Analytical investigation on the detrended fluctuation analysis (DFA)

Marc Höll, Holger Kantz

Max Planck Institute for the Physics of Complex Systems, Dresden

10-Sept-2015

XXXV Dynamics Days Europe 2015

MAX-PLANCK-GESELLSCHAFT

Outline

Problem: Long range correlations vs non-stationarities in time series

Solution: Detrended fluctuation analysis (DFA)

This talk: Analytical investigation on the DFA

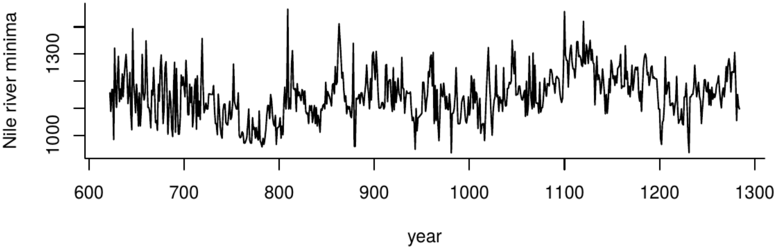
Long range correlations

- Long range memory is an important phenomenon found in many fields (hydrology, climate, finance, DNA sequencing, networks, ...)
- Autocorrelation function $C(s)$ decays slowly with time lag s
- Usually described by a power law

$$C(s) = \frac{\langle (x_t - \mu)(x_{t+s} - \mu) \rangle}{\langle (x_t - \mu)^2 \rangle} \sim s^{-\gamma}$$

  with correlation exponent $0 < \gamma < 1$ and ensemble mean μ
- In simple words: the process forgets its initial condition extremely slowly

- Visual characteristics of a series:
  - local trends
  - periods with small and large values
- Example: Yearly minimum water levels of the Nile River at the Roda Gauge from 622-1281 (Tousson, 1925, p. 366-385)



- First evidence of this phenomenon was found by HURST studying time series data of river flows (1951) → Hurst effect
- He used the rescaled-range analysis
- Other stationary methods:
  - Autocorrelation function analysis
  - Spectral Analysis
  - Fluctuation Analysis
- Two possible explanations for the Hurst effect:
  - Long range correlations (first theoretical model: fractional gaussian noise by MANDELBROT 1968)
  - Short range correlated model with additional non-stationarities (first investigated by BHATTACHARYA 1983)

Autocorrelation function analysis

- Example: AR(1) process with linear trend

$$x_t = \text{AR}(1) + bt$$

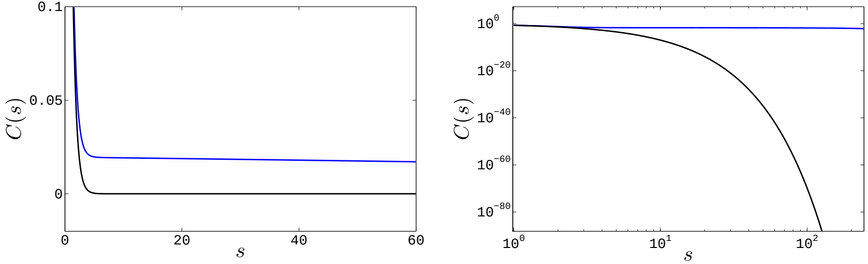
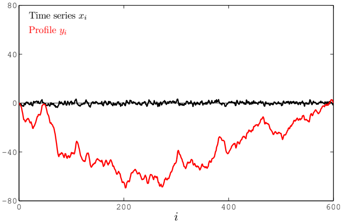


Figure : Stationary process (black) and with trend $b = 5 \cdot 10^{-4}$ (blue)

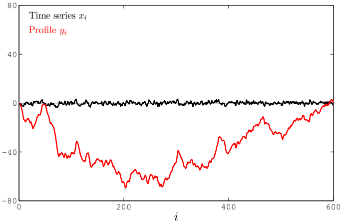
Problem: How to distinguish between long range correlations and non-stationarities in time series?

Detrended fluctuation analysis

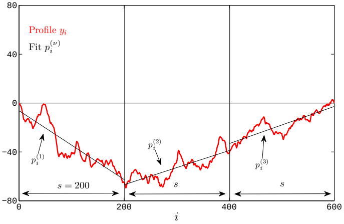
- Widely used method for the detection of long range correlations in non-stationary time series
- Invented by Peng et al. (1994)
- Applied to such diverse fields of interest as
  - DNA
  - human gait
  - heart rate dynamics
  - weather records
  - economical time series
  - ...



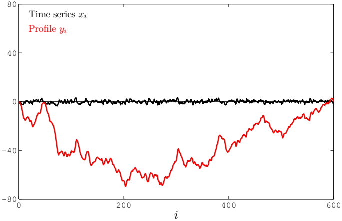
(1) Profile $y_i = \sum_{k=1}^{i} x_k$



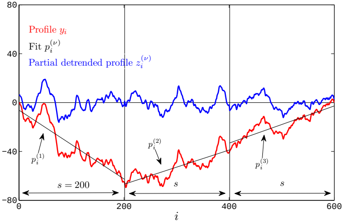
(1) Profile $y_i = \sum_{k=1}^{i} x_k$



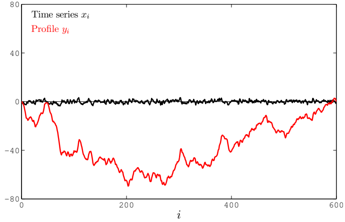
(2) Detrending $y_i - p_i^{(\nu)}$



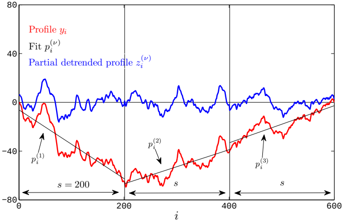
(1) Profile $y_i = \sum_{k=1}^{i} x_k$



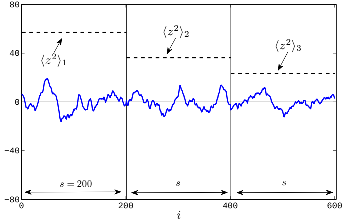
(2) Detrending $y_i - p_i^{(\nu)}$



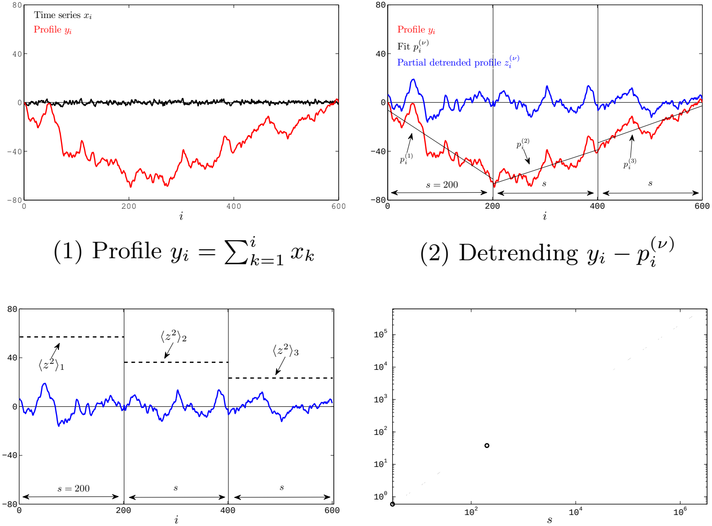
(1) Profile $y_i = \sum_{k=1}^{i} x_k$



(2) Detrending $y_i - p_i^{(\nu)}$



(3) $1/s \sum_{i \in \nu} \left(y_i - p_i^{(\nu)} \right)^2$

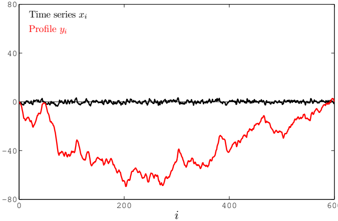


(1) Profile $y_i = \sum_{k=1}^{i} x_k$

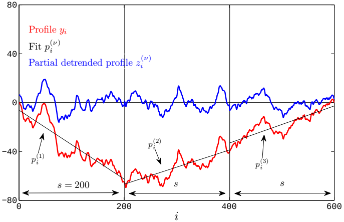
(2) Detrending $y_i - p_i^{(\nu)}$

(3) $1/s \sum_{i \in \nu} \left(y_i - p_i^{(\nu)} \right)^2$

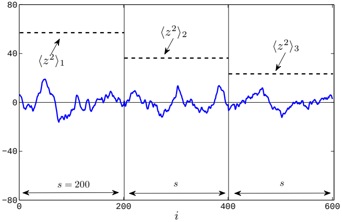
(4) Fluctuation function
$F^2(s) = 1/K \sum_{\nu=1}^{K} (3)$



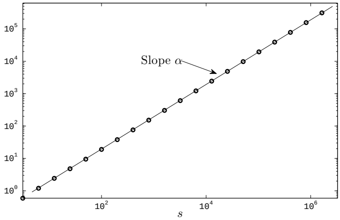
(1) Profile $y_i = \sum_{k=1}^{i} x_k$



(2) Detrending $y_i - p_i^{(\nu)}$



(3) $1/s \sum_{i \in \nu} \left(y_i - p_i^{(\nu)} \right)^2$



(4) Fluctuation function
$F^2(s) = 1/K \sum_{\nu=1}^{K} (3)$

Overview

- Fluctuation function of time series x_t

$$F^2(s) = \frac{1}{K} \sum_{\nu=1}^{K} \left(\frac{1}{s} \sum_{t=1+(\nu+1)s}^{\nu s} (y_t - p_t)^2 \right)$$

- Asymptotic behaviour and fluctuation parameter

$$F(s) \sim s^{\alpha}$$

- Relationship with Autocorrelation function
    - $\alpha = 1/2$ $\Rightarrow$ short memory $C(s) \sim e^{-s/s_c}$
    - $1/2 < \alpha < 1$ $\Rightarrow$ long memory $C(s) \sim s^{-\gamma}$ and $\alpha = 1 - \gamma/2$

Analytical investigation

- Idea: (1) Use original process x_t instead of y_t and p_t. (2) Rewrite
- It is

$$F^2(s) = \frac{1}{s} \sum_{t=1}^{s} \left\langle \left(\sum_{k=1}^{s} x_k \mathcal{P}_{t,k} \right)^2 \right\rangle$$

- For stationary processes

$$F^2(s) = \langle x^2 \rangle \left(W(s) + \sum_{r=1}^{s-1} C(r) L_r(s) \right)$$

with $W(s) \sim s$ and $L_r(s)$ a polynomial in r and s

- Detrending order 0

$$W(s) = \frac{s^2 - 1}{6s}$$

$$L_r(s) = \frac{1}{3s^2}\Big(- r^3 + 3r^2 s + (-3s^2 + 1)r + s^3 - s \Big)$$

- We calculated explicitly $L_r(s)$ for detrending orders 0, 1 and 2

Application

- Analytical calculation of
  - The fluctuation function $F(s)$
  - Additional polynomials

$$F^2(s) = {F_x}^2(s) + {F_{polyn}}^2(s)$$

- Asymptotic behaviour of $F(s)$
- Crossover point of short memory processes

Analytical calculation of the fluctuation function

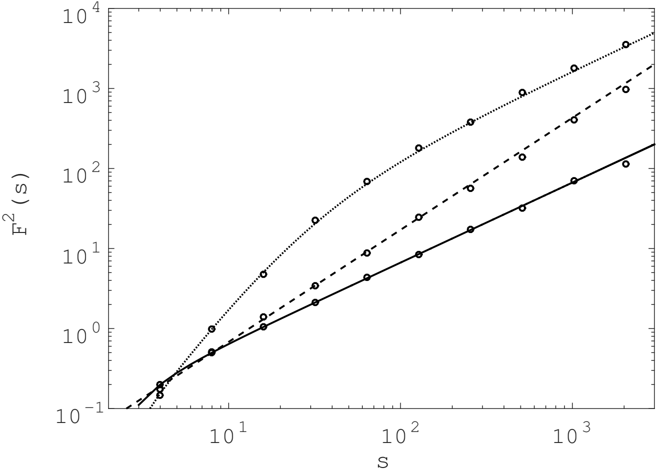


Figure : AR(1) with $a = 0.8$ (dotted), ARFIMA(0,0.2,0) (dashed) and White Noise (straight)

Asymptotic behaviour

- We can show that

$$\sum_{r=1}^{s-1} C(r) L_r(s) \sim \begin{cases} s & \text{for short memory} \\ s^{2-\gamma} & \text{for long memory} \end{cases}$$

  $\Rightarrow$ We can verify the expected parameters
- For large enough s the fluctuation function indicates the correlation
- There is a crossover for short memory processes, but not for long memory

- The crossover point for short memory processes is

$$s_\times = c_q \frac{e^{1/s_c}}{e^{2/s_c} - 1}$$

with $c_0 = 6$, $c_1 = 15$ and $c_2 = 70/3$

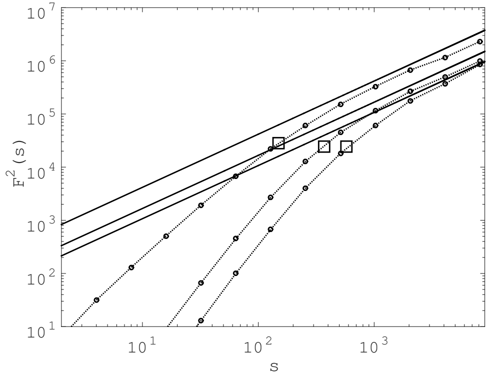


Figure : AR(1) with $a = 0.95$ and different orders of detrending q

Comparison of crossover and correlation time

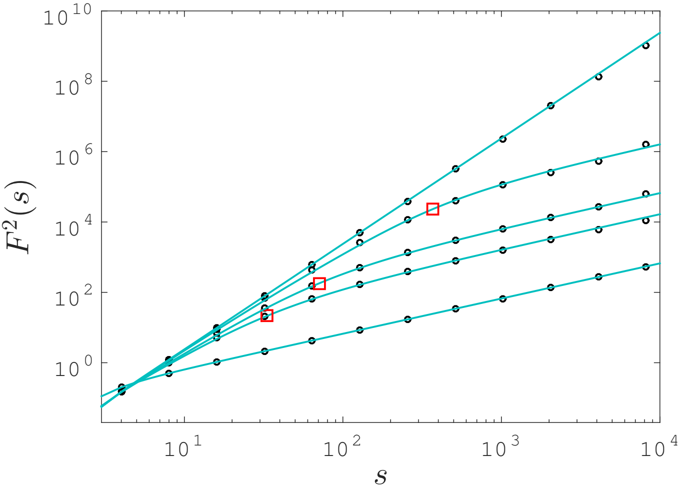


Figure : AR(1) with different s_c (from bottom to top: 0, 4.5, 9.5, 49.5 and random walk) with $q = 1$

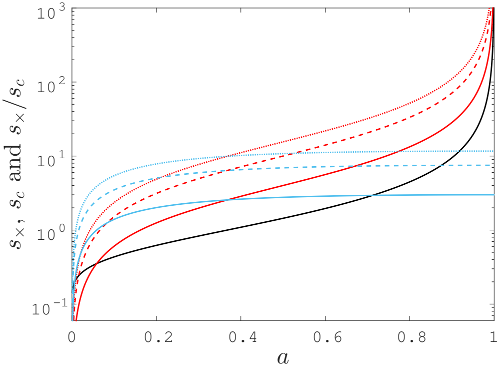


Figure : crossover points $s_\times$, correlation time of AR(1), their ratios

- DFA needs about $(5 \text{ to } 10)c_q$ $(= 30 - 60, 75 - 150, 117 - 233, \dots)$ more data for detecting s_x than a direct evaluation of s_c would need:
  - Disjunct vs overlapping intervals
  - Ratio $s_x/s_c \approx c_q/2 > 1$
- DFA is data consuming

Thank You!

- M. Höll, H. Kantz, European Physical Journal B **88**, 126 (2015)
- M. Höll, H. Kantz, submitted to European Physical Journal B