

**ON THE NATURE OF THE STOCK MARKET:
SIMULATIONS AND EXPERIMENTS**

by

Hendrik J. Blok

B.Sc., University of British Columbia, 1993

M.Sc., University of British Columbia, 1995

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(Department of Physics and Astronomy)

We accept this dissertation as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

November 2000

© Hendrik J. Blok, 2000

Appendix C

Long-range memory: The Hurst exponent

Brownian motion (in one dimension) is a random walk on the line where the step length is given by a mean zero Gaussian (normal) probability distribution. Since each of the steps are independent the cumulative position X is known to obey

$$\langle X(t) - X(0) \rangle = 0 \quad (\text{C.1})$$

$$\langle [X(t) - X(0)]^2 \rangle^{1/2} \propto |t|^{1/2} \quad (\text{C.2})$$

so the standard deviation from the origin grows as $t^{1/2}$. Mandelbrot and Van Ness [103] introduced *fractional* Brownian motion (fBm) as a generalization to processes which grow at different rates t^H

$$\langle [X_H(t) - X_H(0)]^2 \rangle^{1/2} \propto |t|^H \quad (\text{C.3})$$

where $0 < H < 1$ is called the Hurst exponent.

Successive increments ξ_H of a fractional Brownian motion are called fractional Gaussian noise (fGn)

$$\xi_H(t) = X_H(t + \delta) - X_H(t) \quad (\text{C.4})$$

where δ can always be rescaled to one (to be discussed). The autocorrelation function (which measures the covariance of a data series with itself at some lag τ) is formally defined as

$$C(\tau) \equiv \frac{\langle [\xi_H(t) - \langle \xi_H(t) \rangle] [\xi_H(t - \tau) - \langle \xi_H(t - \tau) \rangle] \rangle}{\left\{ \langle [\xi_H(t) - \langle \xi_H(t) \rangle]^2 \rangle \langle [\xi_H(t - \tau) - \langle \xi_H(t - \tau) \rangle]^2 \rangle \right\}^{1/2}}. \quad (\text{C.5})$$

For an fGn process the definition is [103,104]

$$C(\tau) = \frac{1}{2} \left(|\tau + 1|^{2H} - 2|\tau|^{2H} + |\tau - 1|^{2H} \right) \quad (\text{C.6})$$

which is obviously zero for $H = 1/2$ (except for $\tau = 0$ where the autocorrelation is always one) while for general $H \neq 1/2$ and large τ

$$\begin{aligned} \lim_{\tau \rightarrow \infty} C(\tau) &\propto \tau^{2H} \left[(1 + \tau^{-1})^{2H} - 2 + (1 - \tau^{-1})^{2H} \right] \\ &\propto \tau^{2H} \left[(1 + 2H\tau^{-1} + H(2H - 1)\tau^{-2}) - 2 + (\dots) \right] \\ &\propto \tau^{2H} [\tau^{-2}] \\ &\propto \tau^{2H-2} \end{aligned} \quad (\text{C.7})$$

so correlations decay slowly and the resulting fractional Brownian motion exhibits long memory effects. Correlations are positive for $H > 1/2$ (*persistence*) and negative for $H < 1/2$ (*antipersistence*) as shown in Fig. C.1. (Note that fBm is not the only framework for generating long range memory effects: for instance, fractional ARIMA(0,d,0) processes also exhibit scaling with an exponent $H = d + 1/2$ [105].)

As for standard Brownian motion, all fBm series are self-affine [105,106]

$$X_H(at) \stackrel{d}{=} a^H X_H(t) \quad (\text{C.8})$$

meaning that the series appears statistically identical under rescaling the time axis by some factor a and the displacement X_H by a^H . Hence, fBm lacks any characteristic time scale and when generating or sampling an fBm series, an arbitrary step length of one unit may be used without loss of generality [107]. Self-affine signals can be described by a *fractal* dimension D which is related to the Hurst exponent through $D = 2 - H$ for fBm [108,109]. (The fractal dimension D can be loosely interpreted as the “number of dimensions” the signal fills up. For example, notice that in Fig. C.1 the $H = 0.1$ signal “fills in” significantly more space than $H = 0.9$ and, consequently, has a higher fractal dimension.)

The power spectrum (defined as the amplitude-squared contributions from the frequencies $\pm f$, $S(f) \equiv |F_H(f)|^2 + |F_H(-f)|^2$ where F_H is the Fourier transform of X_H [20]) of fBm also demonstrates scaling behaviour. The exact spectrum is difficult to compute but for low frequencies it can be approximated by a power law $S(f) \sim 1/f^{2H+1}$ [105] (see Fig. C.2) which corresponds to long-term spatial correlations. *Flicker* or $1/f^\alpha$ noise with $\alpha \approx 1$ is ubiquitous in nature (see Ref. [12] and references therein) and some of it may be attributable to long-memory fBm processes [110]. Note that from the definition of the Fourier transform the derivative of fBm, fractional Gaussian noise, also has a low frequency power law spectrum but with exponent reduced by 2, i.e.. $1/f^{2H-1}$.

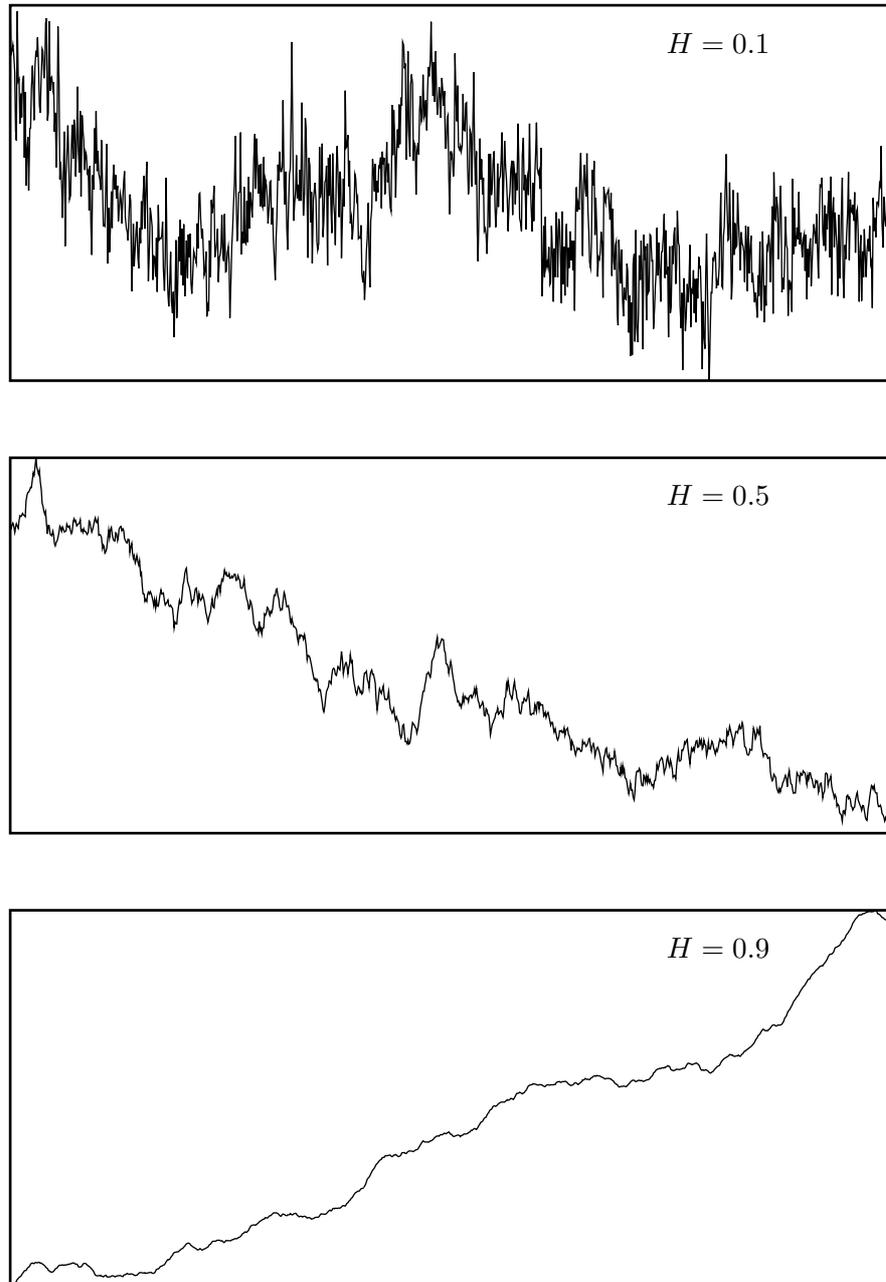


Figure C.1: Sample fractional Brownian motion time series with different Hurst exponents: antipersistent $H = 0.1$ (top) has negative long-range correlations, uncorrelated $H = 0.5$ (center) is standard Brownian motion, and persistent $H = 0.9$ (bottom) has positive long-range correlations.

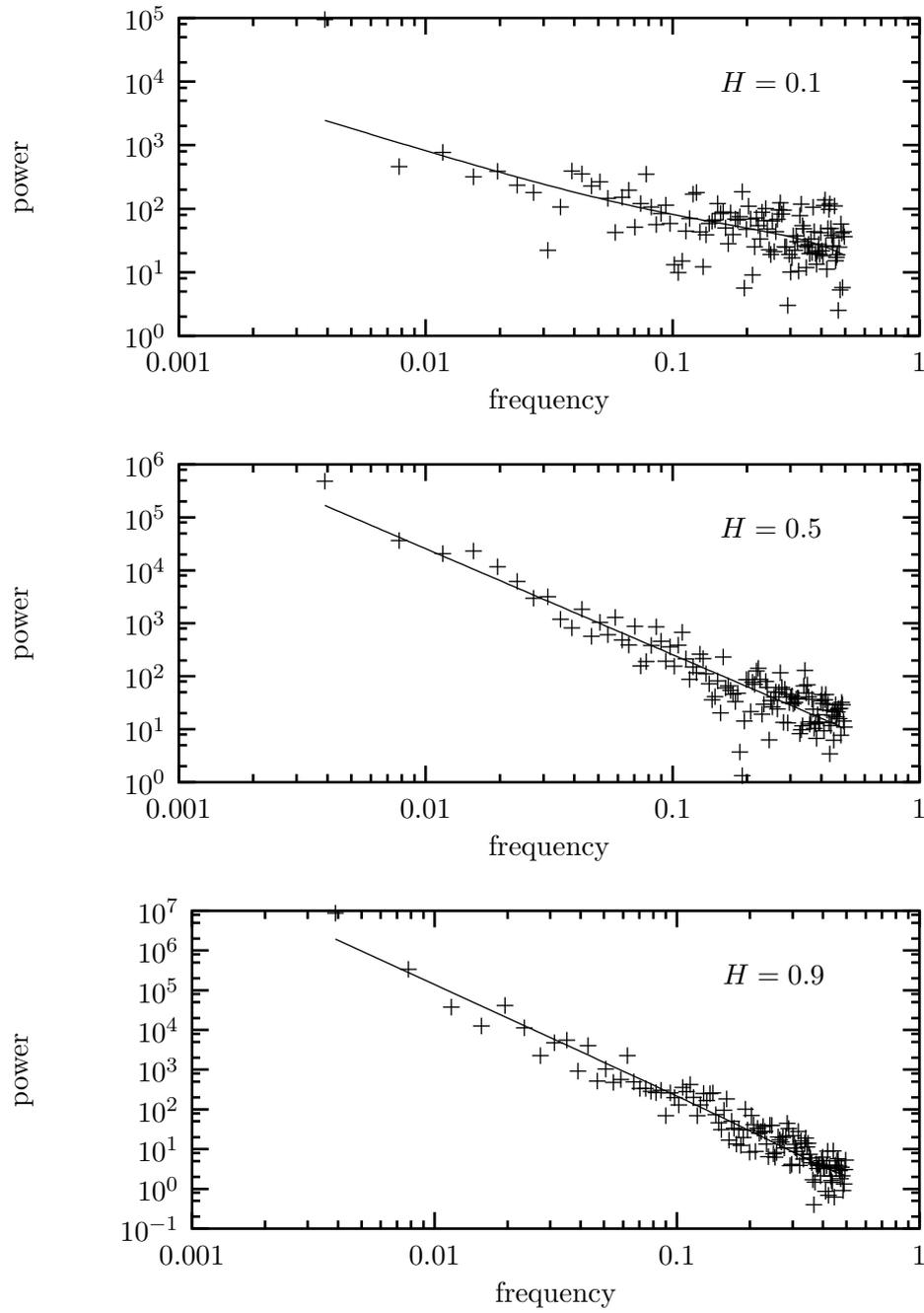


Figure C.2: Power spectral densities for the fractional Brownian motion time series shown in Fig. C.1. The points are from finite samples of 1000 points each and the line represents the theoretical spectrum. For low frequencies the power spectrum is well approximated by a power law $1/f^{2H+1}$.

Fractional Brownian motion has been criticized because it lacks a physical interpretation and because the process has an unrealistic *infinite* memory [111,112]. However, it suits our purposes here because it is a mathematically elegant extension of standard Brownian motion which introduces long-range memory effects and can be characterized by a single parameter H . Hence, it is an ideal experimental control for testing procedures of measuring the Hurst coefficient in real data sets.

C.1 Synthesis

Before we can test various methods of estimating the Hurst exponent, we need some *control* data sets with known exponents. This data must be synthesized from the first principles of fractional Brownian motion as defined above. The computational difficulty is that for $H \neq 1/2$ fBm has an infinite dependence on its history so approximations are required. A number of generators have been proposed [104,106,107,113] but most are slow and/or inaccurate. One of the most common techniques, Successive Random Addition (SRA) [114] is very fast but its correlation function does not match that of fBm.

Another technique, the Spectral Synthesis Method (SSM) [82], uses the scaling behaviour $1/f^{2H+1}$ of the power spectrum to generate synthetic data in frequency space and then inverse Fourier transform the data to recover the desired time series. Although simple and fast—the Fast Fourier Transform (FFT) algorithm only requires on the order of $N \log N$ operations—it fails because the power law in the frequency domain only applies for low frequencies, as mentioned above.

I prefer the generator by Vern Paxson [115] because it is quick and accurate. It also uses a Fourier transform but it uses an accurate approximation to the fBm correlation function to generate a proper power spectrum.

The basic algorithm for generating a data set of N points with Hurst exponent H follows:

1. Find the smallest integer N_8 which is a power of 2 and is not smaller than $8N$.
2. Generate a discrete power spectrum for $f_i = i/N_8, i = 1, \dots, N_8/2$ using Paxson's equations [115, Eqs. (4-6)] given here for convenience (see Ref. [116] for derivation):

$$S(f) = \mathcal{A}(f, H) \left[|2\pi f|^{-2H-1} + \tilde{\mathcal{B}}_3''(f, H) \right] \quad (\text{C.9})$$

where

$$\mathcal{A}(f, H) = 2 \sin(\pi H) \Gamma(2H + 1) (1 - \cos(2\pi f)) \quad (\text{C.10})$$

and

$$\tilde{\mathcal{B}}_3''(f, H) = [1.0002 - 0.000842f] \tilde{\mathcal{B}}_3'(f, H) \quad (\text{C.11})$$

$$\tilde{\mathcal{B}}_3'(f, H) = \tilde{\mathcal{B}}_3(f, H) - 2^{-7.65H-7.4} \quad (\text{C.12})$$

are improved approximations of

$$\tilde{\mathcal{B}}_3(f, H) \approx a_1^d + b_1^d + a_2^d + b_2^d + a_3^d + b_3^d + \frac{a_3^{d'} + b_3^{d'} + a_4^{d'} + b_4^{d'}}{8\pi H} \quad (\text{C.13})$$

where

$$\begin{aligned} d &= -2H - 1 \\ d' &= -2H \\ a_k &= 2\pi(k + f) \\ b_k &= 2\pi(k - f). \end{aligned} \quad (\text{C.14})$$

3. Choose a zero-amplitude null component of the power spectrum $S(0) = 0$ to detrend the fGn increments in real space (zero mean).
4. Multiply each component of the power spectrum by a Poisson distributed uniform deviate η_f with mean $\langle \eta \rangle = 1$

$$S(f) \leftarrow \eta_f S(f). \quad (\text{C.15})$$

This simulates the noise associated with a real data series, for which uncertainties in the power spectrum are multiplicative [20, p. 552].

5. Construct the complex Fourier space representation of the series f_i , $i = -N_8/2, \dots, +N_8/2$ from the power spectrum using random phases $0 \leq \theta_f < 2\pi$

$$F_H(f) = \sqrt{S(|f|)} e^{i\theta_f}. \quad (\text{C.16})$$

Randomizing the phases does not disturb the power spectrum and ensures the finite-sample correlation function converges to the proper theoretical form in the limit $N \rightarrow \infty$ [107].

6. Compute the inverse Fourier transform $\xi_H(t_i)$, $i = 1 \dots N_8$ and discard the imaginary components to get a fractional Gaussian noise series

$$\xi_H(t) = \Re \left(\mathcal{F}^{-1} [F_H(f)] \right). \quad (\text{C.17})$$

7. Pick a random subset of length N of the series and discard the remainder. This minimizes wrap-around effects from the Fast Fourier Transform [20, 110, 117, 118] and gives the illusion of a non-stationary series (to simulate real data, for which the stationarity may be difficult to decide). Note that Paxson does not consider subsampling in his original algorithm.

8. Finally, to convert to a fractional Brownian motion, simply integrate

$$X_H(t) = X_H(t-1) + \xi_H(t). \quad (\text{C.18})$$

Paxson's method is accurate [115], computationally simple, and fast (most of the computation is in the Fourier transform so it still only requires on the order of $N \log N$ operations).

C.2 Analysis

One's first instinct to check for long-range correlations in a data set may be to simply test how quickly the autocorrelation function (Eq. C.5) decays with large lags. This proves to be a poor choice however, because antipersistent data is difficult to distinguish from uncorrelated, the correlations can be mistaken for noise fluctuations around zero.

A method to reliably estimate the Hurst coefficient from a time series would be a useful method of testing for and quantifying long-range correlations. The oldest and still most common method is due to Hurst [119] who noticed that the range R of the depth (or cumulated influx) of water behind a dam over a span of time τ was related to the standard deviation S of the influx over the same period through

$$R/S \propto \tau^H \quad (\text{C.19})$$

where H should be $1/2$ for random, uncorrelated processes [120]. Hurst's method, Rescaled Range or R/S analysis, was to sample non-overlapping subsets of length τ from a time series and calculate the average R/S statistic. Repeating over a wide range of τ -values and recording the data on double-logarithmic graph the Hurst exponent should emerge as the slope of a straight line, $\log(R/S) = H \log \tau + C$.

Unfortunately, despite its extensive usage [21, 76, 111, 121–123] Hurst's rescaled range analysis has been shown to be a poor estimator of H [108, 112, 118, 120] with a consistent bias towards $H = 0.7$ and requiring a large data set for convergence.

Another common technique for testing for correlations is shuffling the order of the data and comparing the statistics of the original data with the shuffled. Shuffling destroys the correlations in the data but care must be taken to detrend the data as well. Persistent data series are characterized by large low-frequency components which make the data series appear non-stationary (notice, for example, the trend in the $H = 0.9$ series in Fig. C.1). Shuffling without first removing this trend would not destroy these low-frequency correlations which extend throughout the entire dataset. Shuffling is a valuable way of testing for correlations but, in itself, doesn't specify any statistical techniques for distinguishing the original from the shuffled

series, and we have already seen that the correlation function and rescaled range analysis are inadequate.

A number of alternatives have been proposed including autocorrelation analysis [82], Fourier analysis [82, 110], and maximum likelihood estimators. The advantage of the latter is that they are not graphical techniques but numerical—they simply return the best estimate of the Hurst exponent directly. Unfortunately, they require (at least) an assumption about the form of the long-range dependence (such as fBm or fractional ARIMA) and perform poorly if the assumption is incorrect [124].

Each of the above methods suffers from biases and slow convergence (a large dataset is required to reduce the bias). However, two methods have been consistently better, requiring smaller datasets and exhibiting less bias [109, 117, 125]: dispersional analysis and scaled-window variance analysis. Both of these methods are graphical, producing a power law relationship from which the exponent can be read off as the slope of the line when using double-logarithmic axes.

C.2.1 Dispersional analysis

Dispersional analysis, also known as the Aggregated Variance method [105], averages the differenced fGn series over bins of width τ and calculates the variance of the averaged dataset. Given a fGn series $\xi_H(i)$, $i = 1, \dots, N$ a particularly simple but effective version of the algorithm follows:

1. Set the bin size to $\tau = 1$.
2. Calculate the standard deviation of the N data points and record the point $(\tau, \tau \cdot \sigma_\tau)$.
3. Average neighbouring data points and store in the original dataset

$$\xi_H(i) \leftarrow \frac{1}{2} [\xi_H(2i-1) + \xi_H(2i)] \quad (\text{C.20})$$

and rescale N and τ appropriately

$$\begin{aligned} N &\leftarrow N/2 \\ \tau &\leftarrow 2\tau. \end{aligned} \quad (\text{C.21})$$

4. As long as more than four data points remain ($N > 4$) return to Step 2. (The reader may prefer to require more than four bins to reduce noise.)
5. Perform a linear regression on the log-log graph

$$\log(\tau \cdot \sigma_\tau) = H \log \tau + C; \quad (\text{C.22})$$

the calculated slope is the best estimate of H .

Recording $\tau \cdot \sigma_\tau$ in Step 2 instead of just the standard deviation σ_τ is not standard but it simplifies the regression because the Hurst exponent can be simply be read off the graph instead of $H - 1$.

Fig. C.4 shows that Dispersional analysis performs significantly better than rescaled range analysis.

C.2.2 Scaled Window Variance analysis

The other method well-received method, Scaled Window Variance analysis (SWV), also known as Detrended Fluctuation Analysis [126] or Residuals of Regression [105], applies to the cumulated fBm series instead. Given a fBm series $X_H(i)$, $i = 1, \dots, N$ my own variation of the algorithm follows:

1. Split the series into $M \equiv \lfloor N/\tau \rfloor$ (where $\lfloor x \rfloor$ is the floor operator—returning the greatest integer not greater than x) evenly-spaced bins of size $\tau = 16$ (SWV is inaccurate for smaller τ [109])

$$X_H^{(k)}(j) = X_H((k-1)\kappa + j), \quad j = 1 \dots \tau \quad (\text{C.23})$$

where

$$\kappa = \left\lfloor \frac{N - \tau}{M - 1} \right\rfloor \quad (\text{C.24})$$

This allows the option of setting a minimum and maximum on the number of bins $M_{min} \leq M \leq M_{max}$. Setting M_{min} larger than N/τ will necessarily result in overlapping bins but this effect has been tested [118] and the benefit of the larger sample size outweighs the influence of cross-correlations introduced.

2. For each bin k , $k = 1 \dots M$, detrend the local series by subtracting off

$$\overline{X}_H^{(k)}(j) = mj + b \quad (\text{C.25})$$

Three options for calculating the trendline have been tested [125]:

- (a) No detrending: $m = 0, b = 0$. This is only recommended for $N < 2^9$ data points.
- (b) Bridge detrending. Form a line between the first and last point in the bin:

$$\begin{aligned} m &= \frac{1}{\tau-1} \left[X_H^{(k)}(\tau) - X_H^{(k)}(1) \right] \\ b &= X_H^{(k)}(1) - m \end{aligned} \quad (\text{C.26})$$

Recommended for $N > 2^{12}$ data points.

(c) Linear detrending. Perform a linear least-squares regression over the entire bin to calculate m and b . Recommended for intermediate N .

3. Calculate the residuals after subtracting off the trend line

$$\widehat{X}_H^{(k)}(j) = X_H^{(k)}(j) - \overline{X}_H^{(k)}(j) \quad (\text{C.27})$$

4. Calculate the standard deviation of the residuals in each bin $\sigma_\tau^{(k)}$ and compute the average and standard deviation of these samples

$$\begin{aligned} \sigma_\tau &\equiv \langle \sigma_\tau^{(k)} \rangle \\ \Delta_\tau &\equiv \sqrt{\langle (\sigma_\tau^{(k)} - \sigma_\tau)^2 \rangle} \end{aligned} \quad (\text{C.28})$$

5. If the average standard deviation σ_τ is non-zero convert it to a log-scale $\sigma_{\log \tau} \equiv \log \sigma_\tau$ and plot $\sigma_{\log \tau} \pm \Delta_{\log \tau}$ versus $\log \tau$. The uncertainty on the log scale can be approximated by

$$\Delta_{\log \tau} \approx \frac{\Delta_\tau}{\sigma_\tau} \log e \quad (\text{C.29})$$

6. Double the bin size $\tau \leftarrow 2\tau$ and repeat while $N > 2\tau$.

7. Perform a linear regression on the log-log graph $\log \sigma_\tau = H \log \tau + C$; the calculated slope is the best estimate of H .

Sample fits using the SWV method (with at least $M_{min} = 4$ bins) are shown in Fig. C.3 using the same data as before. Notice that these data sets are rather small (1000 data points each) but even so, the accuracy is remarkably good. When compared with Hurst's rescaled range analysis (see Fig. C.4) it becomes clear that the SWV method is superior (also edging out the Dispersional method).

Another good feature of the SWV method is that, like all graphical techniques, it clearly reveals multifractal behaviour. At some critical scale, the fractal dimension may crossover to a new value. This is characterized in graphical techniques by a discontinuity in the slope of the log-log graph. In particular, transitions to $H = 0.5$ are often observed for large τ , indicating a transition from correlated behaviour over short time scales to uncorrelated on long time scales. The memory duration can simply be read off the graph as the transition point τ .

C.2.3 Lévy Flight

Despite its advantages, SWV fails in one respect: it is unable to distinguish between long-range correlations and uncorrelated Lévy flight. Lévy flight is similar

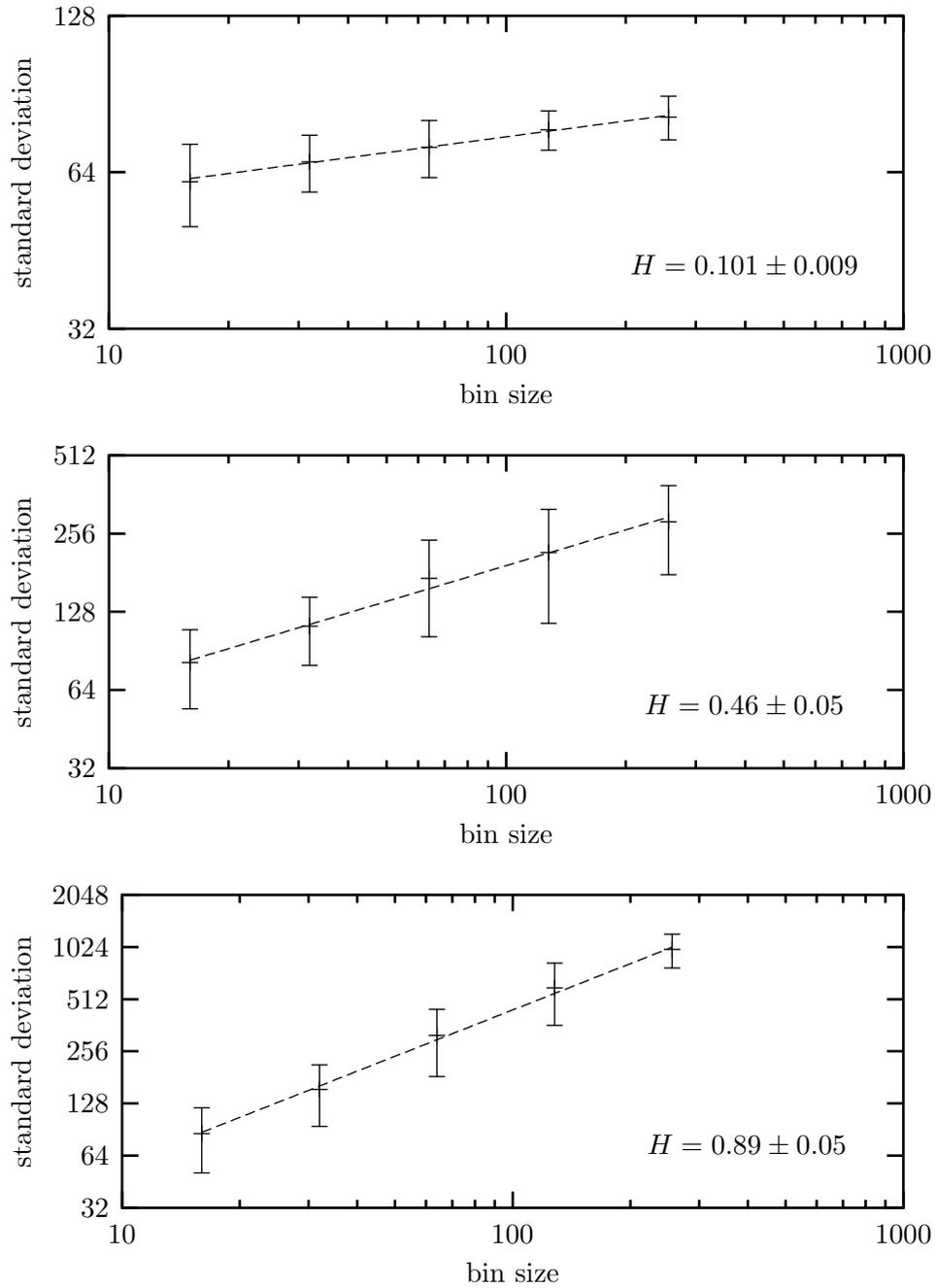


Figure C.3: Scaled window variance analyses for the fractional Brownian motion time series shown in Fig. C.1 (exact $H=0.1$, 0.5 , and 0.9 , respectively). The estimated values of H shown represent the best fit slopes of the lines. The analysis used $M_{min} = 4$ (see the text).

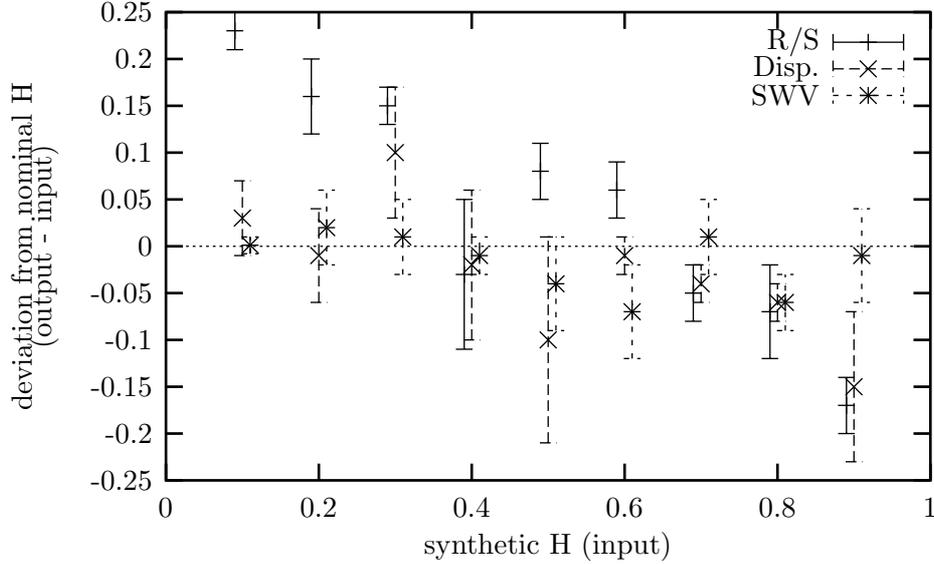


Figure C.4: Comparison of Hurst estimators using synthetic datasets of 1000 points each. The scaled-window variance method (SWV, *) performs significantly better than rescaled range analysis (R/S , +) and marginally better than dispersional analysis (Disp., \times). (The points are offset slightly to improve readability.)

to (traditional) Brownian motion in that it is a cumulated series of independent, identically-distributed (*iid*) increments, but in this case, the increments are Lévy distributed instead of normally distributed.

Normal or Gaussian distributions are well known to obey the following *stability* property: if x_1 and x_2 are both Gaussian-distributed random variables then their sum

$$x \equiv x_1 + x_2 \quad (\text{C.30})$$

is also Gaussian-distributed. Paul Lévy discovered a general class of distributions which have the stability property [83,111,127]. Lévy distributions generally have no closed analytical form but can be defined in terms of their characteristic function $f(k)$ (the Fourier-space representation of the probability distribution) [127,128]

$$\ln f(k; \alpha, \beta) = \begin{cases} -|k|^\alpha (1 - i\beta \tan \frac{\pi\alpha}{2} \text{sign}(k)) & \alpha \neq 1 \\ -|k| \left(1 + \frac{2}{\pi} i\beta \ln |k| \text{sign}(k)\right) & \alpha = 1 \end{cases} \quad (\text{C.31})$$

where $0 < \alpha \leq 2$ is a characteristic exponent and $-1 \leq \beta \leq 1$ is the skewness. Special cases of the Lévy distribution include the Gaussian ($\alpha = 2$, $\beta = 0$) and Cauchy ($\alpha = 1$, $\beta = 0$) distributions.

The stability property says that the sum of a large number of *iid* Lévy random variables will also be a Lévy random variable with the same α and β in apparent violation of the Central Limit Theorem. The paradox is resolved by recognizing that Lévy distributions with $\alpha < 2$ have power-law tails far from the origin

$$p(x) \sim \frac{1}{|x|^{\alpha+1}} \text{ as } |x| \rightarrow \infty \quad (\text{C.32})$$

so the variance $\langle x^2 \rangle$ is infinite for $\alpha < 2$ whereas the Central Limit Theorem assumes a finite variance.

Another interesting property of Lévy distributions is that the cumulative Lévy flight X_α is self-affine, scaling as

$$X_\alpha(at) \stackrel{d}{=} a^{1/\alpha} X_\alpha(t) \quad (\text{C.33})$$

which parallels the scaling relation Eq. C.8 for fractional Brownian motion. A consequence of this is that Hurst coefficient estimators which depend on this scaling property may erroneously predict positive long-term correlations with

$$H = 1/\alpha \quad (\text{C.34})$$

when applied to uncorrelated Lévy flights with $1 < \alpha < 2$.

To test for this effect, data sets of 1000 symmetric ($\beta = 0$) Lévy distributed random variables with were synthetically generated (using a simple and elegant algorithm explained in Ref. [128]) and cumulated to produce a one-dimensional Lévy flight. The synthetic data was then analyzed using the rescaled range, dispersive, and scaled-window variance techniques. The results shown in Fig. C.5 indicate that SWV is sensitive to Lévy noise whereas R/S and Dispersion are not. (Note also that the Fourier spectrum of Lévy flight still approximates a power law $1/f^2$ with exponent 2 (indicating no correlations).)

C.3 Conclusions

In summary, to test for long-range correlations in a data set Dispersive analysis is recommended. If more precision is required (especially for H near 1) and the increments are Gaussian-distributed, a scaled window variance analysis should be performed.

In this discussion we have explored tests for long-range correlations in fractional Brownian motion (correlated with Gaussian increments) and Lévy flight (uncorrelated with non-Gaussian increments). These two extensions to Brownian motion are not exclusive. Fig. C.6 shows how that they are complementary notions

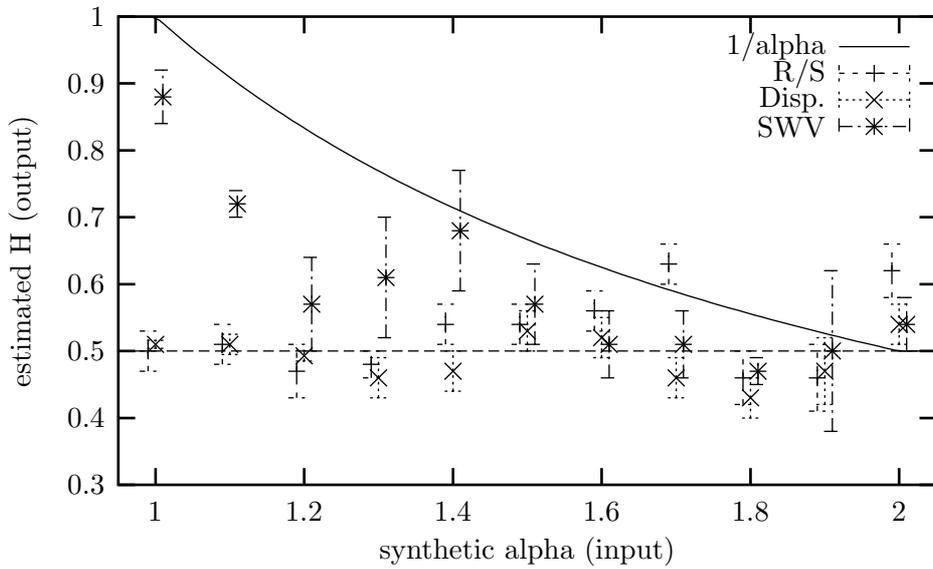


Figure C.5: Comparison of Hurst estimators on uncorrelated Lévy flight with characteristic exponent α using synthetic datasets of 1000 points each. Rescaled range (R/S , +) and dispersal analysis (Disp., \times) perform well but scaled window variance analysis (SWV,*) performs poorly, especially for small α , tending towards the $1/\alpha$ curve. (The points are offset slightly to improve readability.)

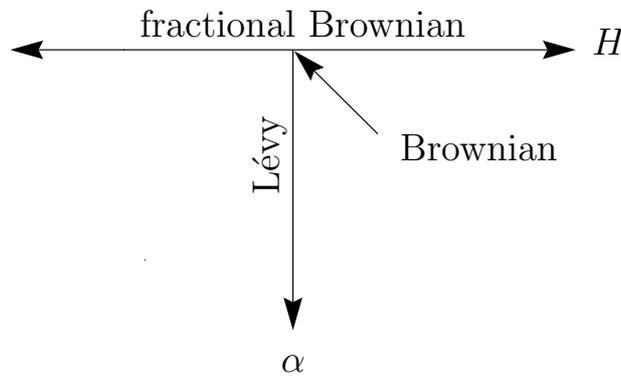


Figure C.6: Schematic representation of relation between fractional Brownian motion and Lévy flight. Traditional Brownian motion sits at the intersection ($H = 1/2$, $\alpha = 2$). The natural extension into the two-space is fractional Lévy motion which has correlated, non-Gaussian increments.

and the two ideas can be combined to produce fractional Lévy motion (fLm) with correlated, non-Gaussian increments. There is very little literature on the subject but it may be a useful model for some natural phenomena [129]. I am unaware of any efficient algorithm to synthesize fLm but it would begin with the correlation function in Eq. C.5, which still applies.