Self-Affine Timeseries Analysis[1]

Rik Blok[2]
Department of Zoology, UBC

March 25, 2003

Abstract

A brief introduction to Lévy flight and fractional Brownian motion from the experimentalist's perspective. Simple tools to analyze these timeseries, the Zipf plot and dispersional analysis, are presented. As a demonstration, these tools are applied to financial and meteorological data to determine the Lévy and Hurst exponents.

Outline

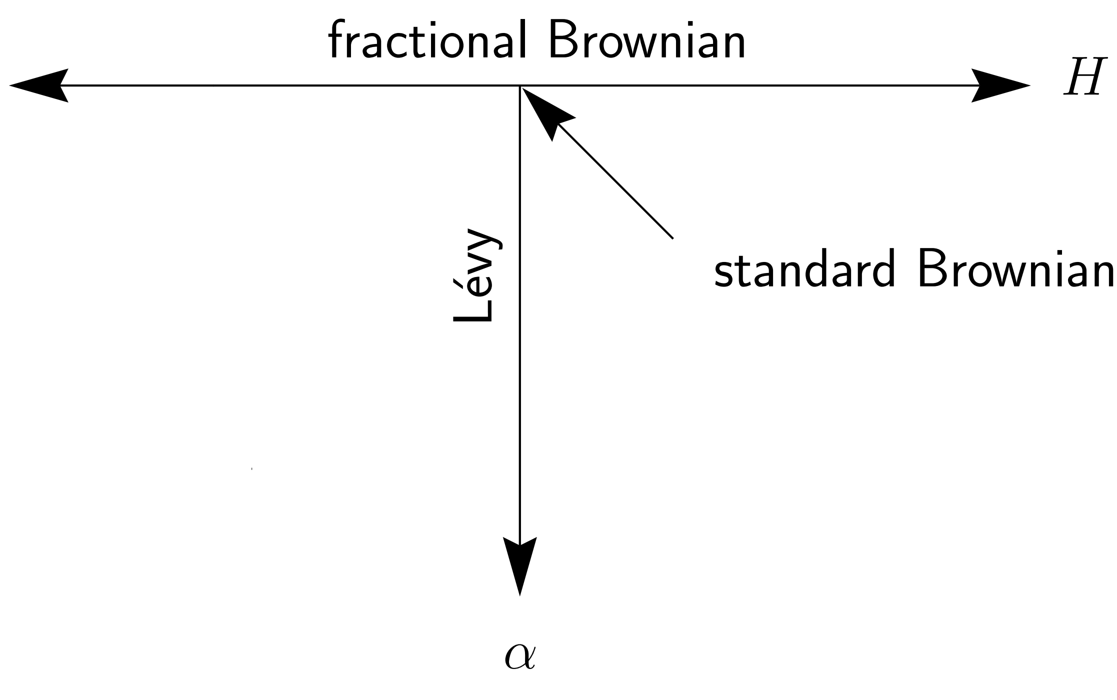
---

[1] `http://rikblok.cjb.net/lib/blok03.html`

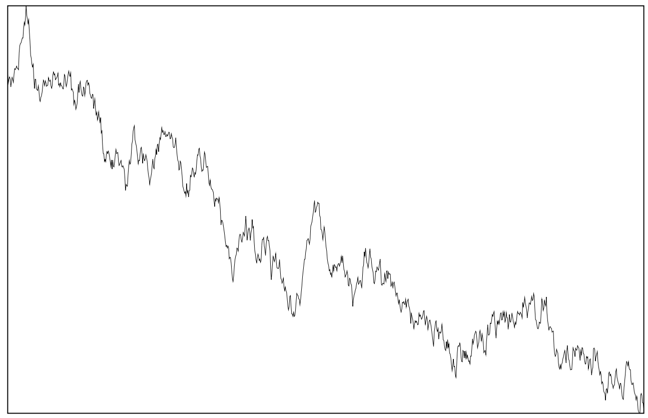
[2] `mailto:rikblok@shaw.ca`

1 Brownian motions

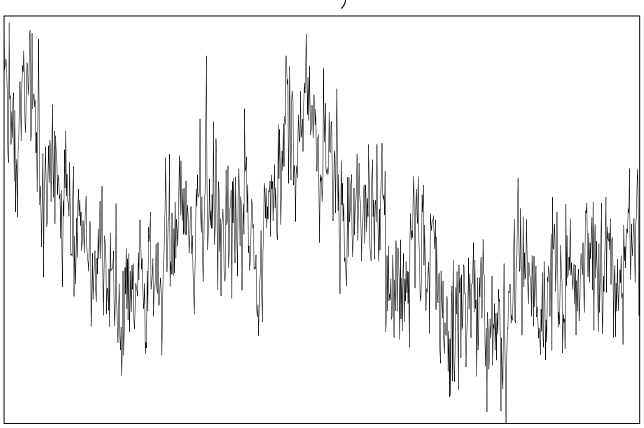
- *Standard Brownian motion* = uncorrelated Gaussian increments (finite variance), $\alpha = 2$, $H = 1/2$
- *Fractional Brownian motion (fBm)* = finite variance but correlations extend over entire history, $H \neq 1/2$ ($0 < H < 1$)
- *Lévy flight* = uncorrelated increments but divergent variance, $\alpha < 2$ (mean also diverges if $\alpha < 1$)



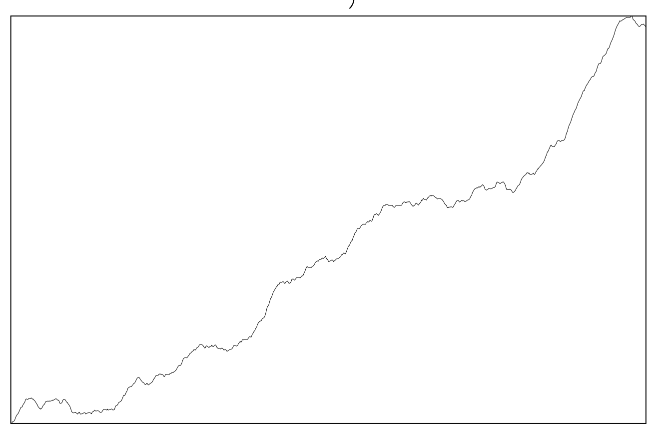
$H = 0.5, \alpha = 2$



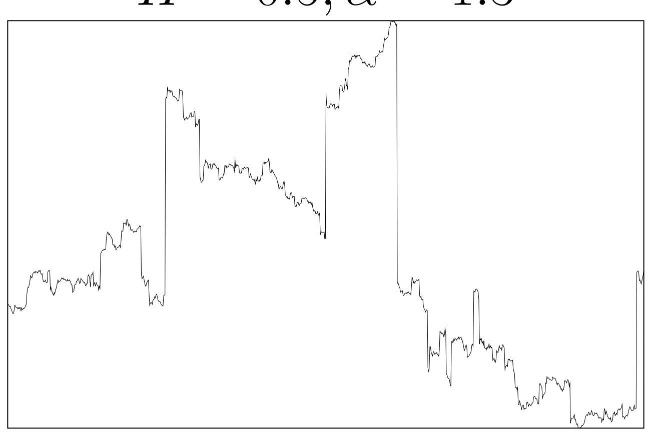
$H = 0.1, \alpha = 2$



$H = 0.9, \alpha = 2$



$H = 0.5, \alpha = 1.3$

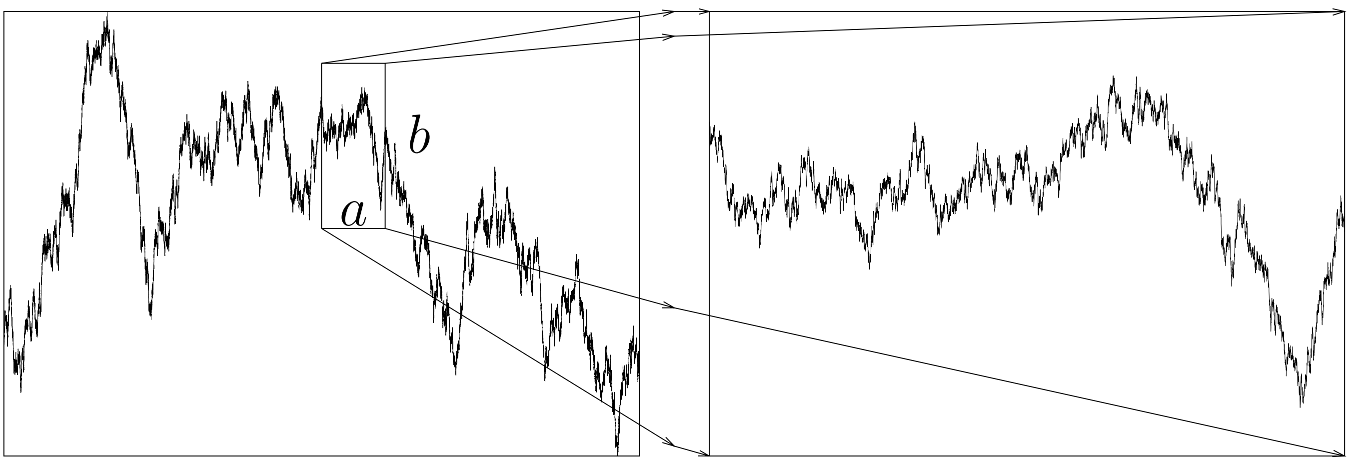


1.1 Self-affine

If timescale zoomed by factor a then process looks statistically identical by scaling series by:

fBm $\Rightarrow b = a^H$

Lévy $\Rightarrow b = a^{1/\alpha}$



My goal is to explain how to obtain parameters H and α from empirical data.

Warning: some tools to calculate exponent rely on self-affinity and are unable to distinguish between fBm and Lévy flight. They will return exponent which could be either H or $1/\alpha$.

2 Lévy flight

Like standard Brownian motion but with overabundance of very large jumps.

$\alpha = 1.3$

If distribution of increments $r(t) = x(t) - x(t - \Delta)$ with stepsize Δ denoted by $p(r)$ then tails of distribution decay as a *power law*:

$$p(r) \sim \frac{1}{|r|^{\alpha+1}} \text{ as } |r| \to \infty \tag{1}$$

for $0 < \alpha < 2$. (For $\alpha = 2$ tails are Gaussian.)

Can use this property to recover α from dataset.

Easier to work with cumulative distributions $C_\pm(r)$,

$$\begin{aligned} C_\pm + (r) &= \int_r^\infty p(r')dr' &= \text{prob. sample } > r \quad (2) \\ C_-(r) &= \int_{-\infty}^r p(r')dr' &= \text{prob. sample } < r. \quad (3) \end{aligned}$$

Then

$$C_\pm(r) \sim \frac{1}{|r|^\alpha} \text{ as } |r| \to \infty. \quad (4)$$

Want to fit this distribution to empirical data. But first...

Power-law tails mean variance diverges. Cannot be true for finite dataset.

Since variance finite, should obey Central Limit Theorem on largest scales. Find good fitting function is [1, 2]

$$\log C_\pm(r) = -\alpha \log |r| - \beta |r| + \gamma, \quad (5)$$

because fit is linear in parameters (α, β, γ).

Effect of finite system size is to truncate power law tail by an exponential cut-off at $r_c \sim 1/\beta$. Well established, empirically.

2.1 Data analysis: Zipf plot

Good choice because the Zipf plot method will not be tricked by other self-affine signals, like fBm (since it shuffles the data).

Recipe

1. Rank order increments r.
2. (Transposed) Zipf plot: Rank versus r.
3. Fit Eq. (5) to data.
4. Interpret results.

Comments

Rank order (sort) increments in both increasing/decreasing orders (to analyze both tails). $\mathrm{Rank}_i \approx NC(r_i)$ for each tail.

Reasonable choice for uncertainty in $\log(\mathrm{Rank})$ is

$$\sigma_i = \sqrt{\frac{N - \mathrm{Rank}_i}{N \cdot \mathrm{Rank}_i}}, \tag{6}$$

from binomial distribution. (Not certain if this improves fit.)

Only want to fit over tail. Start at 2 std. devs., $\mathrm{Rank}_{lo} = 2.5\% N$, then increase lower bound of fit, lo, to minimize reduced chi-squared statistic.

Interpreting results

Must have $0 < \alpha < 2$ [3].

Power law (straight line on log-log graph) must hold over at least one order of magnitude to be significant,

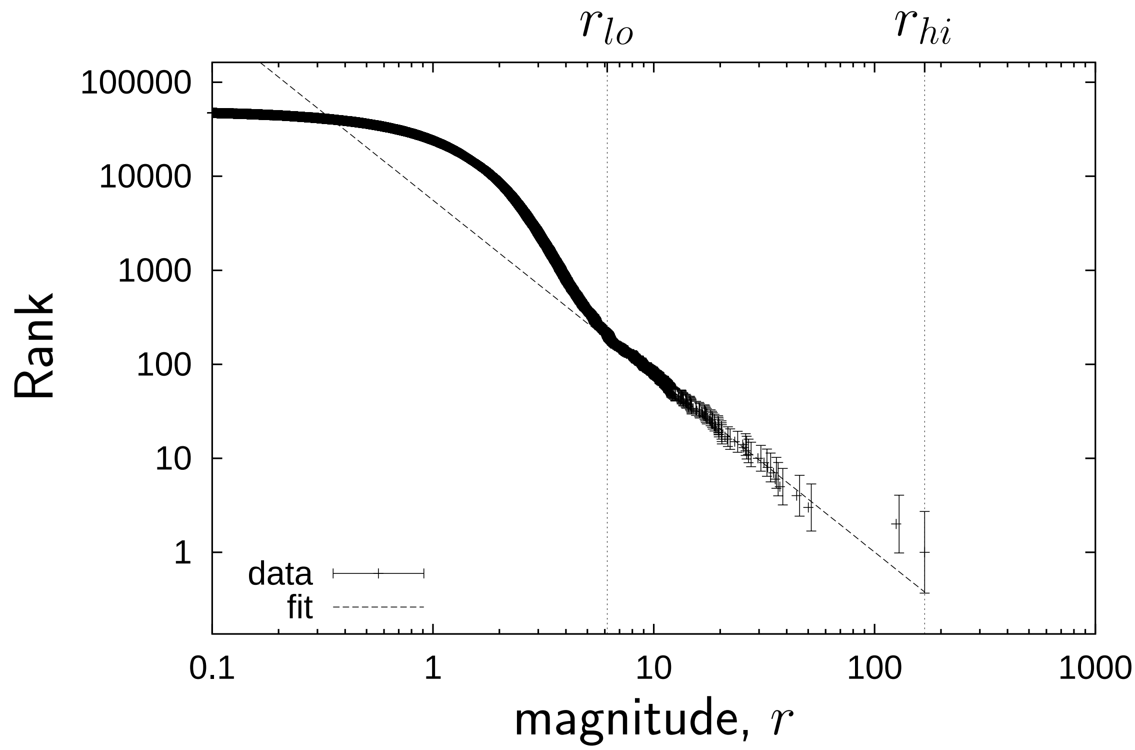
$$\mathrm{OM} = \log_{10} \frac{\min(r_{hi}, r_c)}{r_{lo}} > 1. \tag{7}$$

If either condition fails then Lévy tail not significant so take $\alpha \equiv 2$ (Gaussian).

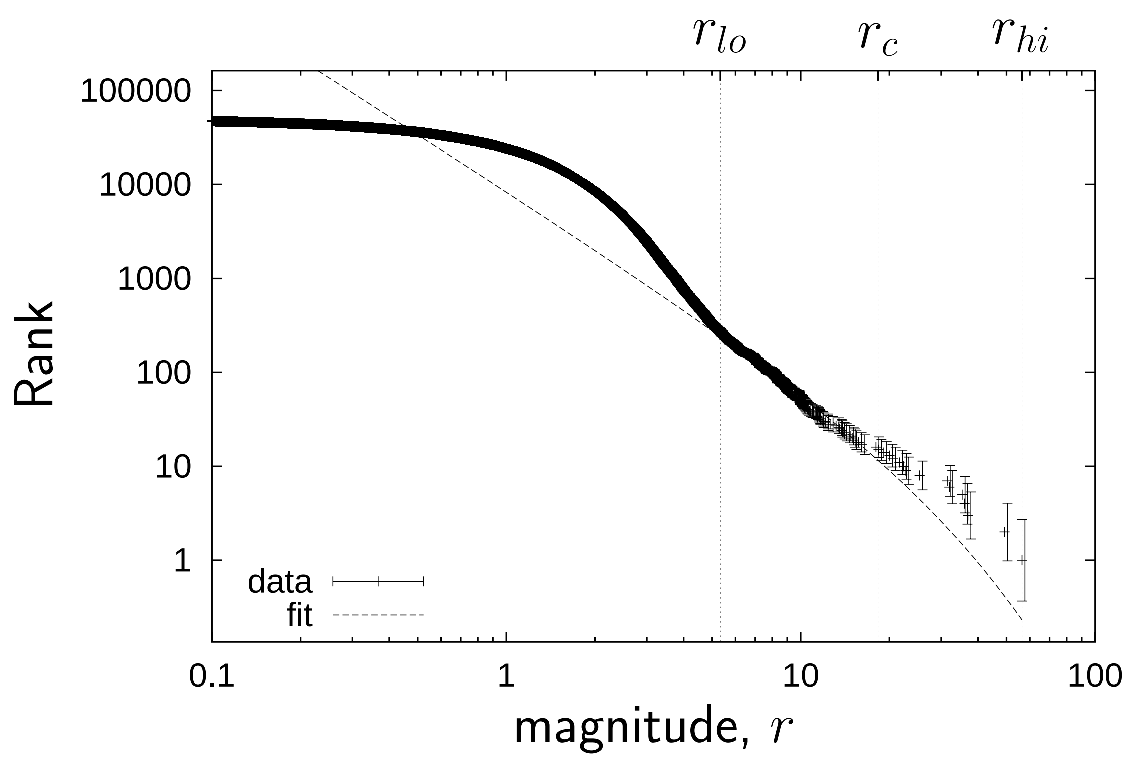
2.2 Test: Synthetic Lévy series

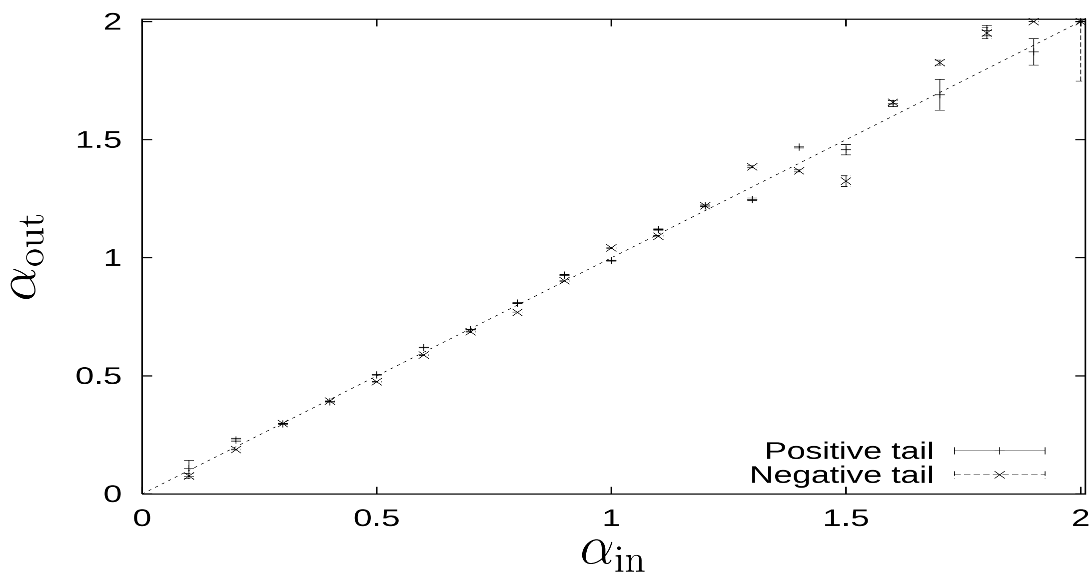
Synthesized timeseries of $N = 100,000$ datapoints for $\alpha_{in} = 0.1, 0.2, \ldots, 2.0$.

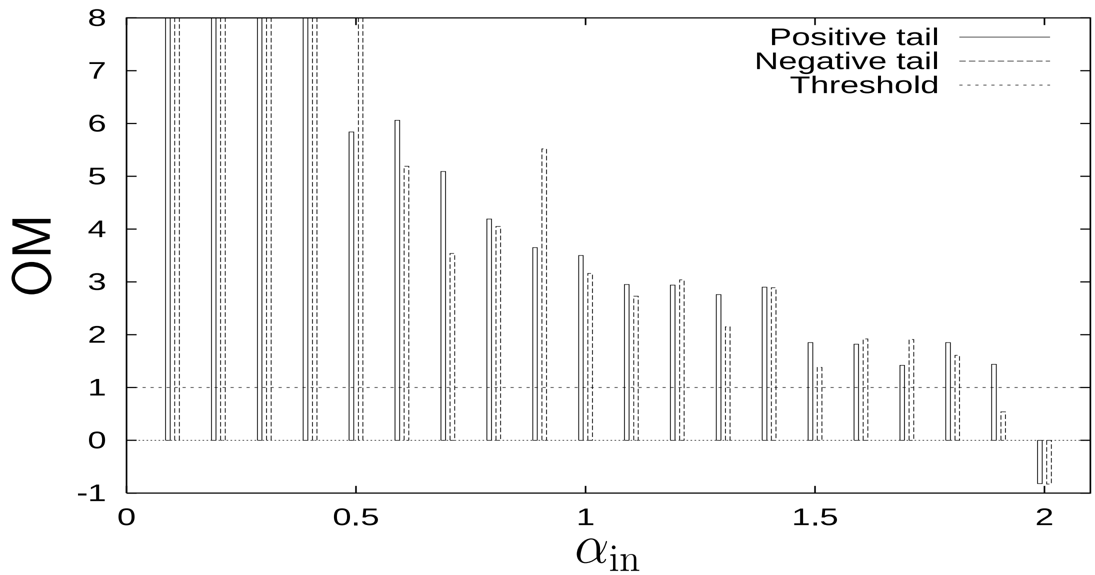
$\alpha_{in} = 1.9$, positive tail, $\alpha_{out} = 1.87$, OM$= 1.4$



$\alpha_{in} = 1.9$, negative tail, $\alpha_{out} = 2.0$, OM$= 0.5$





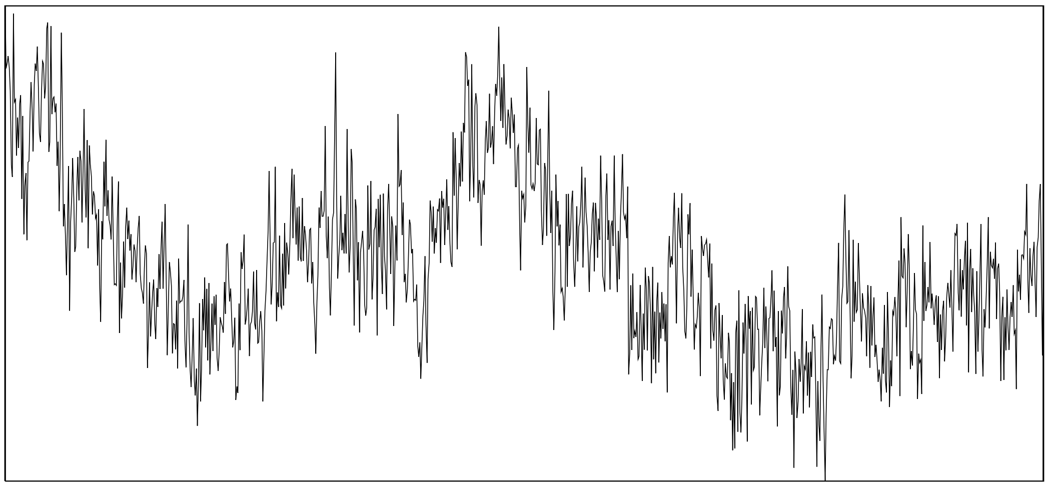


So one out of 40 tails is mis-characterized: for $\alpha_{in} = 1.9$ found OM< 1 so could not confirm Lévy tail. Need very large datasets to distinguish α near 2 from Gaussian.

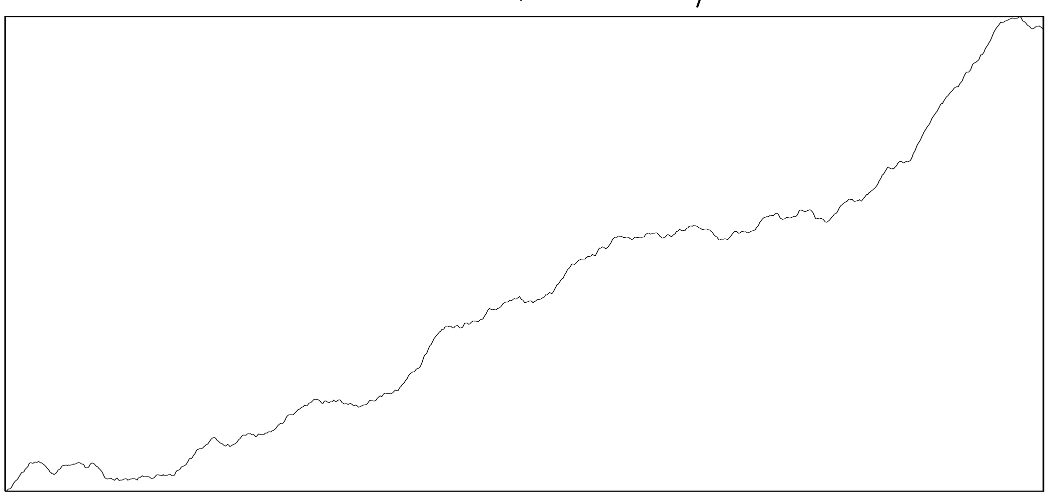
Also tested synthetic fBm series—always returned OM< 1 indicating no Lévy tail.

3 Fractional Brownian motion

Antipersistent, $H < 1/2$



Persistent, $H > 1/2$

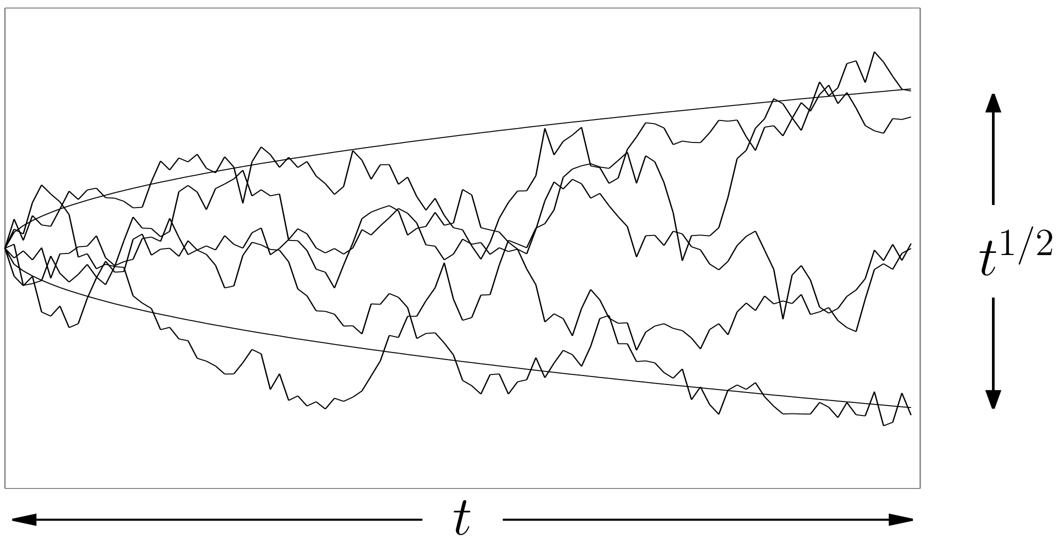


Fractal dimension, $D = 2 - H$, is space filled by signal.

Correlations extend over entire history of series.

3.1 Diffusion

Standard Brownian motion, or random walk, diffuses as $\sigma \sim t^{1/2}$.



In general, fBm diffuses as $\sigma_H \sim t^H$. ($H > 1/2 \Rightarrow$ superdiffusion, $H < 1/2 \Rightarrow$ subdiffusion.) Can use this to estimate H from dataset.

3.2 Data analysis: Dispersion

Methods *not* to use (and why):

- Rescaled range (Hurst, R/S): strong bias $H \to 0.7$.
- Scaled window variance/detrended fluctuation: tricked by Lévy flight.

- Autocorrelation: only for persistent series $H > 1/2$. (Might hold for antipersistent series but would need *huge* datasets, eg. $10^6 - 10^9$ points.)

Dispersional analysis will not confuse Lévy flight with fBm and works for all $0 < H < 1$ with moderately sized datasets.

Slight bias if $H > 0.9$. (Underestimates H.)

Recipe

1. Get increments r.
2. Dispersional analysis.
3. Fit curve.
4. Interpret results.

Comments

As before, work with the increments r_i, called fractional Gaussian noise (fGn).

Again, need $N > 1,000$ data points. Prefer $N > 10,000$.

Dispersional analysis

Start with bin size $L = 1$.

Estimate of diffusion on scale L is given by

$$\sigma(L) \sim \sqrt{\mathrm{Var}\left[r\right]}. \tag{8}$$

Plot deviation, $\sigma(L)$ versus bin size L on log-log scale.

Double bin size $L \to 2L$ and compute increment r over new L,

$$r_i = r_{2i-1} + r_{2i}. \tag{9}$$

Repeat.

Interpreting results

On log-log scale should have linear relationship

$$\log \sigma(L) = H \log L + C \tag{10}$$

where slope is Hurst exponent H.

In practice, finite data series means memory finite so best to skip last five datapoints when fitting.

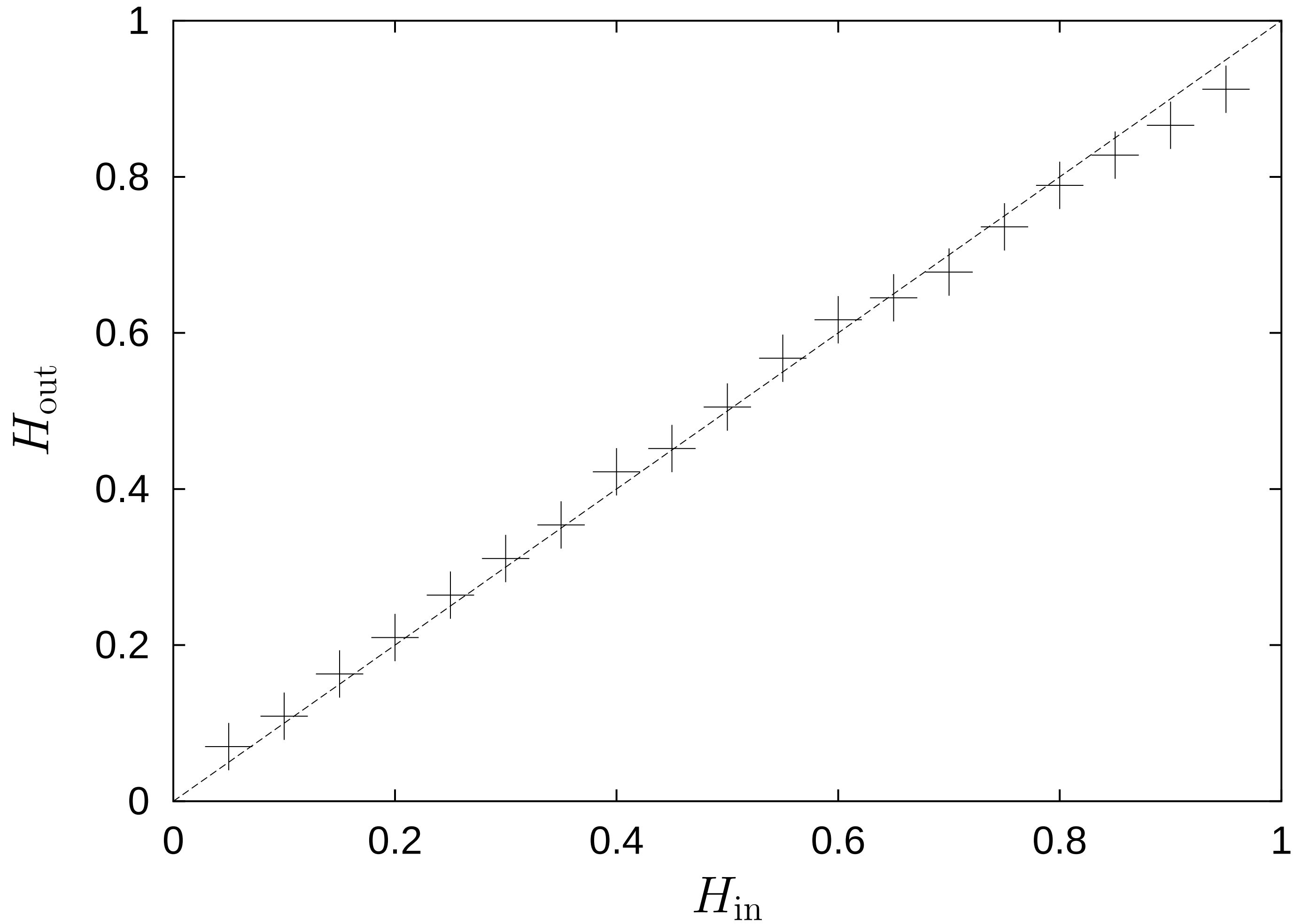
Series might be multifractal, with distinct H values on different L-scales.

If concerned H may be due to artifacts, shuffle data to break correlations and reanalyze. Should get $H \approx 1/2$.

3.3 Test: Synthetic fBm series

Synthesized timeseries of $N = 100,000$ datapoints for $H = 0.05, 0.10, \ldots, 0.95$.

Compare fitted H_{out} to input H_{in}.



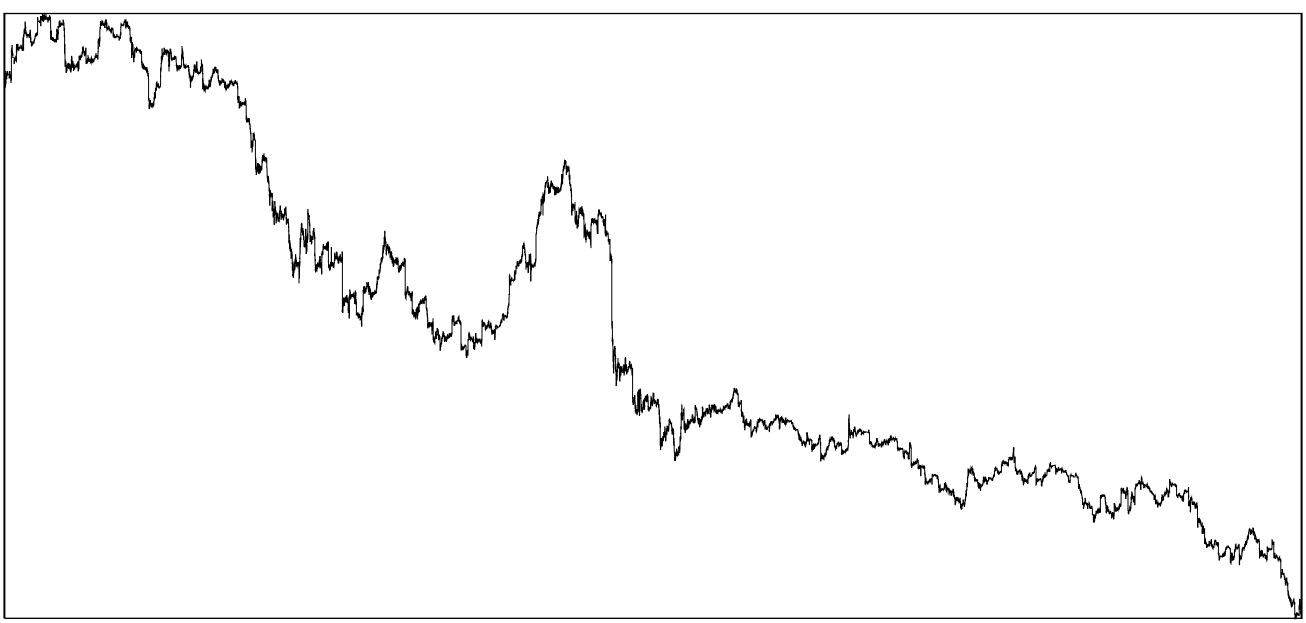
Also tested synthetic Lévy flight series. Returned $H = 0.49 \ldots 0.51$ for all α.

4 Empirical examples

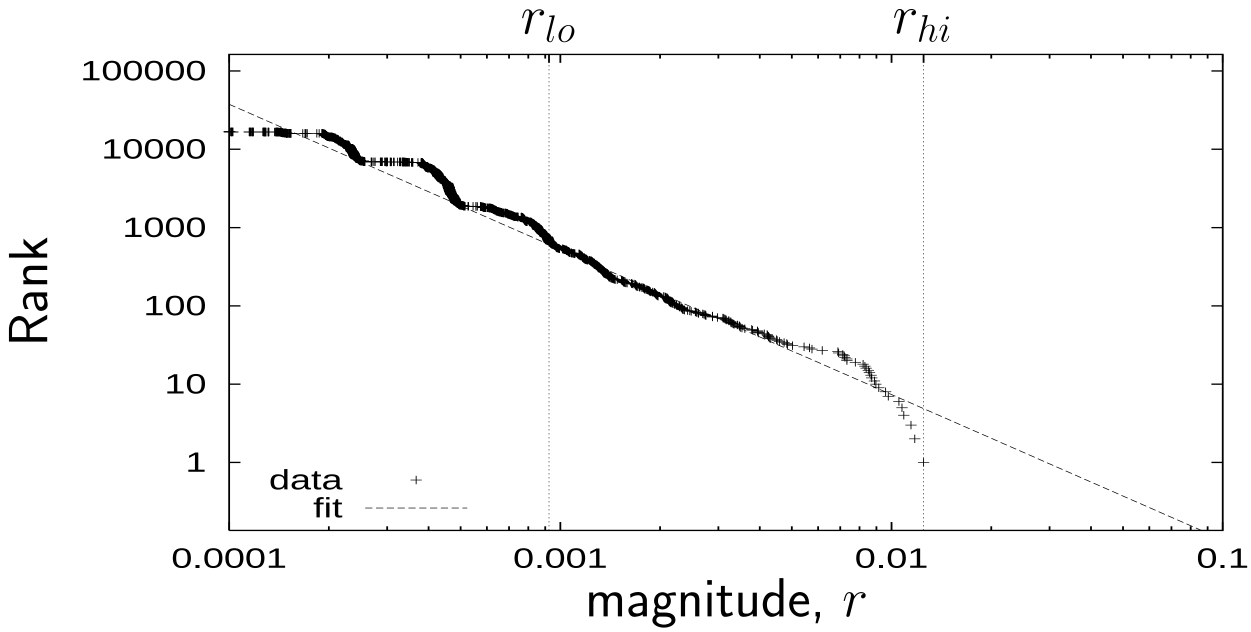
4.1 Swiss Franc versus U.S. Dollar exchange rate

Tickwise data [4] sampled at 1 minute intervals. $N = 99,985$.

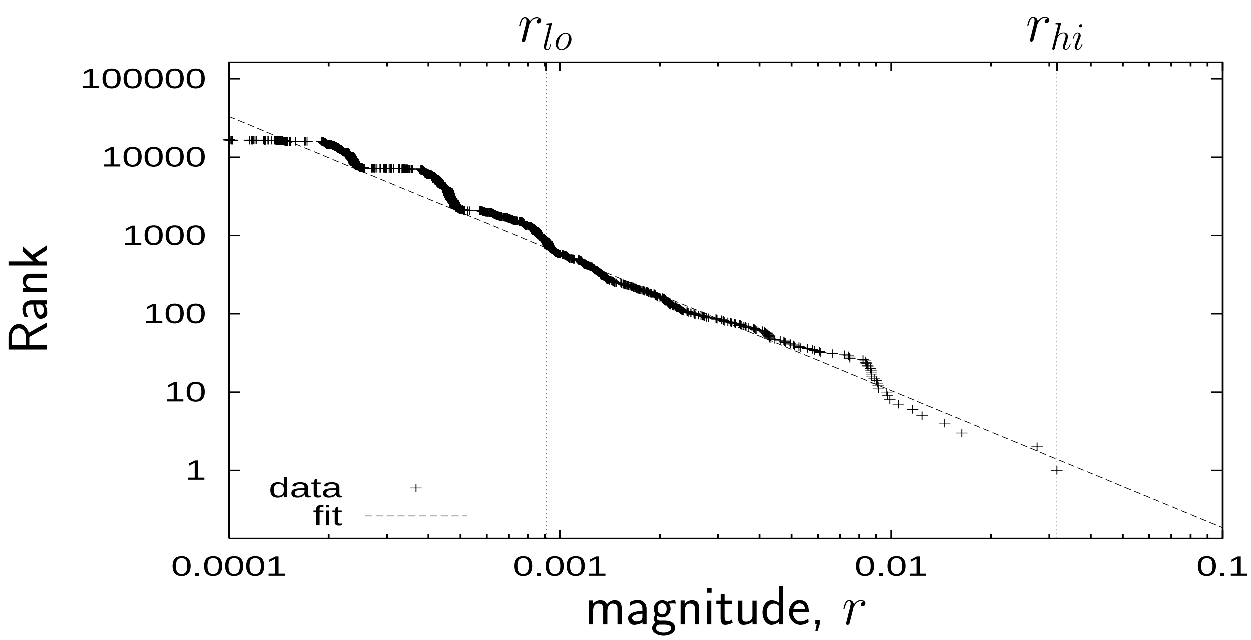
Price is multiplicative process so convert to log-price before processing. (Brownian motions are additive.)



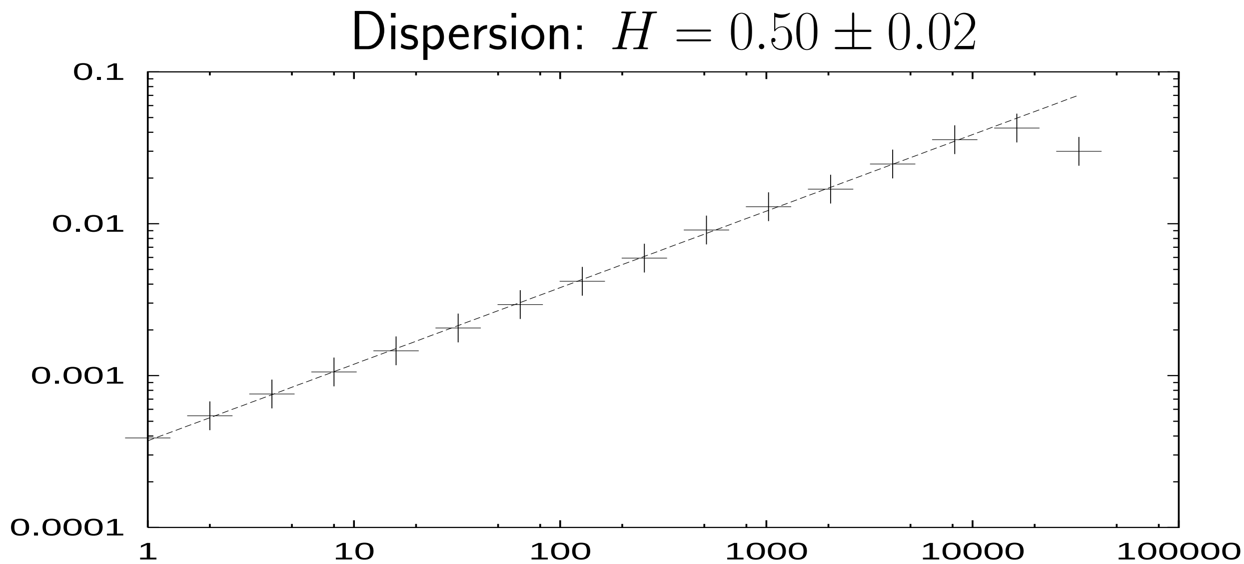
Zipf plot: + tail, $\alpha = 1.85 \pm 0.02$, OM$= 1.1$



Zipf plot: − tail, $\alpha = 1.75 \pm 0.01$, OM$= 1.4$



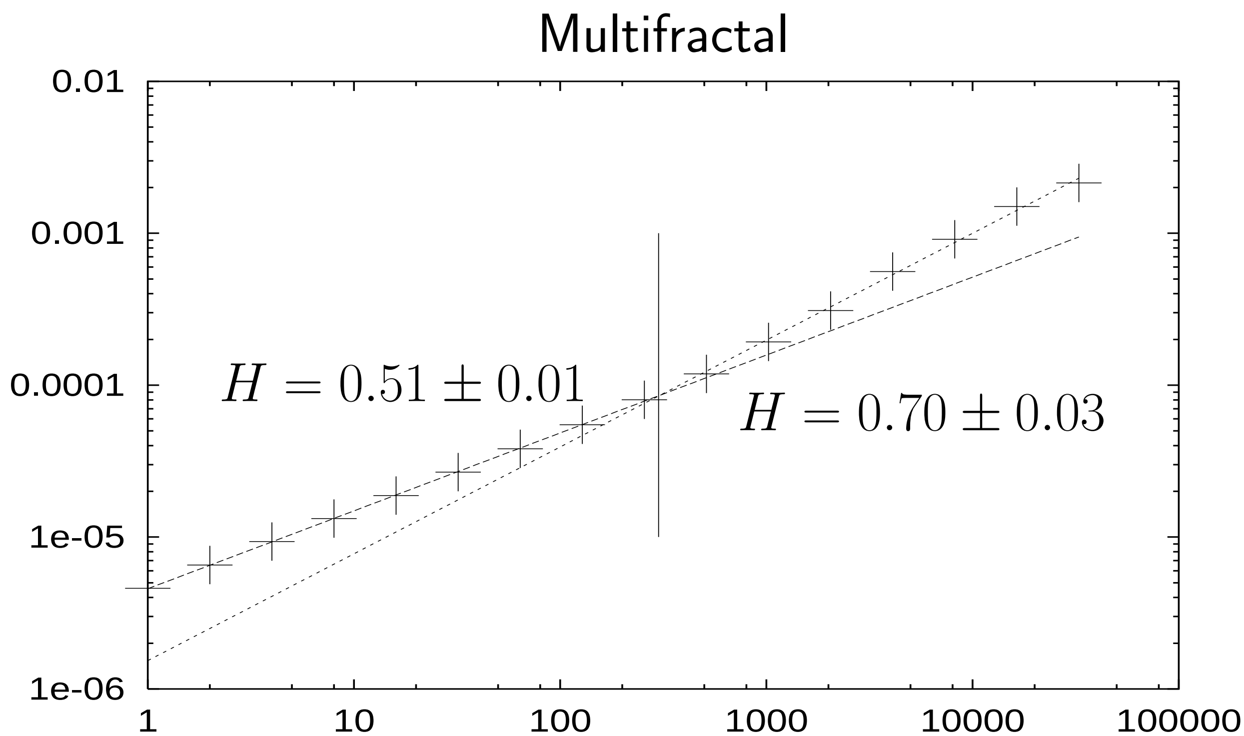
Evidence of a stable Lévy distribution with exponent $\alpha \approx 1.8$.



No memory in price/return history.

Volatility

However, consider squared returns r^2, known as volatility(?) [5, 6]. Measures how much price is fluctuating without regard for direction of movements.



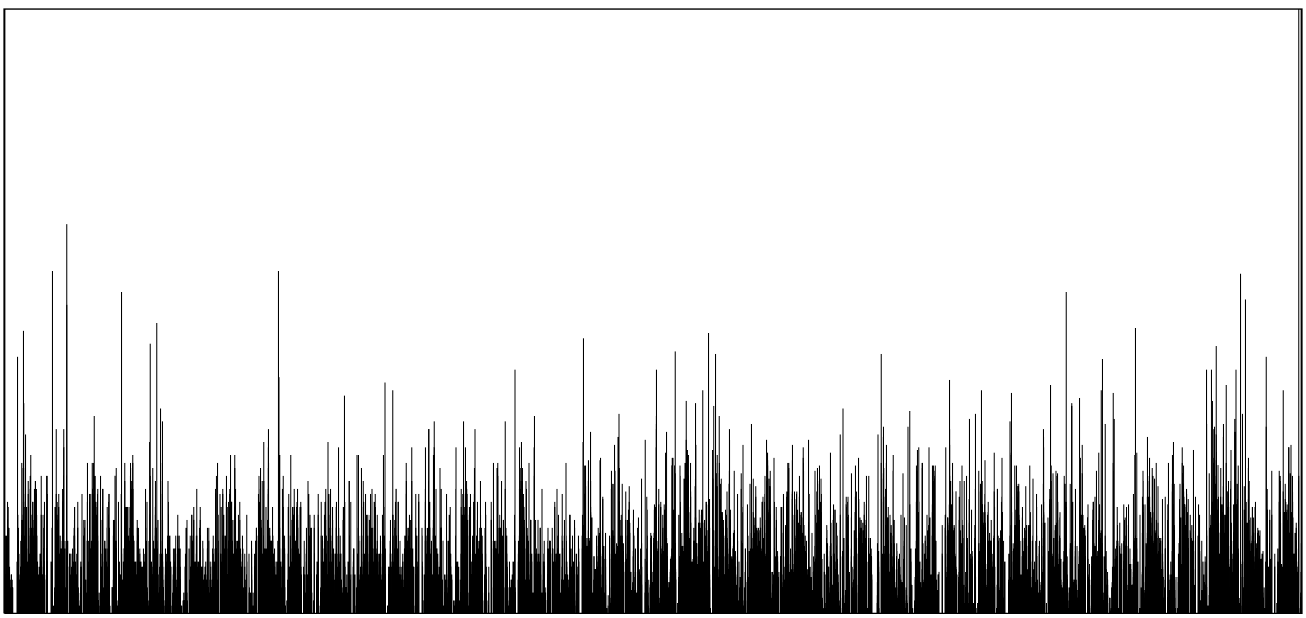
No memory on short timescales but crosses over to positively correlated volatility for timescales > 300 minutes.

Not an artifact of Lévy distribution—synthetic series maintained $H \approx 1/2$ when squared.

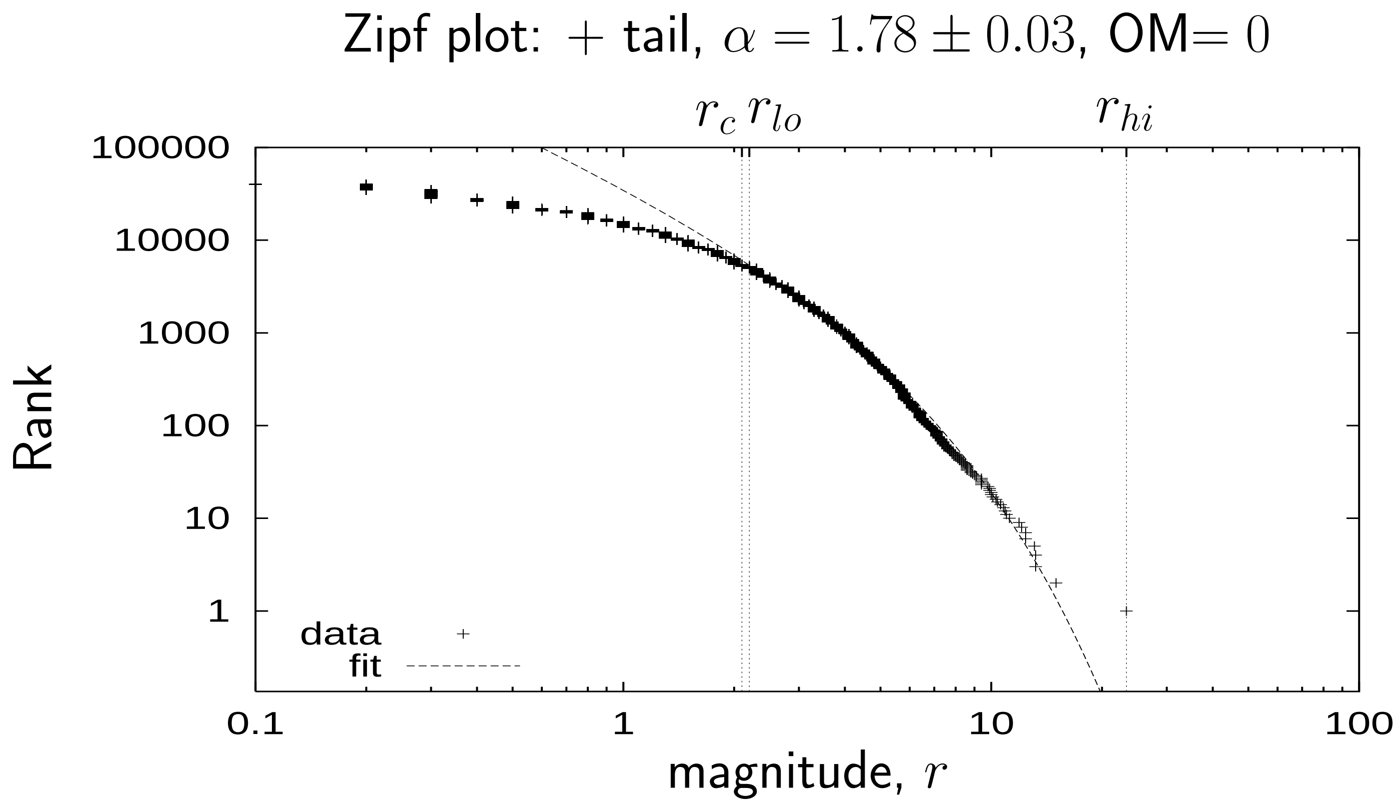
In summary, series exhibits fat tails with Lévy exponent $\alpha \approx 1.8$, no memory in returns, but persistence in volatility ($H = 0.7$) for timescales longer than ∼5 hrs.

4.2 Vancouver precipitation

Hourly precipitation at Vancouver International Airport, 1960–1999. $N = 335,273$.

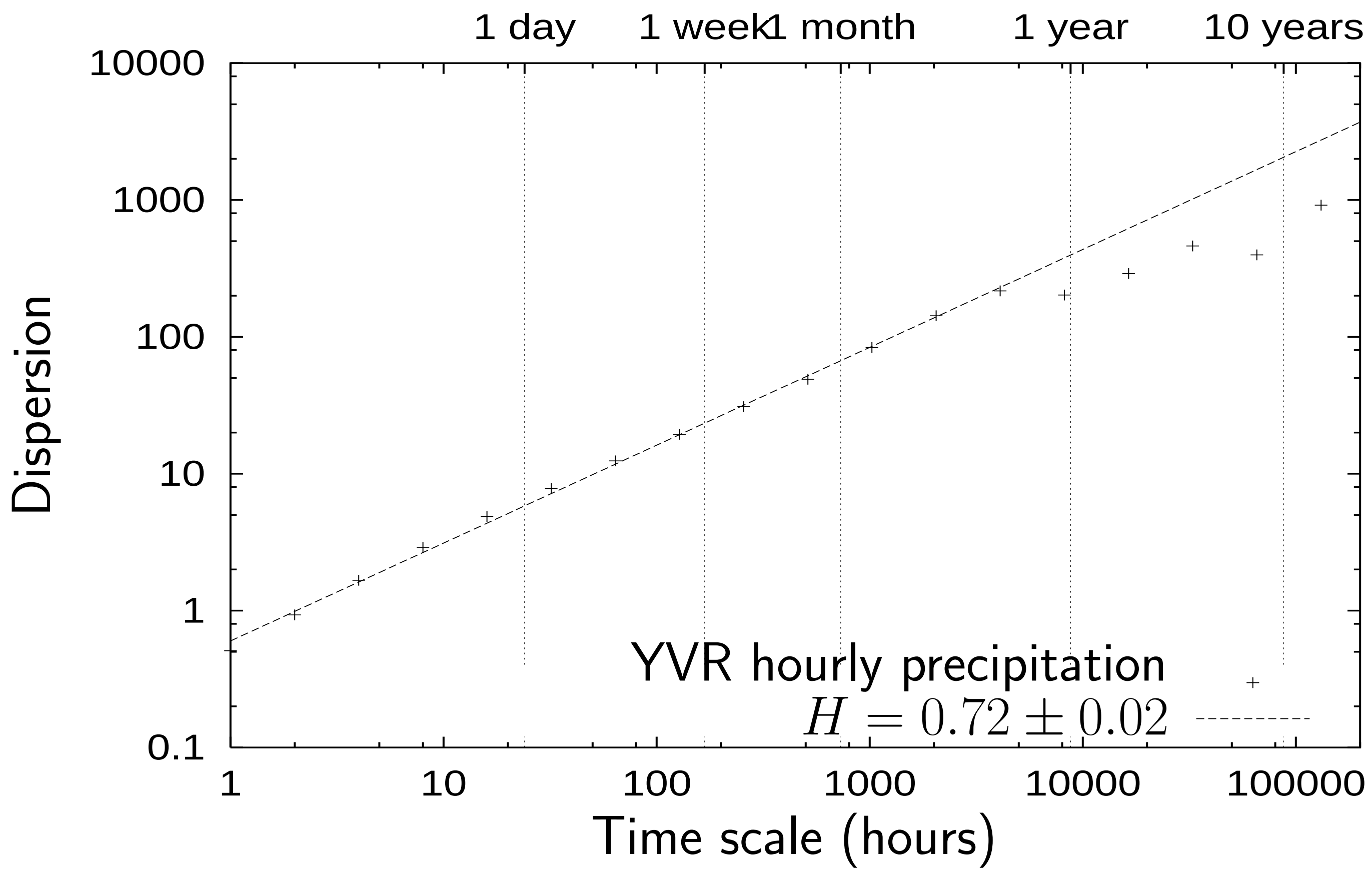


Bounded below by zero so only expect possible Lévy distribution for positive tail:



OM< 1 so does not appear to be Lévy (despite $\alpha < 2$).

Dispersional analysis:



Precipitation has a long ($\sim 1/2$ year) memory. Can confirm this by shuffling data and repeating analysis. Gives $H \approx 1/2$ so effect is due to correlations.

References

[1] Ismo Koponen. Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Phys. Rev. E*, 52:1197–9, 1995.

[2] Hendrik J. Blok. Statistical properties of financial timeseries. PIMS-MITACS Math Finance Seminar, U.B.C., `http://rikblok.cjb.net/lib/blok02.html`, 2002.

[3] Rafał Weron. Levy-stable distributions revisited: Tail index > 2 does not exclude the Levy-stable regime. *Int. J. Mod. Phys. C*, 12(2):209–23, 2001. arXiv:cond-mat/0103256.

[4] Swiss Franc-U.S. Dollar tickwise exchange rate data, 1985–1991. Available from `http://www.stern.nyu.edu/~aweigend/Time-Series/Data/SFR-USD.Tickwise.gz`, provided by Andreas Weigend.

[5] Yanhui Liu, Pierre Cizeau, Martin Meyer, C.-K. Peng, and H. Eugene Stanley. Correlations in economic time series. *Physica A*, 245:437–40, 1997.

[6] Rosario N. Mantegna, Zoltán Palágyi, and H. Eugene Stanley. Applications of statistical mechanics to finance. *Physica A*, 274:216–221, 1999.