Semiparametric bootstrap approach to hypothesis tests and confidence intervals for the Hurst coefficient

Peter Hall[1] Wolfgang Härdle[2]

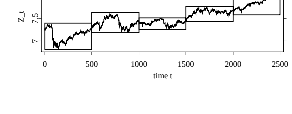
Torsten Kleinow[3] Peter Schmidt[4]

[1] Centre for Mathematics and its Applications,
Australian National University, Canberra

[2+3] Institute of Statistics and Econometrics,
Faculty of Economics and Business Administration,
Humboldt–University at Berlin

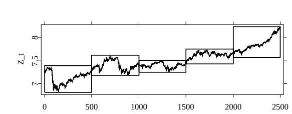
[2+3] Sonderforschungsbereich 373

[4] Quantitative Research,
Bankgesellschaft Berlin AG



Overview

- Motivation
- Self-Similar processes, Long Memory
- Statistical methodology (Model, R/S Analysis)
- Distribution of $\hat{H} - H$
- Bootstrap algorithm
- Application to financial data

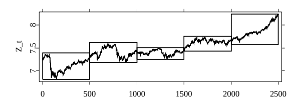


Motivation

- observe discounted asset prices: $(S(n))_{n\in\{1,\dots,N\}}$

- define $X(n) := \log(S(n)) \qquad \forall n \in \{1,\dots,N\}$ and $\dot{X}(n) := X(n+1) - X(n) \qquad \forall n \in \{1,\dots,N-1\}$

- if $(\dot{X}(n))_{n\in\{1,\dots,N\}}$ is stationary, define $\forall n \in \{1,\dots,N-k\}$ Autocorrelation: $\rho(k) = Corr(\dot{X}(n+k), \dot{X}(n))$



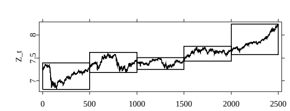
Model

- Model: $(X(n))_{n\in\{1,\dots,N\}}$ are discrete observations of a
  - Diffusion $(X(t))_{t\geq 0} \Longrightarrow \rho(k)=0$

$$\begin{aligned} dX(t) &= \mu(X(t),t)dt + \sigma(X(t),t)dB(t) \\ X(0) &= x \end{aligned}$$

  - self-similar process $(X(t))_{t\in\mathbb{R}}$ with stationary increments
    $\Longrightarrow$ for $N \longrightarrow \infty$ and $K \longrightarrow \infty$

$$\sum_{k=-K}^{K} \rho(k) \longrightarrow 0 \quad \text{or} \quad \sum_{k=-K}^{K} \rho(k) \longrightarrow \infty$$



Self-similar processes

self-similar:

$$c^{-H}(X(ct))_{t\in R} =_d (X(t))_{t\in \mathbb{R}} \qquad \forall c > 0$$

stationary increments:

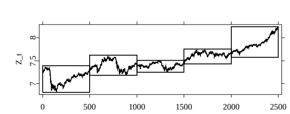
$$\rho(k) = \frac{1}{2}\{(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\}$$

for $0 < H < 1$, $H \neq \frac{1}{2}$

$$\frac{\rho(k)}{H(2H-1)k^{2H-2}} \to 1 \qquad k \longrightarrow \infty$$

$$\sum_{k=-\infty}^{\infty} \rho(k) = 0 \qquad \forall\ H \in \left(0, \frac{1}{2}\right)$$

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty \qquad \forall\ H \in \left(\frac{1}{2}, 1\right)$$



Statistical Methodology

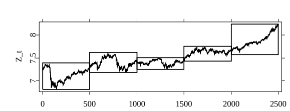
- $X(t) = g(Y(\epsilon t), t)$ for all $t \in \mathcal{I}$
  - Y is a Gaussian process whose sample path has fractal dimension $D = 2 - H$, $EY(t) \equiv 0$,

    $$E(Y(s+t) - Y(s))^2 = c|t|^\alpha + O\left(|t|^{\alpha+\beta}\right)$$

    uniformly in $s \in [0, 1]$ as $t \longrightarrow 0$ for constants $c > 0$, $\alpha = 2H \in \left(\frac{1}{2}, 2\right)$ and $\beta > min\left(\frac{1}{2}, 2 - \alpha\right)$.
  - g is a smooth function

    $$g_{j_1 j_2} := \frac{\delta g}{(\delta y)^{j_1} (\delta t)^{j_2}}$$

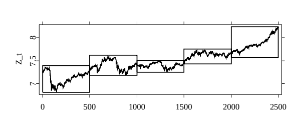
    are bounded for each $j_1, j_2 \geq 0$ and g_{10} does not vanish
  - H is called the Hurst coefficient.



R/S Analysis 1

- observations: $X(0), \ldots, X(N)$
- interval lengths $n \in \{l_1 m, \ldots, l_k m\} \subset I\!N$, $\max\{l_1 m, \ldots, l_k m\} \leq N$, $m \in I\!N$
- number of intervals of length n: $A := \frac{N}{n}$
- for every A and every $t \in [(a-1)n, an]$, $a = 1, \ldots A$ define:

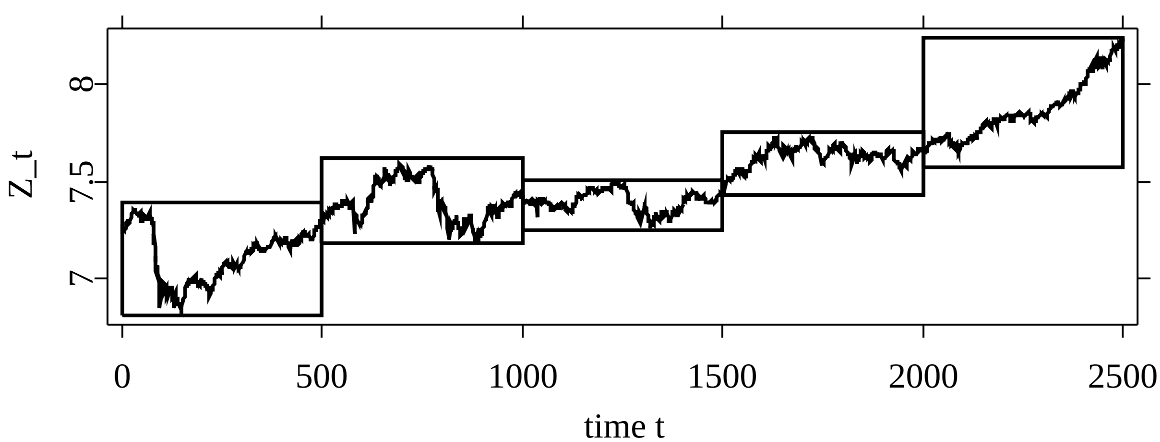
$$Z_t := X_t - \frac{t}{n}(X_{an} - X_{(a-1)n})$$

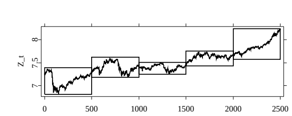


R/S Analysis 2

$$R_a = \max_{(a-1)n \le t \le an} Z_t - \min_{(a-1)n \le t \le an} Z_t$$

$$S_a := \sqrt{\frac{1}{n} \sum_{t=(a-1)n}^{an} (X_t - m_a)^2}$$



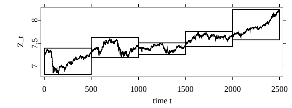


R/S Analysis 3

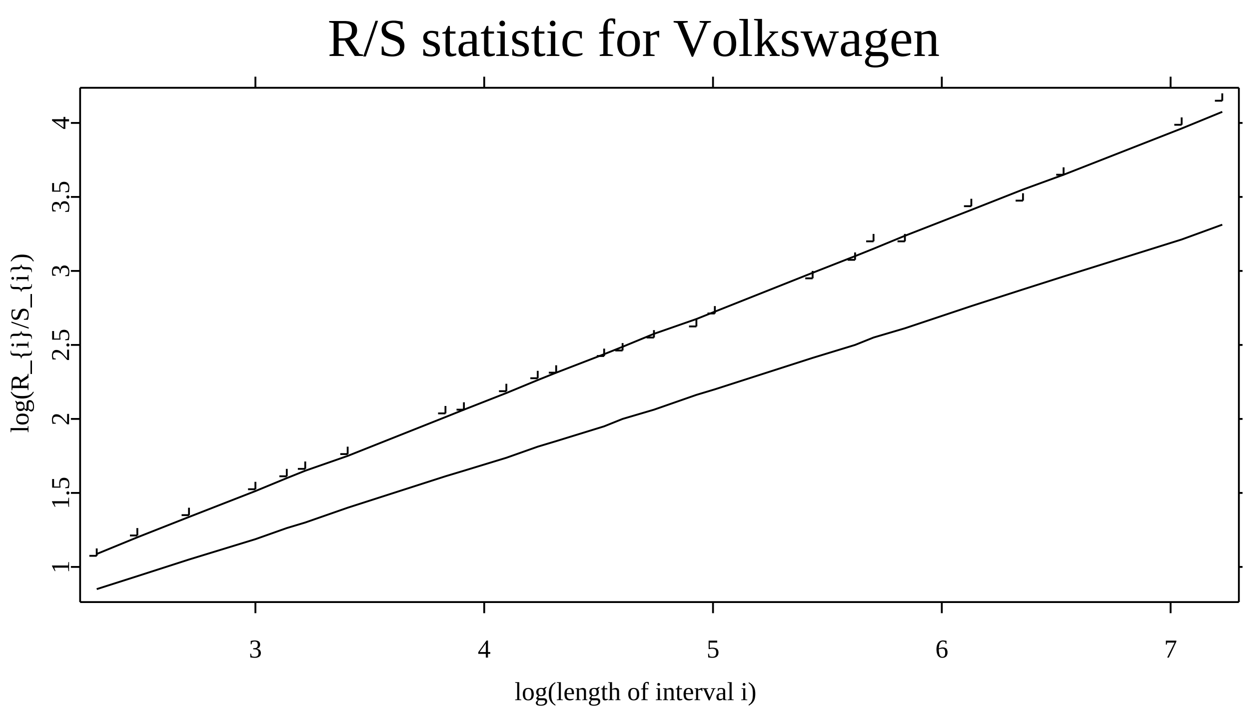
$$\left(\frac{R}{S}\right)_n := A^{-1} \sum_{a=1}^{A} \frac{R_a}{S_a}$$

define $\hat{H}$:

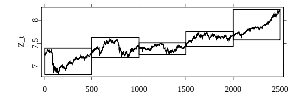
$$\log\left(\left(\frac{R}{S}\right)_n\right) = \hat{H} \log n + b$$



R/S Analysis - Example

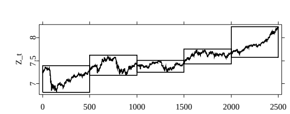


$$\hat{H} = 0.6049$$



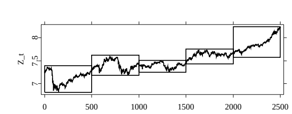
Distribution of $\hat{H} - H$

- Limiting distribution ($N \longrightarrow \infty$, for fixed k):
  - Normal: $0 < H \leq \frac{3}{4}$
  - Rosenblatt: $\frac{3}{4} < H < 1$
- Monte Carlo approach
  - greater accuracy then asymptotic approximation
  - stochastic fluctuation of the “S” part of the R/S Analysis influence the true distribution of $\hat{H}$ but not the limiting distribution



Classical bootstrap approach

- Simulation of a parametric model
  - The model is not completely known
- nonparametric bootstrap
  - the data are strongly dependend
  - in general it is not possible to find the iid "disturbances"



Distribution of $\hat{H} - H$

- assume $\epsilon := \frac{1}{N} \longrightarrow 0$ and $m = m(\epsilon) \longrightarrow \infty$ as $\epsilon \longrightarrow 0$ with $\frac{1}{m} + m\epsilon = O(\epsilon^{\alpha})$ $(\alpha > 0)$

- define: $\xi := m\epsilon$ and

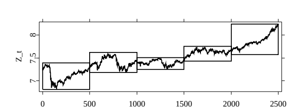
$$t_\xi = \begin{cases} \xi^{2(1-H)} & : \quad \text{if } 3/4 < H < 1 \\ \left(\xi \log \xi^{-1}\right)^{1/2} & : \quad \text{if } H = 3/4 \\ \xi^{1/2} & : \quad \text{if } 0 < H < 3/4\,, \end{cases}$$

Then: $\epsilon \longrightarrow 0 \quad \Longrightarrow \quad t_\xi \longrightarrow 0$

and

$$\hat{H} - H = t_\xi Z_\xi$$

where Z_ξ has a proper limiting distribution independent of g. $(X(t) = g(Y(\epsilon t), t))$



Bootstrap algorithm

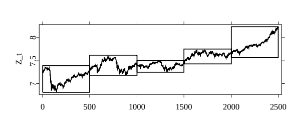
- estimate $\hat{H}$
- simulate a number of processes $(X(t)^*)_{t \in I\!R}$ with self-similarity parameter $\hat{H}$ and estimate $\hat{H}^*$ for every simulated process
- take the empirical distribution of $\hat{H}^* - \hat{H}$ as an approximation of $\hat{H} - H$

How to choose $(X(t)^*)_{t \in I\!R}$?

- structur of process $(X(t))_{t \in I\!R}$ does not influence the limiting distribution of $\hat{H}$

$\Longrightarrow$ it would be the same as for an elementary self-similar gaussian process

$\Longrightarrow$ fractional Brownian Motion



Financial Application

- Data: 6900 observations of 24 german stocks (included in DAX) from January 1973 to June 1999

- Test: $H_0 : H = \frac{1}{2}$ $\qquad$ $H_1 : H \neq \frac{1}{2}$

asset	$\hat{H}$	p-value
BMW	**0.5851**	**0.05**
Daimler	**0.5859**	**0.05**
Mannesmann	**0.5856**	**0.05**
Preussag	**0.5884**	**0.035**
Siemens	**0.6007**	**0**
Volkswagen	**0.6049**	**0**

$$\text{p-value} := P\left[|\hat{H}^* - E\hat{H}^*| > |\hat{H} - E\hat{H}^*| \,\Big|\, H_0 \right]$$