappa(s-t)} + \theta(1 - e^{-\kappa(s-t)}) \right) \right) = \theta^2 + \frac{\theta \sigma_v^2}{2\kappa} e^{-\kappa(s-t)},
\]
and $EV_t = \theta$. Therefore, we can obtain that

$$Cov(V_t, V_s) = \frac{\theta \sigma^2_v}{2 \kappa} e^{-\kappa(s-t)},$$
$$Cov(\Delta V_t, \Delta V_s) = \frac{\theta \sigma^2_v}{2 \kappa} \left(2e^{-\kappa(s-t)} - e^{-\kappa(s-t-\delta)} - e^{-\kappa(s-t+\delta)}\right)$$
$$= \frac{\theta \sigma^2_v}{2 \kappa} (2e^{-\kappa h} - e^{-\kappa(h-\delta)} - e^{-\kappa(h+\delta)}) = C_1 e^{-\kappa h},$$

where the constant $C_1$ is defined as $(2 - e^{\kappa \delta} - e^{-\kappa \delta}) \frac{\theta \sigma^2_v}{2 \kappa}$, and

$$\sum_{h=1}^{\infty} |Cov(V_t, V_{t+h})| = \frac{\theta \sigma^2_v}{2 \kappa} \sum_{h=1}^{\infty} e^{-\kappa h} < \infty, \sum_{h=1}^{\infty} |Cov(\Delta V_t, \Delta V_{t+h})| = |C_1| \sum_{h=1}^{\infty} e^{-\kappa h} < \infty.$$

We then conclude that the process $V_t$ (and also its incremental process) has a short time memory.

2. We approximate the pricing model

$$dY_t^H = \mu dt + \sqrt{V_t} W_t^{H_y},$$

where $W_t^{H_y}$ is fractional Brownian motion, by

$$\Delta Y_t^H := \mu + \sqrt{V_t} \Delta W_t^{H_y}$$

for $\Delta W_t^{H_y} := W_{t+1}^{H_y} - W_t^{H_y}$ which is fractional Gaussian noise corresponding to $W_t^{H_y}$.

Further, since the process $W_t^{H_y}$ is independent of the process $V_t$, we obtain that

$$E(\sqrt{V_t} \Delta W_t^{H_y}) = E\sqrt{V_t} E\Delta W_t^{H_y} = 0,$$

and

$$Cov(\Delta Y_t^H, \Delta Y_s^H)$$
$$= Cov\left(\sqrt{V_t} \Delta W_t^{H_y}, \sqrt{V_s} \Delta W_s^{H_y}\right) = E(\sqrt{V_t} \Delta W_t^{H_y} \sqrt{V_s} \Delta W_s^{H_y})$$
$$= E(\Delta W_t^{H_y} \Delta W_s^{H_y}) E\left(\sqrt{V_t} \sqrt{V_s}\right) = \gamma_{H_y}(s-t) E\sqrt{V_t V_s},$$
where \( \gamma_H(s) = \frac{1}{2} (|s + \delta|^{2H} + |s - \delta|^{2H} - 2|s|^{2H}) \) is the ACVF of a fractional Gaussian noise with the self-similarity index \( H \) and the time lag of the fractional Brownian motion (corresponding to this fGn) as \( \delta \). Consequently,

- if \( H = 0.5 \), for \( s \neq t \), \( \text{Cov}(\Delta Y^H_t, \Delta Y^H_s) = 0 \) and \( \Delta Y^H_t \) is a random walk.

- if \( 0 < H < 0.5 \),

\[
0 \leq |\text{Cov}(\Delta Y^H_t, \Delta Y^H_s)| \leq |\gamma_H(s - t)| \{ E(V_t) \} \frac{1}{2} \{ E(V_s) \} \frac{1}{2} = |\gamma_H(s - t)| \theta,
\]

and

\[
\sum_{h=1}^{\infty} |\text{Cov}(\Delta Y^H_t, \Delta Y^H_{t+h})| < \infty,
\]

meaning that \( \Delta Y^H_t \) has a short memory.

- if \( 0.5 < H < 1 \), we first define the quantities and check the facts below:

- \( -2c_1 V_s |V_t| \sim \chi^2(2q + 2, 2u_t) \), for \( s - t = h > 0 \), \( c_1 = \frac{2\kappa}{\sigma^2(1 - e^{-\kappa h})} \), \( u_t = \frac{2\kappa}{\sigma^2} \frac{e^{-\kappa h}}{1-e^{-\kappa h}} V_t \), and \( q = \frac{2\kappa \theta}{\sigma^2} - 1 \).

- \( c_3 = c_1 e^{-\kappa h} \), and \( c_4 = \frac{c_3}{c_2} = \frac{1-e^{-\kappa t}}{1-e^{-\kappa t}} e^{-\kappa h} > 0 \).

- Given a random variable \( X \) following a non-central \( \chi^2(K, \lambda) \) with degrees of freedom \( K > 0 \) and non-centrality parameter \( \lambda > 0 \), its density function is given by

\[
f_X(x) = e^{-\frac{\lambda + x}{2}} \sum_{j=0}^{\infty} \left( \frac{\lambda}{4} \right)^j j! \Gamma \left( \frac{K + 2j + 1}{2} \right),
\]

and

\[
E \left( X^{\frac{1}{2}} \right) = \sqrt{2} e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \left( \frac{\lambda}{4} \right)^j j! \Gamma \left( \frac{K + 2j + 1}{2} \right).
\]

- The fractional moment of a random variable \( X \) following a Gamma distribution with mean \( \alpha/\beta \) and variance \( \alpha/\beta^2 \) for \( j \geq 0 \) and a constant \( a > 0 \), is given by
Then, for $h = s - t > 0$,

\[
E\sqrt{V_t V_s} = \frac{1}{\sqrt{2c_1}} E\sqrt{V_t} E\left(\sqrt{2c_1 V_s | V_t}\right) \\
\geq \frac{1}{\sqrt{c_1}} \frac{\Gamma(q + 1.5)}{\Gamma(q + 1)} E\left(\sqrt{2V_t e^{-u_t}} \sum_{j=0}^{\infty} \frac{u_t^j}{j!} \frac{\Gamma(q + j + 1.5)}{\Gamma(q + j + 1)} \right) \\
\geq \frac{1}{\sqrt{c_1}} \left(\frac{\Gamma(q + 1.5)}{\Gamma(q + 1)}\right)^2 \left(\frac{\sigma_v^2}{2\kappa}\right)^{0.5} \left(1 - e^{-\kappa h}\right)^{q+2} \\
\geq \frac{\Gamma(q + 1.5)^2}{\Gamma(q + 1)} \left(\frac{\sigma_v^2}{2\kappa}\right) \left(1 - e^{-\kappa h}\right)^{q+2}
\]

Therefore, for $0.5 < H^y < 1$,

\[
Cov(\Delta Y^H_t, \Delta Y^H_{t+h}) = \gamma_{H^y}(s-t) E\sqrt{V_t V_s} \geq \gamma_{H^y}(s-t) \frac{\sigma_v^2}{2\kappa} (1 - e^{-\kappa h})^{q+2} \left(\frac{\Gamma(q + 1.5)}{\Gamma(q + 1)}\right)^2,
\]

and

\[
\sum_{h=1}^{\infty} |Cov(\Delta Y^H_t, \Delta Y^H_{t+h})| \geq \sum_{h=1}^{\infty} Cov(\Delta Y^H_t, \Delta Y^H_{t+h}) = \infty.
\]

\[\blacksquare\]

**4.8 Future Work**

**4.8.1 Time-varying structure**

Motivated by the study of S&P500, stochastic volatility models with time-varying memory structures are interesting future topics. To help with this type of research, processes
with time-varying memory structure should be proposed or studied. Therefore, as the basis of future study, a Gaussian process with changed values of $H$ are introduced below. (Ayache et al., 2000.)

• **Definition 4.8.1** *(Hölder Function of Exponent $H$)* Let $(X, d_X)$ and $(Y, d_Y)$ be two metric spaces. A function $F : X \to Y$ is called a Hölder function of exponent $h \geq 0$, if for each $x, y \in X$ such that $d_X(x, y) < 1$, we have

$$d_Y(F(x), F(y)) \leq kd_X(x, y)^h$$

for $x, y \in X$ and some constant $k > 0$. (Lutton and Lévy Véhel, 1998.)

• **Definition 4.8.2** Let $H : [0, \infty) \to [a, b] \subset (0, 1)$ be a Hölder function of exponent $\beta > 0$.

– *(Moving Average Definition)* For $t \geq 0$, the following random function is called multifractional Brownian motion (mBm) with functional parameter $H$ (W denotes ordinary Brownian motion):

$$W_{H(t)}(t) = \int_{-\infty}^{0} [(t-s)^{H(t)-1/2} - (-s)^{H(t)-1/2}]dW(s) + \int_{0}^{t} (t-s)^{H(t)-1/2}dW(s).$$

– *(Harmonizable Representation)* For $t \geq 0$, the following function is called multifractional Brownian motion:

$$W_{H(t)}(t) = \int_{\mathbb{R}} e^{it\xi - 1} |\xi|^{-H(t)+1/2}dW(\xi).$$

Contrarily to fBm, the almost sure Hölder exponent of mBm is allowed to vary along the trajectory, a useful feature when one needs to model processes whose regularity evolves in time. Following the moving average definition or the Harmonizable representation of mBm, this process is a zero mean Gaussian process whose increments
are in general neither independent nor stationary. When $H(t) = H$ for all $t$, mBm degenerates to fBm with the exponent $H$. Since the incremental process of mBm in general are not stationary, it is not likely to define stochastic integrals for mBm in general, but only for some specific forms of the function $H(t)$, e.g. a step function.

• Moreover, the covariance structure of mBm is stated in following proposition (Ayache et al., 2000).

**Proposition 4.8.3** Let $X(t)$ be a standard mBm (i.e. such that the variance at $t = 1$ is 1) with functional parameter $H(t)$. Then,

$$\text{Cov}(X(t), X(s)) = E(X(t)X(s)) = D(H(t), H(s))(t^{H(t)+H(s)} + s^{H(t)+H(s)} - |t-s|)$$

where

$$D(x, y) = \frac{\sqrt{\Gamma(2x + 1)\Gamma(2y + 1)\sin(\pi x)\sin(\pi y)}}{2\Gamma(x + y + 1)\sin(\pi(x+y)/2)}.$$ 

4.8.2 Dependent $W_t^{H_y}$ and $W_t^{H_v}$ or $W_t$

If we relax the assumption of independence of $W_t^{H_y}$ and $W_t^{H_v}$ or $W_t$, some of the results proved in Section 4.5.2 and Section 4.7.2 hold. In particular, results regarding the covariance of $V_t^H$ or $V_t$ do not change. However, results pertaining to $Y_t^H$ need to be revised. The dependence between $W_t^y$ and $W_t^v$ can also take on different forms and induce different types of interactions between $Y_t^H$ and $V_t^H$ or $V_t$. For example, it has been argued that minor disturbances in the $V_t^H$ or $V_t$ process can introduce a delayed effect in the $Y_t^H$ process.
The R/S statistic that defines the Hurst exponent is often used to determine the long-range dependence in a time series. The self-similarity index defined with a process, determines the memory structure underlying a self-similar process. This dissertation shows that asymptotically, the R/S statistic and the self-similarity index of fractional Brownian motion agree in the expectation sense. This result has been mentioned in the literature but – to our knowledge – without formal proof, see, e.g., Taqqu et al. (1995).

In addition to the R/S statistic, many other estimators of the self-similarity index in fractional Brownian motion have been proposed in the literature. Two of those are based on the aggregated variance of the process. We show that for a fixed number of blocks (with fixed block size), the Aggregated Variance Method (AVM) and the Absolute Value of the Aggregated Series (AVA) method result in estimators of $H$ that converge to $H$ with probability 1. Therefore, both estimators are consistent. To prove these results we have fixed both the number of blocks and the block size. It is possible to show that both results still hold if block size and number are not fixed, as long as they are bounded. Our proof does not extend in a natural way to the situation where either block size or block number become unbounded. From a practical viewpoint, this does not constitute a limitation of these approaches if we are interested in the statistics of observed data series.

We studied the properties of several estimators of the self-similarity index that are mentioned in literature. We used fractional Gaussian noise with different values of $H$ to
be our test processes, and examined the sample variance, the bias and the mean square error of these estimators. The results of our simulation experiment suggest that estimators based on the aggregated data behave well, while the Whittle estimator (WHI) and the regression estimator (REG) also show nice properties.

We then investigated the effect of block size on the $H$ and the time-dependence of the estimated $H$. This time we adopted a real time series, the S&P500, to be our test process. One reason to use the S&P500, is that our final interest is in learning about and modeling real data. The S&P500 has studied extensively and it has been found that the estimates of its underlying memory structure may be affected by block size effects. For example, Peters (1996) found that the memory structure of the monthly S&P500 stock price series follows a curve with a peak around block sizes equal to four years. We used the S&P500 weekly price series instead, and applied the regression estimator which showed nice properties in the earlier simulation study and which is easy to implement. To make our results comparable to Peters findings, we also applied the empirical R/S and the R/S statistic to these weekly price data. Besides investigating the block-size effect and the time effect on the estimated $H$, we also studied the memory structure of these data using a permutation test. From our results, we see that the S&P 500 price data behave more like a random walk or a short time memory process with short or very long block size, but exhibit a long-memory property in the medium block-size range, and shows a decreasing time effect.

As a final remark, this work also focused on some of the standard continuous-time models for a process with a short time memory structure. In particular, after an extensive review of the literature we found that few authors have discussed checking the memory structure of complex continuous-time market models. For long memory processes, the discussion about continuous-time market models in the literature is even less abundant. To fill this gap, this dissertation also provides some results in the area of continuous-time
models with long time memory where the long memory is incorporated into the main process, i.e. the price process. These long memory processes do not necessarily capture time-dependence in the memory coefficient $H$. There are some stochastic processes that have a time dependent $H$ including multifractional Brownian motion (mBm), but how to develop a dynamical model for processes with time dependent $H$ process is still an open question that deserves more research.
BIBLIOGRAPHY


