Statistical Properties of Financial Timeseries[1]

Rik Blok[2]
Centre for Applied Ethics, UBC

June 6, 2002

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 intraday foreign exchange data to determine the Lévy and Hurst exponents.

Outline

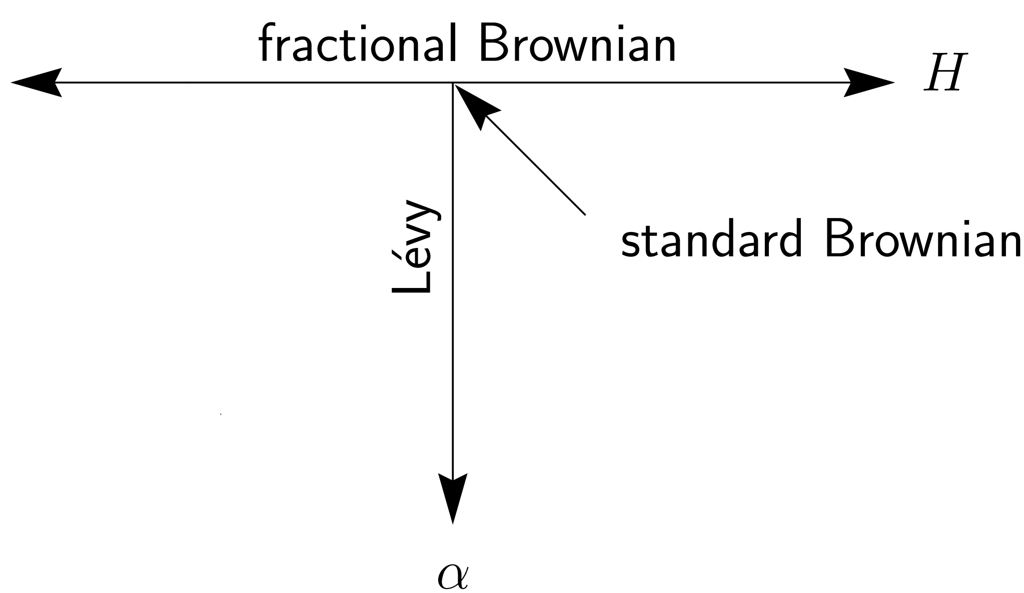
Brownian motions **1**
Self-affine 3
Universal 4
Lévy flight **5**
Finite-size effects: A tail of tails 7
Data analysis: Zipf plot 9
Test: Synthetic Lévy series 12
Fractional Brownian motion **14**
Diffusion 15
Data analysis: Dispersion 16
Test: Synthetic fBm series 19
Empirical example(s) **20**
Swiss Franc versus U.S. Dollar exchange rate 20

[1] http://rikblok.cjb.net/lib/blok02.html

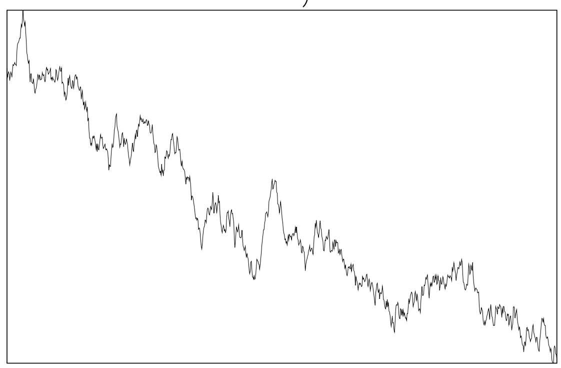
[2] mailto:rikblok@shaw.ca

Brownian motions

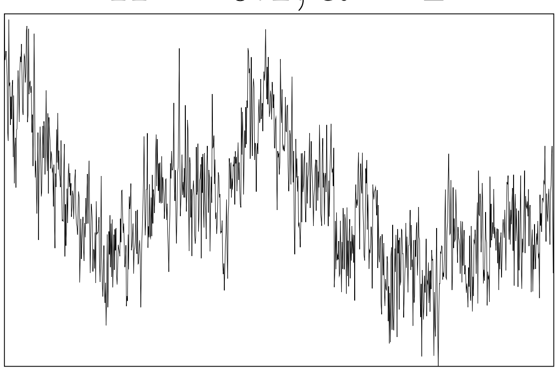
- *Standard Brownian motion* = uncorrelated Gaussian increments (finite variance), $\alpha = 2$, $H = 1/2$
- *Fractional Brownian motion (fBm)* = finite variance but correlations extend over entire history, $H \neq 1/2$ $(0 < H < 1)$
- *Lévy flight* = uncorrelated increments but divergent variance, $\alpha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



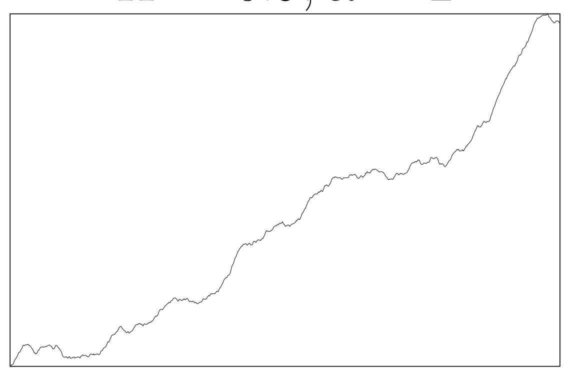
$H = 0.5, \alpha = 2$



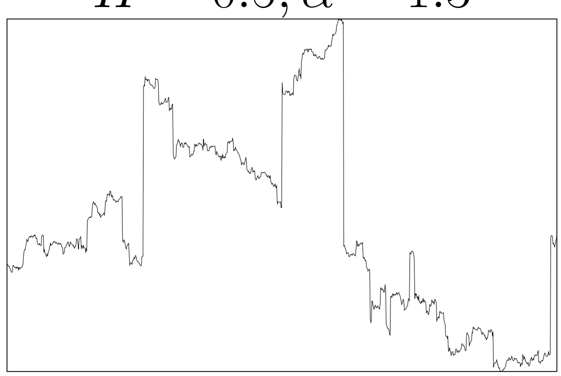
$H = 0.1, \alpha = 2$



$H = 0.9, \alpha = 2$

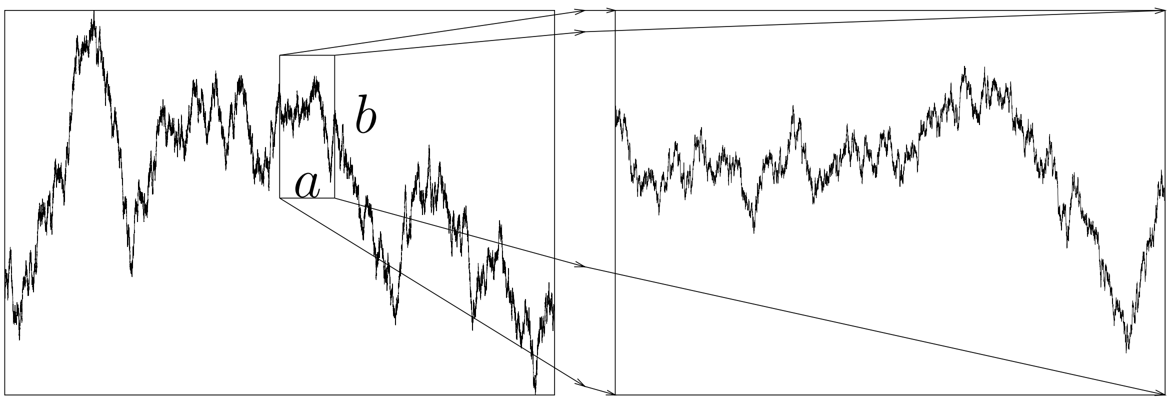


$H = 0.5, \alpha = 1.3$



Self-affine

If timescale zoomed by factor a then process looks statistically identical by scaling series by:
fBm $\Rightarrow b = a^H$
Lévy $\Rightarrow b = a^{1/\alpha}$



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 $1/\alpha$.

Universal

Q: Why should we be interested in these series?

A: Are observed in natural and man-made systems. Parameters H and α are often *universal*, independent of system details.

Sorry: I had planned to have some examples here but ran out of time. I hope to include some before uploading to the web.

My purpose is to explain how to obtain parameters from empirical data.

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.

Challenge: Power law tail derives from low frequency limit of Fourier representation (I think). A higher order expansion would be *invaluable*!

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

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

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...

Finite-size effects: A tail of tails

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

Since variance finite, should obey Central Limit Theorem on largest scales.

Can account for this by modifying the fitting function,

$$C_{\pm}(r) \sim \frac{f(r/r_c)}{|r|^{\alpha}}, \tag{5}$$

where new parameter r_c indicates onset of finite-size effects.

Need to fit on log-scale to emphasize tail.

For normal distribution

$$\log f(x) = \log\left\{\frac{1}{2}\left[1-\operatorname{erf}(x/\sqrt{2})\right]\right\} \tag{6}$$

$$\approx -\log 2 + \sqrt{\frac{2}{\pi}}\,x + \mathcal{O}(x^2) \tag{7}$$

which should describe cutoff of tails. (*This is an argument, not a proof.*)

So good fitting function is [1]

$$\boxed{\log C_{\pm}(r) = -\alpha \log|r| - \sqrt{\frac{2}{\pi}}\,\frac{|r|}{r_c} + \gamma,} \tag{8}$$

because fit is linear in parameters (α, β, γ) if we use $\beta = 1/r_c$. (γ is normalization constant.)

Effect of finite system size is to truncate power law tail by an exponential cut-off. Well established, empirically.

Data analysis: Zipf plot

Good choice because the Zipf plot method will not be tricked by other self-affine signals, like fBm.

Recipe

1. Rank order increments r.
2. Zipf plot.
3. Fit curve.
4. Interpret results.

1. Rank order increments r

Difference series $r(t) = x(t) - x(t - \Delta)$ at highest possible resolution Δ. If discrete process may be better to sample every event instead of regular intervals [2, 3, 4, App. B].

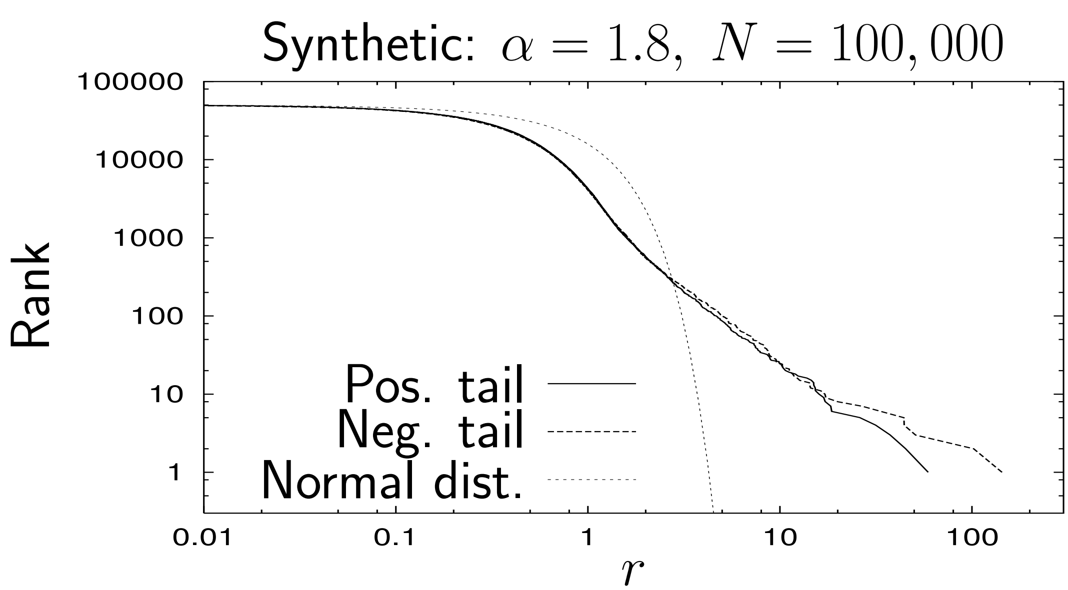
Need $N > 1,000$ data points. Prefer $N > 10,000$.

Trick: Center increments around median
$r(t) = r(t) - \mathrm{med}\,[r]$. Helpful if you suspect $\alpha < 1$.

Rank order (sort) increments in both increasing/decreasing orders (to analyze both tails).

2. Zipf plot

For each sort direction plot Rank vs. $|r|$ on log-log scale. (Traditional Zipf plot is transposed $x \leftrightarrow y$.). Eg.



3. Fit curve

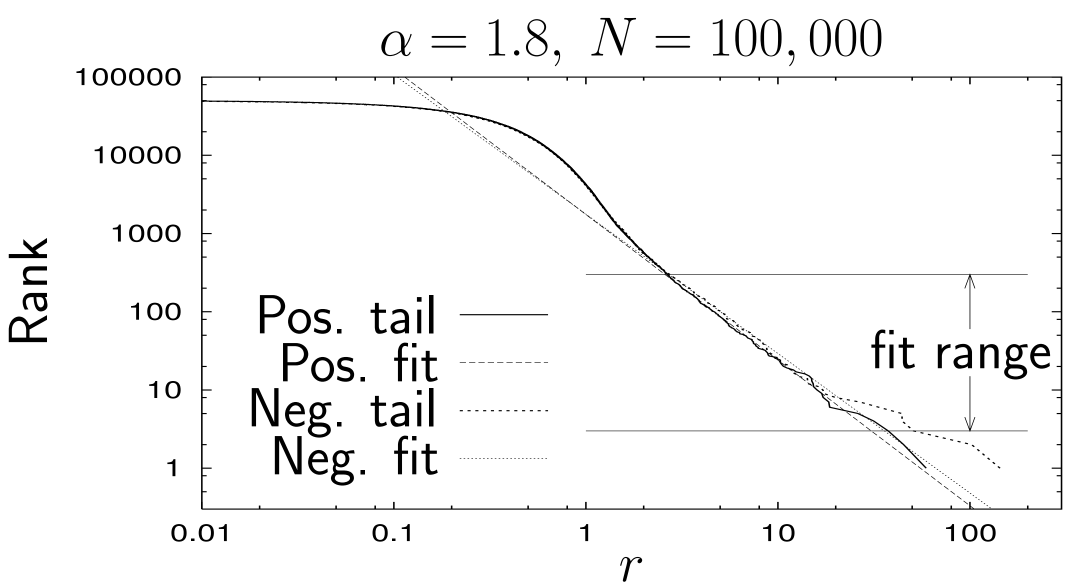
As N increases should get $\mathrm{Rank}_i \to NC(r_i)$ for each tail.

So can fit

$$-\alpha \log r - \sqrt{\frac{2}{\pi}}\frac{r}{r_c} + \gamma \tag{9}$$

to $\log \mathrm{Rank}$ via parameters α, r_c and γ.

Only want to fit over tail. Rule of thumb: fit over y-range $3 < \mathrm{Rank} < 0.003N$ (works for $0.4 \lesssim \alpha \lesssim 1.9$). Eg.



4. Interpret results

Must have $0 < \alpha < 2$ [5].

Power law (straight line on log-log graph) must hold over at least one decade (x-axis) to be significant.

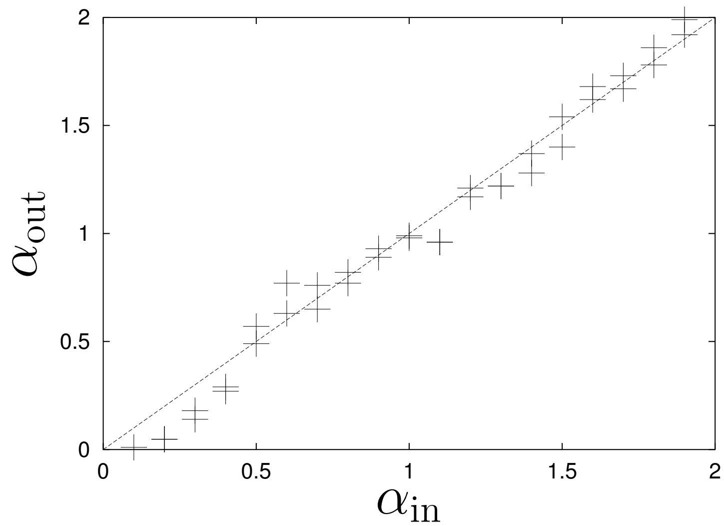
If either condition fails then Lévy tail not significant so take $\alpha \equiv 2$ (Gaussian).

Test: Synthetic Lévy series

Synthesized timeseries of 100,000 datapoints for $\alpha = 0.1, 0.2, \ldots, 2.0.$[3]

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

[3] Contact me (rikblok@shaw.ca) if you want to know how to generate synthetic Lévy or fBm datasets.

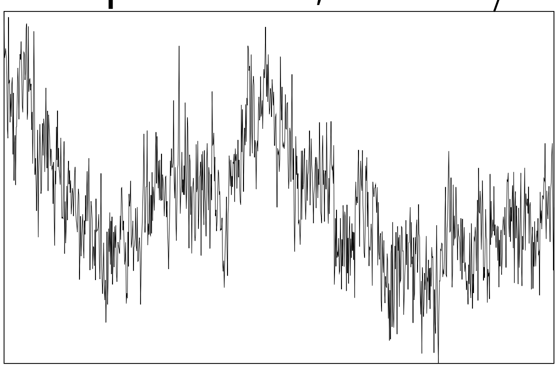


Each fitted a power law over at least one decade except $\alpha_{\rm in} = 2$ (Gaussian; returned $\alpha_{\rm out} \approx 10$).

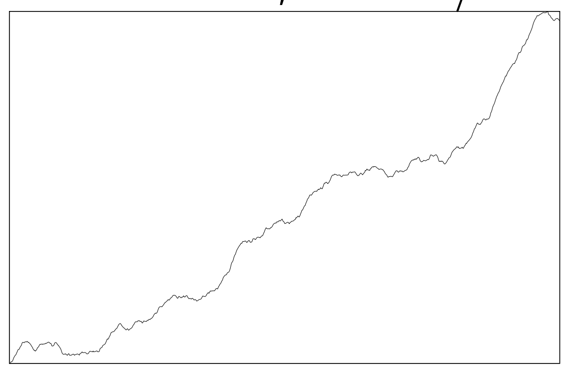
Also tested synthetic fBm series—always returned $\alpha_{\rm out} > 2$ indicating no Lévy tail.

Fractional Brownian motion

Antipersistent, $H < 1/2$



Persistent, $H > 1/2$

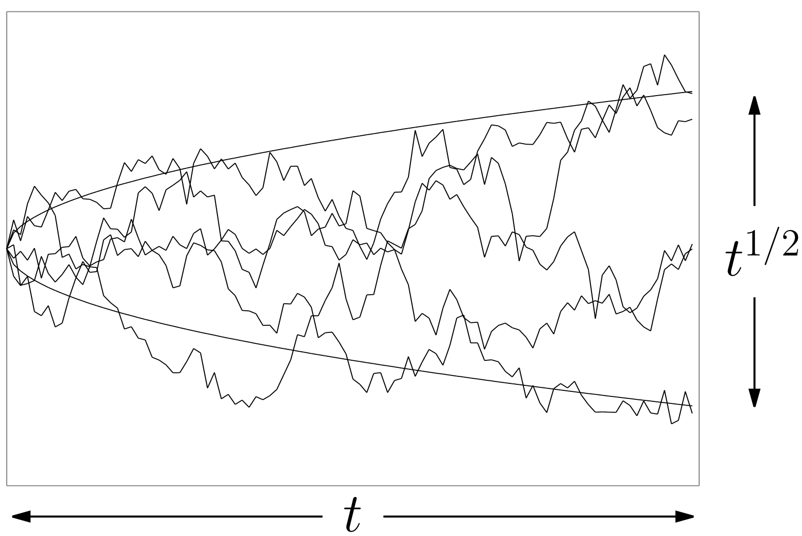


Fractal dimension, $D = 2 - H$, is space filled by signal.

Correlations extend over entire history of series.

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.

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 $H > 1/2$.

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

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

Recipe

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

1. Get increments r

Difference series $r(t) = x(t) - x(t - \Delta)$. Increments of fBm called fractional Gaussian noise (fGn).

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

2. Dispersional analysis

Average r over bins of length L (initially $L = 1$),

$$r_i^{(L)} = \sum_{j=iL}^{(i+1)L-1} r_j. \tag{10}$$

Estimate of diffusion on scale L is given by[4] (*trust me!*)

$$\sigma(L) \sim L\sqrt{\mathrm{Var}\left[r^{(L)}\right]}. \tag{11}$$

[4] Normally, dispersional analysis doesn't multiply by L so the slope of the graph is $H - 1$.

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

Double bin size $L \rightarrow 2L$ and repeat.

3. Fit curve

On log-log scale should have linear relationship

$$\log \sigma(L) = H \log L + C \tag{12}$$

where slope is Hurst exponent H.

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

4. Interpret results

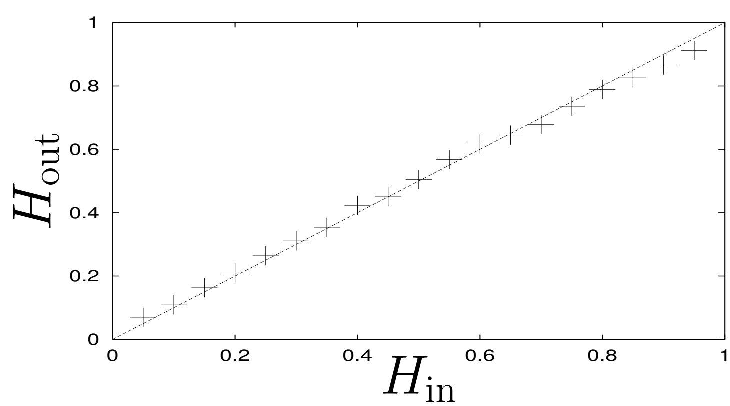
Only thing to be aware of is that series might be multi-fractal, with distinct H values on different L-scales.

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

Test: Synthetic fBm series

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

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



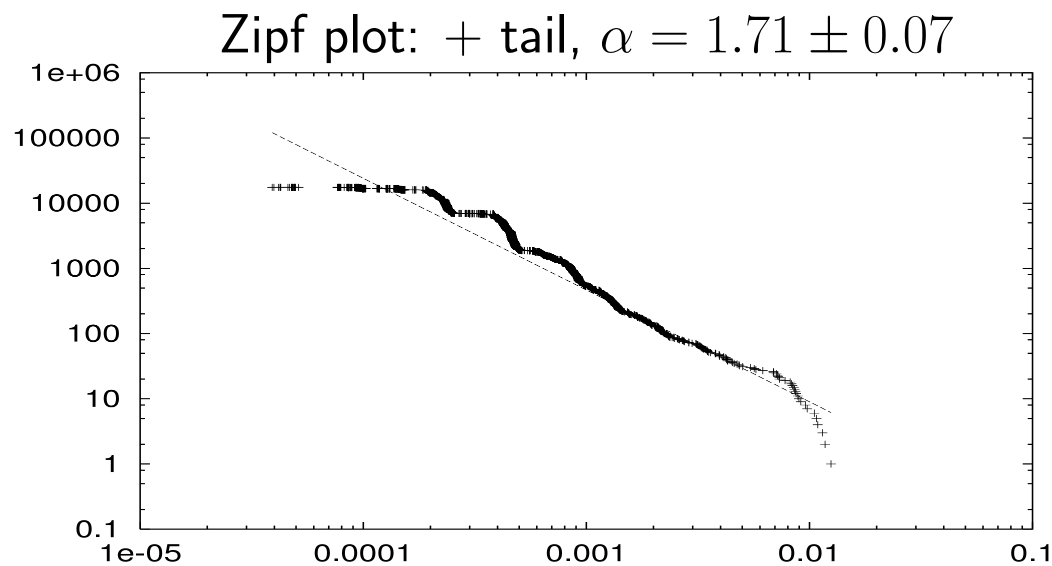
Also tested synthetic Lévy flight series. Returned $H = 0.49 \ldots 0.51$ for all α.

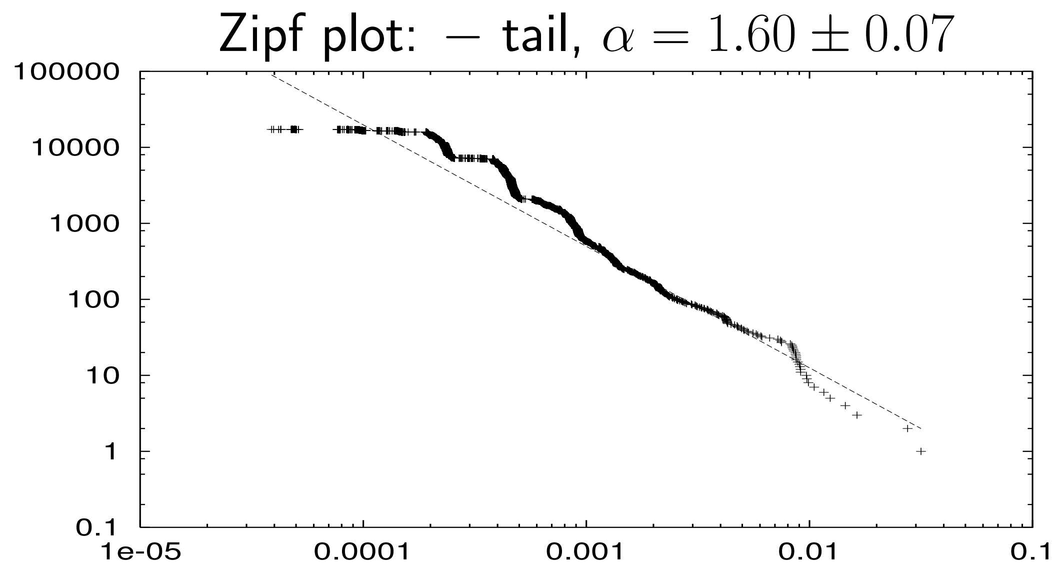
Empirical example(s)

Swiss Franc versus U.S. Dollar exchange rate

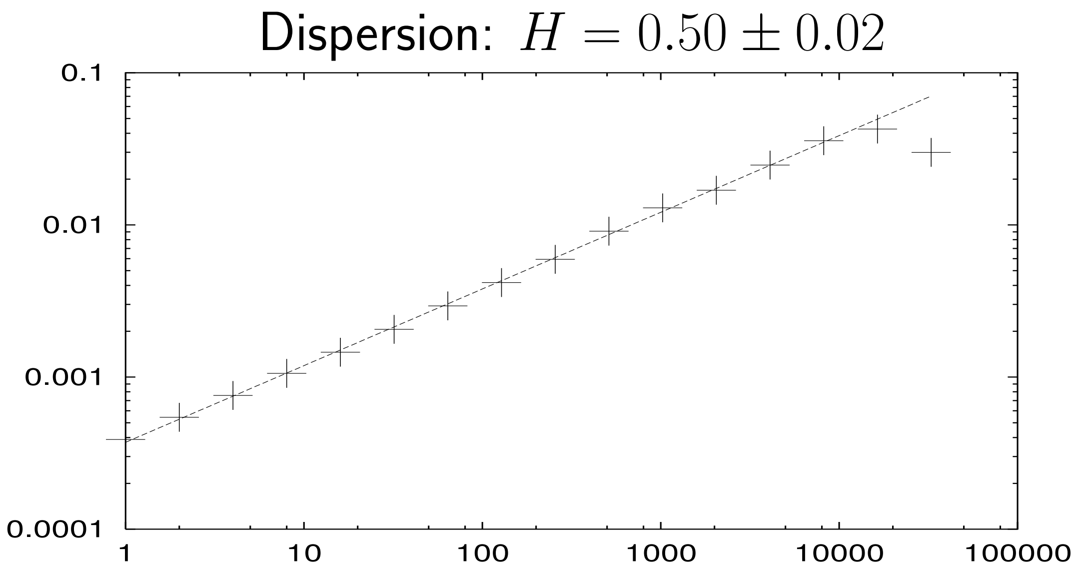
Tickwise data [6] sampled at 1 minute intervals. $N = 99,985$.

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





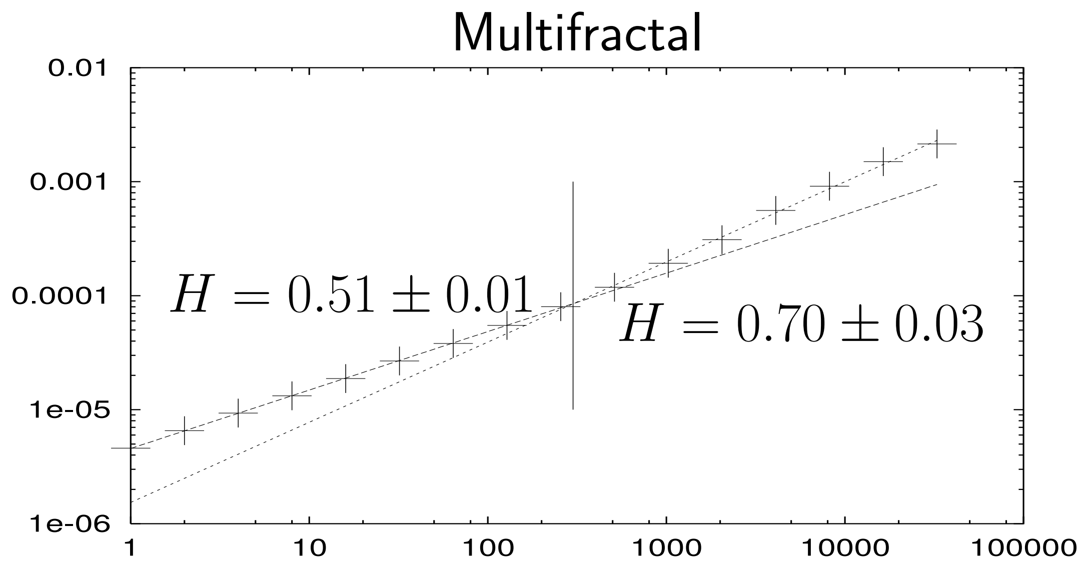
Evidence of a stable Lévy distribution with exponent $\alpha \approx 1.65$.



No memory in price/return history.

Volatility

However, consider squared returns r^2, known as volatility(?) [7, 8]. 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 exponent $\alpha = 1.6 - 1.7$, no memory in returns, but persistence in squared returns ($H = 0.7$) for timescales longer than ∼5 hrs.

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] Peter K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41:135–55, 1973.

[3] Zoltán Palágyi and Rosario N. Mantegna. Empirical investigation of stock price dynamics in an emerging market. *Physica A*, 269:132–9, 1999.

[4] Hendrik J. Blok. *On the nature of the stock market: Simulations and experiments*. PhD thesis, University of British Columbia, 2000. http://rikblok.cjb.net/lib/blok00b.html, arXiv:cond-mat/0010211.

[5] 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.

[6] 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.

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

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