

Time Series Analysis

Chapter 8: Long memory and ARFIMA



Daniel Traian PELE

Bucharest University of Economic Studies
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Romanian Academy, Institute for Economic Forecasting
MSCA Digital Finance

Acronyms used in this material (1/2)

- **ACF**: Autocorrelation Function
- **ADF**: Augmented Dickey–Fuller (test)
- **AI**: Artificial Intelligence
- **AIC**: Akaike Information Criterion
- **AR**: AutoRegressive (model)
- **ARCH**: AutoRegressive Conditional Heteroskedasticity
- **ARFIMA**: AutoRegressive Fractionally Integrated Moving Average (model)
- **ARIMA**: AutoRegressive Integrated Moving Average (model)
- **ARMA**: AutoRegressive Moving Average (model)
- **BET**: Bucharest Exchange Trading (index)
- **BIC**: Bayesian Information Criterion
- **BNR**: Banca Națională a României (the National Bank of Romania)
- **CPI**: Consumer Price Index
- **DFA**: Detrended Fluctuation Analysis
- **EODHD**: EOD Historical Data (market-data provider)
- **FFT**: Fast Fourier Transform

Acronyms used in this material (2/2)

- **FIGARCH**: Fractionally Integrated GARCH
- **FRED**: Federal Reserve Economic Data
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **GPH**: Geweke–Porter-Hudak (estimator)
- **HAR**: Heterogeneous AutoRegressive (model)
- **HICP**: Harmonised Index of Consumer Prices
- **IGARCH**: Integrated GARCH
- **LW**: Local Whittle (estimator)
- **MA**: Moving Average
- **ML**: Machine Learning
- **OLS**: Ordinary Least Squares
- **RMS**: Root Mean Square
- **RMSE**: Root Mean Squared Error
- **RV**: Realised Volatility
- **SD**: Standard Deviation
- **SE**: Standard Error
- **US**: United States

Today's question and route

- **Question**: how long does a shock stay in a series? ARMA models (Chapter 2) forget at an exponential rate; some series forget much more slowly
  - examples: the Nile floods, inflation, unemployment, the volatility of stock returns
- **Route** of the chapter
  - short and long memory: exponential against hyperbolic decay of the ACF
  - fractional differencing $(1-L)^d$ and the ARFIMA(p,d,q) model
  - estimating d: R/S, DFA, GPH, local Whittle, maximum likelihood; forecasting
  - long memory in volatility; spurious long memory from breaks and regimes
- We build on Chapter 1 (ACF, spectrum), Chapter 3 (unit roots) and Chapter 5 (GARCH)

Learning outcomes

- Distinguish short from long memory by the decay of the ACF and by the spectrum near frequency zero
- Compute the weights of $(1-L)^d$ and the autocorrelations of ARFIMA$(0, d, 0)$
- State the stationarity and invertibility ranges of d and the link $H = d + 1/2$
- Estimate d with R/S, DFA, GPH, local Whittle and maximum likelihood, and judge their uncertainty
- Forecast with ARFIMA and compare it fairly with ARMA benchmarks
- Recognise spurious long memory caused by breaks and regime shifts

Reading and tools

- Textbook background: Huang and Petukhina (2022) (ARIMA and GARCH models, the building blocks of this chapter)
  - theory: Brockwell and Davis (1991), Sec. 13.2 (long memory processes); Beran et al. (2013); survey: Baillie (1996)
  - history: Graves et al. (2017), "A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA"
- Python Quantlets of this chapter: Quantlets/Ch_08
  - the estimators are written out in the code (R/S, DFA, GPH, local Whittle, exact ML); `statsmodels` has no ARFIMA class; FIGARCH comes from the `arch` package
- Lecture notebook: open in Google Colab
- Video course: Applied Time Series Analysis with Python

Harold Edwin Hurst and the Nile



H. E. Hurst (1880–1978)

Photo: Elliott & Fry (1953); public domain; Wikimedia Commons

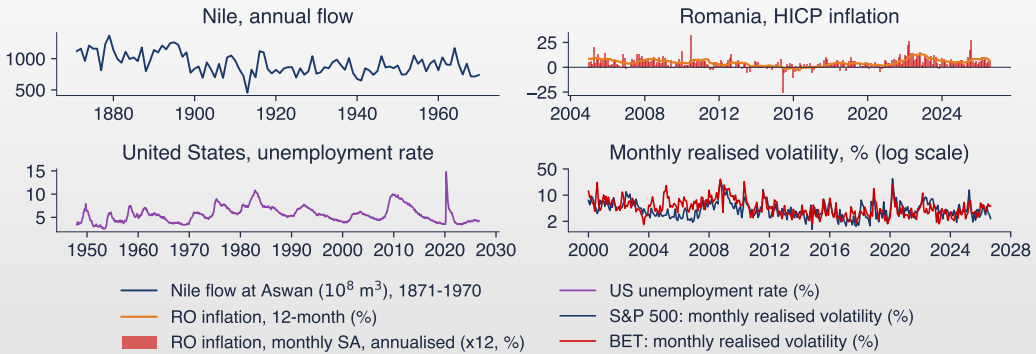
- Hurst spent more than 60 years in Egypt measuring the Nile, to size the reservoirs of the Aswan dams
  - a reservoir must cover the longest run of dry years: its size depends on the **range** of cumulative inflows
- Hurst (1951): the range grows like n^H with $H \approx 0.7$, not like $n^{0.5}$ as for independent years
  - wet years cluster with wet years and dry with dry, over decades: the "Joseph effect" of Mandelbrot and Wallis (1968)
- This is **long memory**: a dependence that fades too slowly for any low-order ARMA model



The Aswan Low Dam (1902), on the Nile

Photo: Karelj (2010); public domain; Wikimedia Commons

Four persistent series



- Nile flow at Aswan, 1871–1970 (statsmodels); Romanian HICP inflation, 12-month and monthly, seasonally adjusted (SA) with the monthly means (Eurostat)
- US unemployment rate (FRED UNRATE); monthly realised volatility of daily S&P 500 and BET returns (EODHD)

TSA_ch8_memory_data

Interpreting the four series

- Each series stays above or below its mean for long stretches
  - Nile: mean 919 (10^8 m^3), long dry spells after 1900; Romanian 12-month inflation peaked at 13.6% in November 2022 and was 6.1% in August 2026
- US unemployment rises fast in recessions (peak 14.8% in April 2020) and falls back slowly; latest 4.2% (September 2026)
  - realised volatility: calm and turbulent years alternate; maxima 27% (S&P 500, March 2020) and 27% (BET, October 2008)
- The question of the chapter: is this a slowly decaying (long) memory, a near unit root, or a sequence of regimes?

Short memory

- Recall (Chapter 1): $\rho(k) = \gamma(k)/\gamma(0)$, the autocorrelation at lag k of a stationary series
  - **short memory**: the autocorrelations are absolutely summable, $\sum_{k=0}^{\infty} |\rho(k)| < \infty$
- Stationary ARMA processes: $|\rho(k)| \leq Cr^k$ with $0 < r < 1$: **exponential** decay
  - $C > 0$: a constant; r: the decay rate per lag (for an AR(1), $r = |\phi|$)
  - AR(1): $\rho(k) = \phi^k$; with $\phi = 0.9$, $\rho(50) = 0.005$: after 50 periods the shock is forgotten
- Spectral view: the spectral density $f(\lambda)$ is finite and positive at frequency $\lambda = 0$
  - $\lambda \in [0, \pi]$: the frequency in radians per period (a cycle of length $2\pi/\lambda$ periods)
  - $\gamma(k)$: the autocovariance at lag k
  - $f(0) = \frac{1}{2\pi} \sum_k \gamma(k)$: the long-run variance is finite; the variance of the sample mean falls like $1/T$

Long memory

- **Long memory** (long-range dependence): $\rho(k) \sim C\, k^{2d-1}$ as $k \to \infty$, with $0 < d < 1/2$
  - a **hyperbolic** (power-law) decay: the exponent $2d - 1$ lies in $(-1, 0)$
  - the sum $\sum_k \rho(k)$ diverges: distant observations still matter together
  - $\sim$: the ratio of the two sides tends to 1; $C, G, c > 0$: constants; d: the memory parameter
- Spectral view: $f(\lambda) \sim G\, \lambda^{-2d}$ as $\lambda \to 0$: a **pole** at frequency zero
  - the low frequencies (slow cycles of all lengths) carry most of the variance
- Consequence: $\mathrm{Var}(\bar{x}_T) \sim c\, T^{2d-1}$ ($\bar{x}_T$: the sample mean of T observations), slower than $1/T$
  - confidence intervals that assume independence are too narrow; the memory parameter d measures how slowly the past fades

Sample ACF against an AR(1)



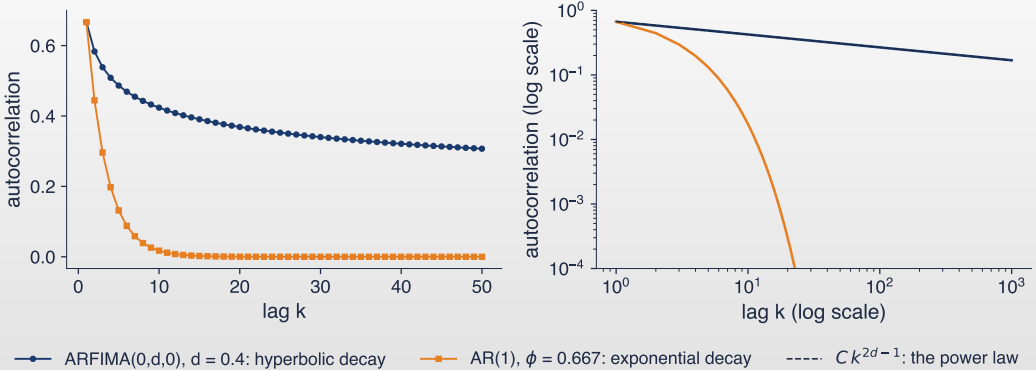
⊡ Bars: sample ACF; dashed line: $\hat{\rho}(1)^k$, the ACF of an AR(1) with the same lag-1 autocorrelation; shaded: $\pm 1.96/\sqrt{T}$

TSA_ch8_memory_data

Interpreting the sample ACF

- Nile: $\hat{\rho}(1) = 0.50$; an AR(1) would give $\rho(10) = 0.0009$, the data give 0.09
  - 11 of the first 30 autocorrelations lie above the white-noise band
- Romanian monthly inflation: $\hat{\rho}(1) = 0.44$, $\hat{\rho}(10) = 0.20$; 22 of 48 lags above the band
  - US monthly inflation: 48 of 48; S&P 500 log realised volatility: 15 of 48
- A low first autocorrelation with a long, slowly decaying tail is the signature of long memory; an AR(1) cannot produce it

Hyperbolic against exponential decay



- Theoretical ACF of ARFIMA$(0, d, 0)$ with $d = 0.4$ (next sections) and of an AR(1) with the same $\rho(1) = d/(1-d) = 0.667$; right: both on log–log axes, with the power law Ck^{2d-1}

TSA_ch8_arfima_processes

Interpreting the two decay laws

- Same start, different tails: at lag 10, 0.424 against 0.017; at lag 50, 0.307 against $1.6 \cdot 10^{-9}$
  - the long-memory ACF is still 0.169 at lag 1000
- On log–log axes the hyperbolic ACF becomes a straight line of slope $2d - 1 = -0.2$; the exponential ACF bends down
  - the slope of a log–log plot is the idea behind every estimator of this chapter
- Sums of autocorrelations: AR(1) $\sum_{k \geq 1} \phi^k = 2.0$; ARFIMA: 33.0 up to lag 100 and 210.4 up to lag 1000, still growing

Recap: Short and long memory

- Short memory: summable ACF, exponential decay, finite spectrum at zero (ARMA)
- Long memory: $\rho(k) \sim Ck^{2d-1}$, non-summable ACF, spectral pole $f(\lambda) \sim G\lambda^{-2d}$
- Hurst and the Nile: ranges grow like n^H, $H > 1/2$
- Inflation and volatility show long, slowly decaying ACF tails

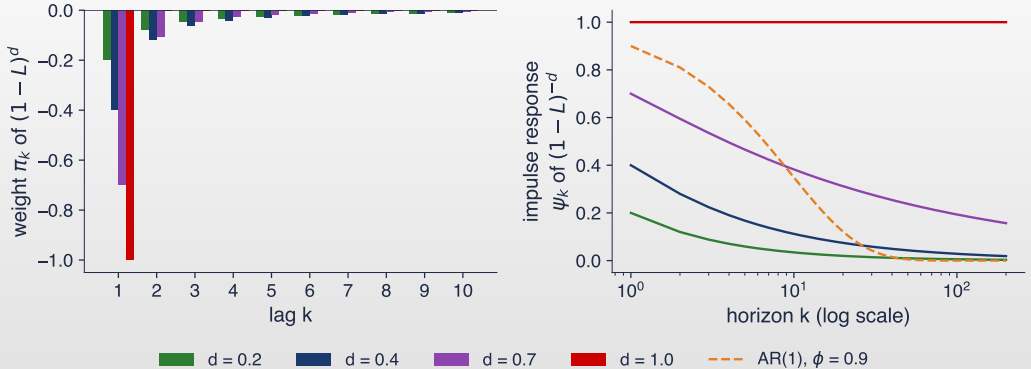
From integer to fractional differences

- Lag operator: $Lx_t = x_{t-1}$; first difference $(1-L)x_t = x_t - x_{t-1}$ (Chapter 3)
  - $d = 0$: the level; $d = 1$: the first difference; $d = 2$: $x_t - 2x_{t-1} + x_{t-2}$
- **Fractional difference** for any real d (binomial series)
  - $(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k}(-L)^k = \sum_{k=0}^{\infty} \pi_k L^k$
  - $(1-L)^d = 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - \dots$
  - recursion: $\pi_0 = 1$, $\pi_k = \pi_{k-1}\,\frac{k-1-d}{k}$; for large k: $\pi_k \approx -\frac{d}{\Gamma(1-d)}\,k^{-1-d}$
  - $\binom{d}{k} = d(d-1)\cdots(d-k+1)/k!$: the binomial coefficient for real d; π_k: the weight of x_{t-k}; $\Gamma(\cdot)$: the gamma function, $\Gamma(n) = (n-1)!$ for integers
- A fractional difference uses **all** past values, with weights that decay hyperbolically: it removes part of the memory, not all of it

Worked example: the weights of $(1-L)^{0.4}$

- Recursion $\pi_k = \pi_{k-1}(k-1-d)/k$ with $d = 0.4$:
  - $\pi_1 = 1 \cdot (0-0.4)/1 = -0.40$
  - $\pi_2 = -0.40 \cdot (1-0.4)/2 = -0.12$
  - $\pi_3 = -0.12 \cdot (2-0.4)/3 = -0.0640$; $\pi_4 = -0.0640 \cdot 3.6/4 = -0.0416$; $\pi_5 = -0.0300$
- So $(1-L)^{0.4} x_t = x_t - 0.4x_{t-1} - 0.12x_{t-2} - 0.064x_{t-3} - 0.0416x_{t-4} - \ldots$
  - the weights sum to $(1-1)^{0.4} = 0$ over all lags: a constant level is removed, as with $d = 1$
- Compare $d = 1$: $\pi_1 = -1$ and all other weights are 0; $d = 0.7$: $\pi_1 = -0.70$, $\pi_2 = -0.105$

Fractional differencing and fractional integration



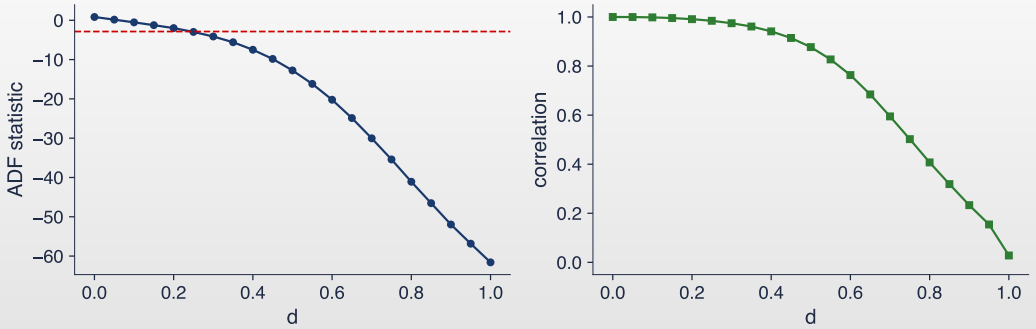
⊡ Left: weights π_k of $(1-L)^d$; right: the impulse responses ψ_k of $(1-L)^{-d}$, i.e. the effect after k periods of a unit shock to a fractionally integrated series, against an AR(1) with $\phi = 0.9$

TSA_ch8_fractional_differencing

Interpreting the weights

- Fractional integration: $x_t = (1-L)^{-d}\varepsilon_t = \sum_k \psi_k \varepsilon_{t-k}$, with $\psi_k = \psi_{k-1}(k-1+d)/k \approx k^{d-1}/\Gamma(d)$
  - after 100 periods a shock still has weight 0.028 ($d = 0.4$) and 0.193 ($d = 0.7$); the AR(1) keeps $2.7 \cdot 10^{-5}$
- Early on the AR(1) remembers more (0.349 at lag 10, against 0.112 for $d = 0.4$); later it forgets much faster
  - $d = 1$ (random walk): $\psi_k = 1$ forever, the shock never fades
- For $0 < d < 1$ the responses tend to zero: the series is **mean-reverting**, even when it is non-stationary ($d \geq 0.5$)

The degree of differencing a price needs



⊡ Log S&P 500, daily, 6 715 days since 2000; $(1-L)^d$ with a fixed window (weights below 10^{-4} dropped), $d = 0, 0.05, \ldots, 1$; left: ADF statistic (Chapter 3); right: correlation with the log level

TSA_ch8_fractional_differencing

Interpreting fractional differencing of a price

- The log level ($d = 0$) has a unit root (ADF 0.84); the return ($d = 1$) is stationary (ADF -61.6) but has correlation 0.03 with the level
  - the first difference throws away all the information about the level
- The ADF statistic crosses the 5% critical value (-2.86) already at $d = 0.25$, where the correlation with the level is still 0.984
  - the fixed window then has 445 weights
- A fractional difference can make a series stationary while keeping most of its memory: a tool used for forecasting features in machine learning (Chapter 9)

Recap: Fractional differencing

- $(1-L)^d = \sum_k \pi_k L^k$, $\pi_k = \pi_{k-1}(k-1-d)/k$: hyperbolically decaying weights
- $(1-L)^{-d}$: impulse responses $\psi_k \approx k^{d-1}/\Gamma(d)$; mean reversion for $d < 1$
- d between 0 and 1 fills the gap between levels and first differences

The ARFIMA(p, d, q) model

- **Definition**: $\phi(L)(1-L)^d(x_t - \mu) = \theta(L)\varepsilon_t$, ε_t white noise with variance σ^2
  - $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$, with roots outside the unit circle (Chapter 2)
  - μ: the mean of x_t; ARFIMA: autoregressive fractionally integrated moving average
- Two parts with two jobs
  - d: the **long-run** memory, the hyperbolic tail of the ACF
  - ϕ, θ: the **short-run** dynamics, the first few autocorrelations
- Special cases: $d = 0$: ARMA(p, q); $d = 1$: ARIMA$(p, 1, q)$ (Chapter 3); d between them: fractional integration, $x_t \sim I(d)$

Granger, Joyeux and Hosking (1980–1981)



Clive W. J. Granger (1934–2009), Nobel Prize 2003

Photo: Olaf Storbeck (2008); CC BY-SA 2.0; Wikimedia Commons

- Granger and Joyeux (1980) and Hosking (1981) introduced fractional differencing independently
  - Granger and Joyeux: aggregating many AR(1) series with different persistence produces long memory
  - Hosking: the ACF, the stationarity and invertibility conditions, and the ARFIMA(p, d, q) family
- Their model connects the Hurst exponent of hydrology with the Box–Jenkins models of econometrics
- Applications followed in inflation Hassler and Wolters (1995), output, interest rates and volatility Ding, Granger and Engle (1993)

The parameter d: what each range means

Range of d	Properties	ACF / shocks
$d \leq -0.5$	not invertible (no AR(∞) form)	–
$-0.5 < d < 0$	stationary, invertible, anti-persistent	negative, summable to $-1/2$
$d = 0$	ARMA, short memory	exponential decay
$0 < d < 0.5$	stationary, long memory	hyperbolic, not summable
$0.5 \leq d < 1$	non-stationary, mean-reverting	shocks fade, slowly
$d = 1$	unit root, ARIMA	shocks are permanent

- Stationarity needs $d < 1/2$, invertibility needs $d > -1/2$ (Hosking (1981)); a non-stationary series with $d < 1.5$ is handled by differencing once: $(1 - L)x_t \sim I(d - 1)$
- Hurst exponent of the stationary increments: $H = d + 1/2$; $H > 1/2$ persistence, $H < 1/2$ anti-persistence

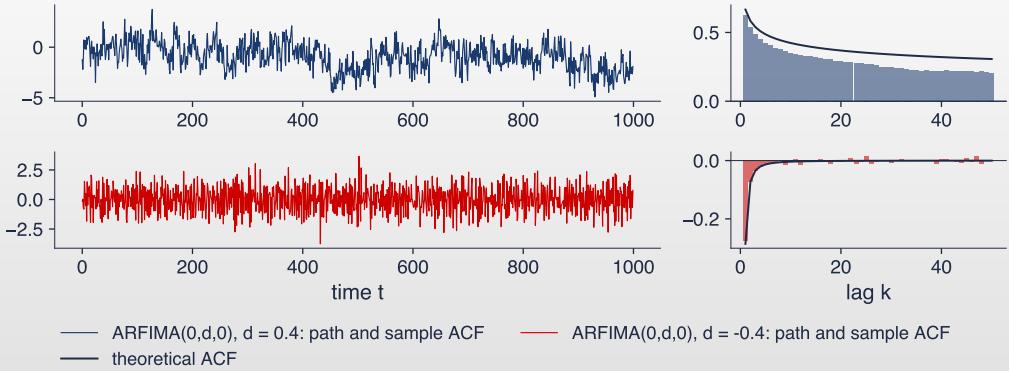
Autocorrelations of ARFIMA$(0, d, 0)$

- For $x_t = (1-L)^{-d}\varepsilon_t$, $-1/2 < d < 1/2$ (Hosking (1981)):
  - variance $\gamma(0) = \sigma^2 \dfrac{\Gamma(1-2d)}{\Gamma(1-d)^2}$
  - autocorrelations $\rho(k) = \dfrac{\Gamma(1-d)\,\Gamma(k+d)}{\Gamma(d)\,\Gamma(k+1-d)}$, computed with $\rho(k) = \rho(k-1)\,\dfrac{k-1+d}{k-d}$
- First lags: $\rho(1) = \dfrac{d}{1-d}$, $\quad \rho(2) = \dfrac{d(1+d)}{(1-d)(2-d)}$
  - large k: $\rho(k) \approx \dfrac{\Gamma(1-d)}{\Gamma(d)}\, k^{2d-1}$, the power law of Section 1 with $C = \Gamma(1-d)/\Gamma(d)$
- Spectral density (i: the imaginary unit): $f(\lambda) = \dfrac{\sigma^2}{2\pi}\,|1-e^{-i\lambda}|^{-2d} = \dfrac{\sigma^2}{2\pi}\,\big(2\sin(\lambda/2)\big)^{-2d} \approx \dfrac{\sigma^2}{2\pi}\lambda^{-2d}$ near 0

Worked example: ARFIMA$(0, 0.3, 0)$

- $\rho(1) = 0.3/0.7 = 0.429$; $\rho(2) = 0.429 \cdot 1.3/1.7 = 0.328$; $\rho(3) = 0.328 \cdot 2.3/2.7 = 0.279$
  - continuing the recursion: $\rho(5) = 0.228$, $\rho(10) = 0.173$
- Variance with $\sigma^2 = 1$: $\Gamma(0.4)/\Gamma(0.7)^2 = 1.316$, larger than the variance of the shocks
  - the memory accumulates past shocks
- An AR(1) with the same $\rho(1) = 0.429$ has $\rho(10) = 0.429^{10} = 0.0002$: three orders of magnitude below 0.173
- $H = d + 0.5 = 0.8$: a persistent series

ARFIMA$(0, d, 0)$ with $d = 0.4$ and $d = -0.4$



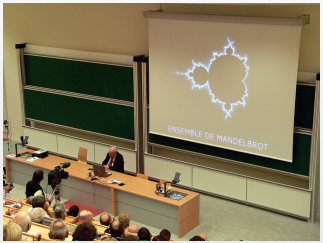
- Simulated paths (1000 observations, exact simulation by circulant embedding) with $\varepsilon_t \sim N(0, 1)$; right: sample ACF of 20 000 observations and the theoretical ACF

TSA_ch8_arfima_processes

Interpreting the two ARFIMA paths

- $d = 0.4$: long excursions above and below zero; the hyperbolic ACF of Section 1; variance 2.07
  - the path looks as if it had trends, although the mean is constant: a reason why long memory is mistaken for non-stationarity
- $d = -0.4$: rapid zig-zag; $\rho(1) = -0.286$, $\rho(2) = -0.071$; variance 1.18
  - anti-persistence: a rise tends to be followed by a fall; typical of over-differenced series, e.g. the difference of a series with $d = 0.6$

Mandelbrot: fractional Brownian motion and H

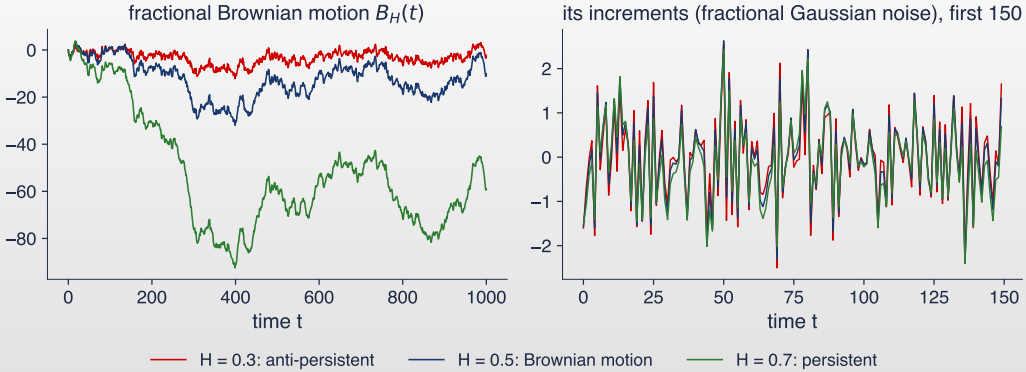


Benoît Mandelbrot (1924–2010)

Photo: David Monniaux (2006); CC BY-SA 3.0; Wikimedia Commons

- **Fractional Brownian motion** $B_H(t)$, $0 < H < 1$ Mandelbrot and Van Ness (1968): Gaussian, $B_H(0) = 0$, $\mathrm{Var}(B_H(t)) = t^{2H}$
  - self-similar: $B_H(ct)$ has the same distribution as $c^H B_H(t)$; $H = 1/2$: Brownian motion
- Its increments, **fractional Gaussian noise** (fGn), are stationary with $\rho(k) = \frac{1}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H})$
  - $\rho(k) \approx H(2H-1)k^{2H-2}$: the same tail as ARFIMA with $d = H - 1/2$
- Mandelbrot brought Hurst's finding into statistics: fGn is the continuous-time cousin of ARFIMA$(0, d, 0)$

Fractional Brownian motion for three Hurst exponents



- The same random numbers for $H = 0.3$, 0.5 and 0.7 (circulant embedding); right: the first 150 increments

TSA_ch8_arfima_processes

Interpreting the fBm paths

- $H = 0.7$: smooth, trending path; lag-1 autocorrelation of the increments $2^{2H-1} - 1 = 0.32$ (sample 0.32)
  - $H = 0.3$: rough path that keeps turning back; $\rho(1) = -0.24$ (sample -0.23)
- The range of the path grows like t^H: the quantity Hurst measured on the Nile
- In time-series modelling we prefer ARFIMA: it adds the short-run ARMA part that fGn lacks

Recap: The ARFIMA model

- $\phi(L)(1-L)^d(x_t - \mu) = \theta(L)\varepsilon_t$: d for the long run, ARMA for the short run
- Stationary for $d < 1/2$, invertible for $d > -1/2$; mean-reverting for $d < 1$
- ARFIMA$(0, d, 0)$: $\rho(1) = d/(1-d)$, $\rho(k) \sim Ck^{2d-1}$; $H = d + 1/2$
- fBm and fGn: the same long memory in Mandelbrot's continuous-time form

Five estimators

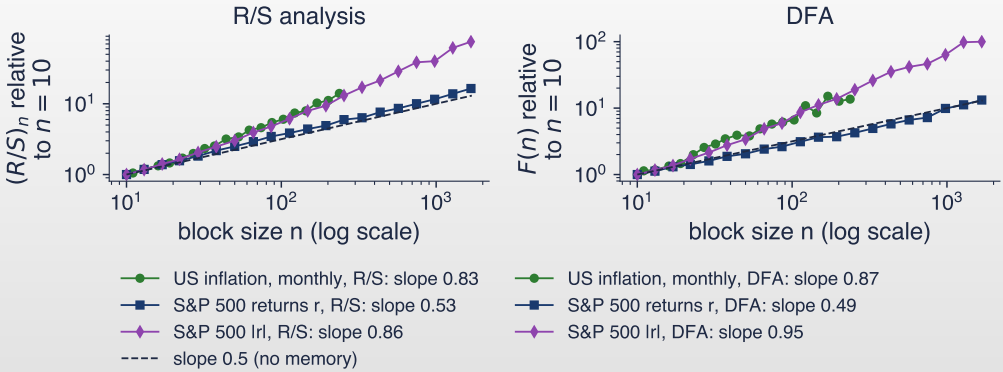
Estimator	Uses	Estimate	Type
R/S Hurst (1951)	range of partial sums in blocks	slope of $\log(R/S)_n$ on $\log n = H$	heuristic
DFA Peng et al. (1994)	detrended fluctuations of the profile	slope of $\log F(n)$ on $\log n = H$	heuristic
GPH Geweke and Porter-Hudak (1983)	lowest m periodogram ordinates	OLS slope of a log-periodogram regression	semiparametric
Local Whittle Robinson (1995)	lowest m periodogram ordinates	maximises a local Gaussian likelihood	semiparametric
Exact ML, Whittle Sowell (1992), Fox and Taqqu (1986)	the whole series	d, ϕ, θ, σ^2 jointly	parametric

- Semiparametric: only the behaviour near frequency zero is modelled, the short-run part is left free; parametric: the whole ARFIMA model must be right
- ML: maximum likelihood; OLS: ordinary least squares; $d = H - 1/2$ converts between the two scales

The R/S statistic, step by step

- Block $x_1, \ldots, x_n$: mean $\bar{x}$, cumulative deviations $Y_k = \sum_{i \le k}(x_i - \bar{x})$
  - range $R = \max_k Y_k - \min_k Y_k$, S: the standard deviation of the block; rescaled range R/S
- Example: $x = (0.5, -0.3, 0.8, -0.2, 0.6, -0.1, 0.4, -0.7)$, $\bar{x} = 0.125$
  - Y_k: 0.375, −0.050, 0.625, 0.300, 0.775, 0.550, 0.825, 0; $R = 0.825 - (-0.050) = 0.875$
  - $S = 0.489$, so $R/S = 1.79$
- **Hurst exponent**: average $(R/S)_n$ over blocks of size n; $E(R/S)_n \approx c\, n^H$, so H is the slope of $\log(R/S)_n$ on $\log n$
  - pitfalls: biased upwards in small blocks Mandelbrot and Wallis (1969); short memory also raises it; Lo (1991) corrects S with autocovariances (modified R/S)
  - **DFA** Peng et al. (1994): fit a line to the profile Y_k in each block; $F(n)$, the RMS (root mean square) deviation, grows like n^H

R/S and DFA on three series



⊡ US monthly inflation (FRED CPIAUCSL, 1947–2026), daily S&P 500 returns r_t and absolute returns $|r_t|$ (2000–2026); both axes logarithmic, curves normalised to 1 at $n = 10$

TSA_ch8_hurst_rs_dfa

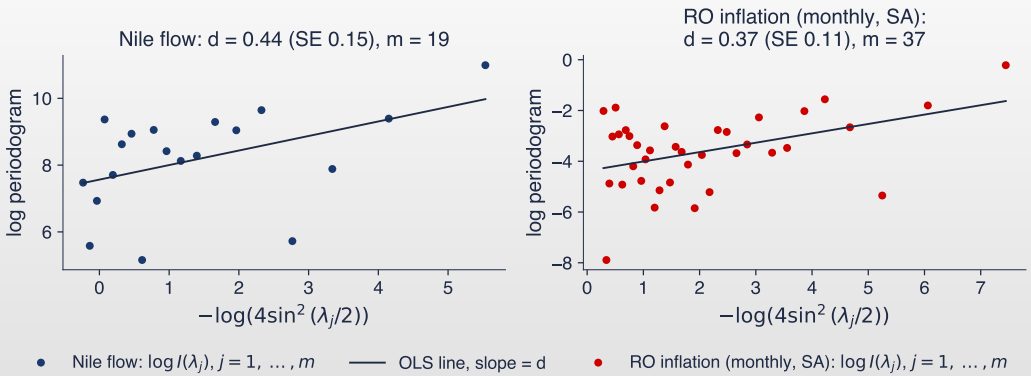
Interpreting the R/S and DFA slopes

- S&P 500 returns: slopes 0.53 (R/S) and 0.49 (DFA), close to 0.5: no memory in returns
  - the weak efficiency of Chapter 3 holds for the level of returns
- Absolute returns: 0.86 and 0.95; US inflation: 0.83 and 0.87
  - both well above 0.5: persistent; a DFA slope near 1 hints at d close to 0.5, the edge of stationarity
- These are graphical tools: they give no standard error and react to short memory and trends; the next estimators are built for inference

The periodogram and the GPH estimator

- **Periodogram** at the Fourier frequencies $\lambda_j = 2\pi j/T$: $I(\lambda_j) = \frac{1}{2\pi T}\left|\sum_t (x_t - \bar{x})e^{-i\lambda_j t}\right|^2$
  - i: the imaginary unit; $|\cdot|$: the modulus of a complex number; $j = 1, \ldots, \lfloor T/2 \rfloor$
  - a noisy estimate of the spectral density $f(\lambda_j)$; λ_j near 0 = slow cycles of period T/j
- Near zero, $f(\lambda) \approx G\,(4\sin^2(\lambda/2))^{-d}$; taking logs:
  - **GPH regression** Geweke and Porter-Hudak (1983): $\log I(\lambda_j) = c + d\left[-\log(4\sin^2(\lambda_j/2))\right] + u_j$, $j = 1, \ldots, m$
  - c: constant; u_j: regression error; $\hat{d}$ = OLS slope; standard error $\pi/\sqrt{24m}$
- **Bandwidth** m: how many low frequencies we use; here $m = \lfloor T^{0.65} \rfloor$
  - small m: little bias from the short-run part, large variance; large m: small variance, but short memory leaks into $\hat{d}$

GPH regressions: the Nile and Romanian inflation



⊡ Dots: log periodogram at the m lowest Fourier frequencies; line: OLS fit. Nile: $T = 100$ years; Romanian monthly inflation (SA): $T = 260$ months, August 2026 the last

TSA_ch8_estimation

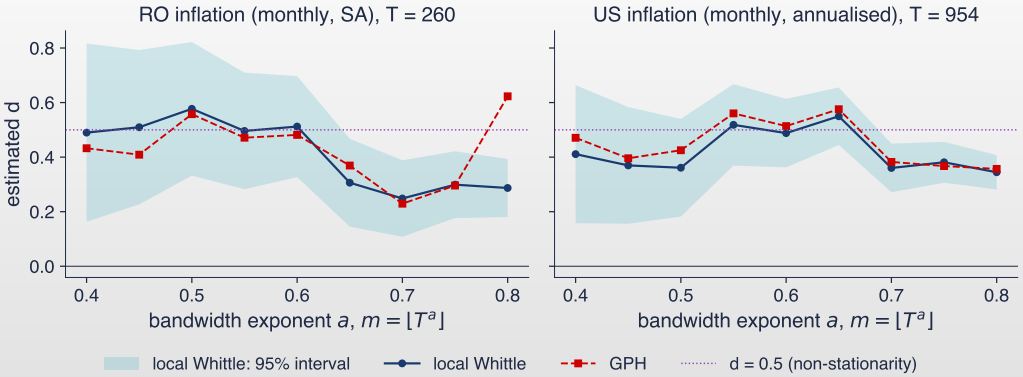
Interpreting the GPH regressions

- Nile: $\hat{d} = 0.44$ (SE 0.15), $m = 19$ frequencies, i.e. cycles longer than 5.3 years
  - local Whittle (next slide): 0.40 (SE 0.11); both significantly above 0, but the 95% intervals [0.15, 0.72] and [0.18, 0.63] also contain 0.5: stationarity is not established
- Romanian inflation: $\hat{d} = 0.37$ (SE 0.11); local Whittle 0.31 (SE 0.08)
  - cycles longer than 7.0 months: a shock to monthly inflation fades over years, not months
- The cloud is wide: each periodogram ordinate is roughly an exponential variable around $f(\lambda_j)$; the slope needs many points

The local Whittle estimator

- Local Whittle (LW) Robinson (1995)
  - assume $f(\lambda) \approx G\lambda^{-2d}$ for $\lambda_1, \ldots, \lambda_m$ and maximise the Gaussian (Whittle) likelihood of these ordinates
  - after concentrating out the constant G of $f(\lambda) \approx G\lambda^{-2d}$:
    $\hat{d} = \arg\min_d \ \log\left(\frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2d} I(\lambda_j)\right) - \frac{2d}{m}\sum_{j=1}^{m} \log \lambda_j$
- Standard error $1/(2\sqrt{m})$, smaller than GPH ($\pi/\sqrt{24m} \approx 0.64/\sqrt{m}$)
  - consistent for $d < 1$, but asymptotically normal with this standard error only for $-1/2 < d < 3/4$ Velasco (1999)
- Both GPH and local Whittle depend on m: always report $\hat{d}$ for several bandwidths

Sensitivity to the bandwidth



⊡ $\hat{d}$ for $m = \lfloor T^a \rfloor$, $a = 0.40, \ldots, 0.80$; shaded: 95% interval of local Whittle. Romanian monthly inflation ($T = 260$, m from 9 to 85) and US monthly inflation ($T = 954$)

TSA_ch8_estimation

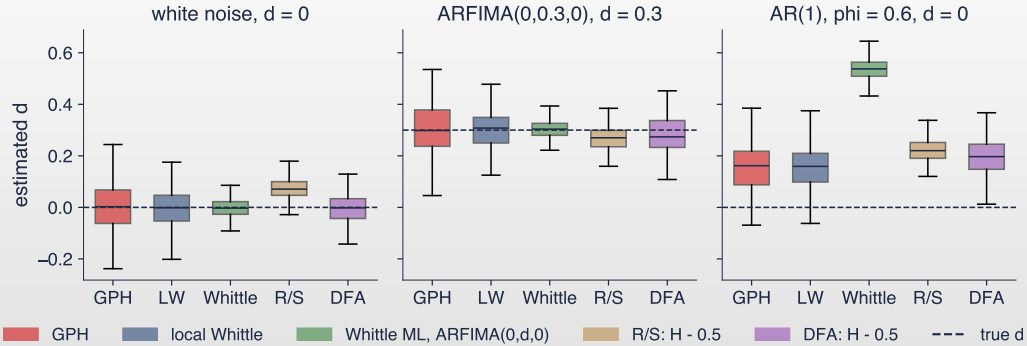
Interpreting the bandwidth plot

- Romania: local Whittle between 0.25 and 0.58, GPH between 0.23 and 0.62
  - with few frequencies the estimate is near 0.5; with more it falls to about 0.3: the answer depends on the choice of m
- United States: local Whittle between 0.34 and 0.55, always well above 0
  - the long sample (1947–2026) gives narrow intervals; the level of d near 0.5 hints at regime changes (Section 7)
- Conclusion: inflation has long memory in both countries; a single number for d hides an uncertainty of about ± 0.15

Maximum likelihood for ARFIMA(p, d, q)

- **Exact ML** Sowell (1992): for Gaussian $x = (x_1, \ldots, x_T)'$ with covariance matrix $\Gamma(\vartheta)$, $\vartheta = (d, \phi, \theta, \sigma^2)$:
  - $\ell(\vartheta) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log\det\Gamma(\vartheta) - \frac{1}{2}(x-\mu)'\Gamma(\vartheta)^{-1}(x-\mu)$
  - ℓ: the log-likelihood; μ: the vector of means; $\Gamma(\vartheta)$: the $T \times T$ matrix of the ARFIMA autocovariances
  - Γ comes from the ARFIMA autocovariances; the Durbin–Levinson recursion evaluates ℓ in $O(T^2)$ operations
- **Whittle** (approximate) ML Fox and Taqqu (1986): replace the likelihood by $-\frac{1}{2}\sum_j \left[\log f(\lambda_j;\vartheta) + I(\lambda_j)/f(\lambda_j;\vartheta)\right]$ over all Fourier frequencies
  - $f(\lambda_j;\vartheta)$: the spectral density implied by the parameters ϑ
  - fast ($O(T\log T)$ with the FFT, the fast Fourier transform), close to exact ML in large samples
- Parametric estimators are efficient (SE about $\sqrt{6}/(\pi\sqrt{T})$ for ARFIMA$(0, d, 0)$) only if the ARMA orders are right; choose p, q by AIC/BIC

Monte Carlo: five estimators of d



- 300 samples of $T = 500$ per design; GPH and local Whittle with $m = \lfloor T^{0.65} \rfloor$; Whittle ML of ARFIMA$(0, d, 0)$; R/S and DFA converted with $d = H - 1/2$

TSA_ch8_monte_carlo

Interpreting the Monte Carlo

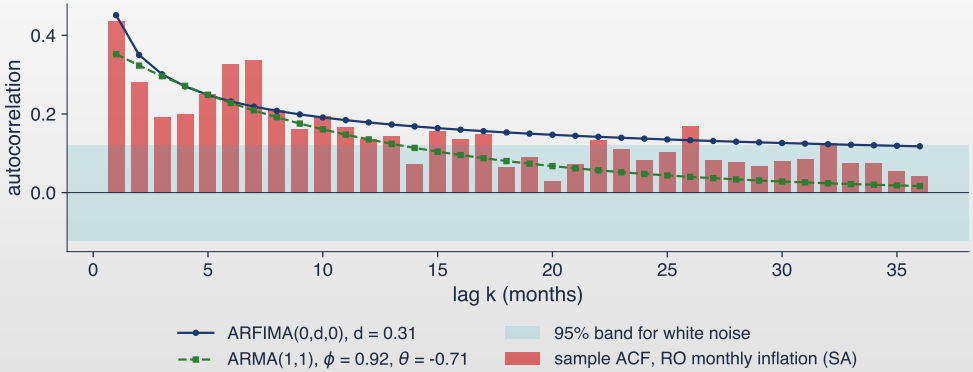
- Correct model: Whittle ML is the most precise (SD 0.04 against 0.08 for local Whittle and 0.10 for GPH at $d = 0.3$)
  - R/S is biased upwards under white noise (mean 0.07) Mandelbrot and Wallis (1969)
- AR(1) with $\phi = 0.6$ and no long memory: ARFIMA$(0, d, 0)$ by Whittle gives 0.54, GPH 0.15, local Whittle 0.15
  - a misspecified parametric model turns short memory into "long memory"; semiparametric estimators suffer less, but are not immune
- Practice: always fit ARFIMA(p, d, q) with $p, q \geq 1$ as alternatives, and check the semiparametric $\hat{d}$ for several m

Romanian inflation: ARMA or ARFIMA?

Model (exact ML)	$\hat{d}$	AR / MA	ℓ	AIC	BIC
ARMA(1,1)	0	0.92 / −0.71	−138.0	282.1	292.8
ARMA(2,0)	0	0.39, ...	−139.9	285.9	296.6
ARFIMA(0,d,0)	0.31	–	−135.9	**275.8**	**282.9**
ARFIMA(1,d,0)	0.27	0.07 / –	−135.6	277.1	287.8
ARFIMA(0,d,1)	0.27	– / 0.08	−135.5	277.1	287.8
ARFIMA(1,d,1)	0.27	−0.03 / 0.10	−135.5	279.1	293.3

- Romanian monthly HICP inflation, seasonally adjusted, $T = 260$ months (2005–2026); ℓ: maximised log-likelihood; bold: the smallest criterion
- ARFIMA$(0, d, 0)$ wins with one parameter: $\hat{d} = 0.31$ (SE 0.047); adding ϕ or θ barely changes ℓ
- The best ARMA needs $\phi = 0.92$ close to 1 and a cancelling MA term to imitate the slow decay

Romanian inflation: fitted autocorrelations



- Bars: sample ACF; lines: the ACF implied by ARFIMA$(0, d, 0)$ and by ARMA(1,1), both estimated by exact ML

TSA_ch8_estimation

Interpreting the fitted ACF

- At lag 12: sample 0.14, ARFIMA 0.18, ARMA(1,1) 0.14; at lag 24: 0.08, 0.14 and 0.05
  - the two models agree on the first year and disagree on the second: their long-horizon forecasts will differ
- The ARFIMA curve matches the slow tail with one parameter; the ARMA curve goes to zero exponentially
- Seasonal adjustment matters: a seasonal peak near the lowest frequencies would distort $\hat{d}$ (Chapter 4)

Long memory in twelve series

Series	T	$H_{R/S}$	H_{DFA}	d_{GPH}	d_{LW} (SE)
Nile	100	0.83	0.64	0.44	0.40 (0.11)
RO inflation, monthly SA	260	0.75	0.84	0.37	0.31 (0.08)
RO inflation, 12-month	260	0.96	1.38	1.14	1.00 (0.08)
US inflation, monthly	954	0.83	0.87	0.58	0.55 (0.05)
US inflation, 1985–2019	420	0.69	0.59	0.07	0.10 (0.07)
US unemployment	944	0.97	1.27	0.82	0.91 (0.05)
S&P 500, r_t	6 714	0.53	0.49	0.02	−0.01 (0.03)
S&P 500, $\lvert r_t \rvert$	6 714	0.86	0.95	0.56	0.53 (0.03)
BET, r_t	6 670	0.58	0.57	0.16	0.12 (0.03)
BET, $\lvert r_t \rvert$	6 670	0.79	0.85	0.40	0.39 (0.03)
EUR/RON, r_t	5 352	0.53	0.52	−0.01	−0.01 (0.03)
EUR/RON, $\lvert r_t \rvert$	5 352	0.82	0.85	0.38	0.39 (0.03)

- $m = \lfloor T^{\mathbf{0.65}} \rfloor$; LW: local Whittle, SE $1/(2\sqrt{m})$ (valid for $d < 3/4$); data: statsmodels, Eurostat, FRED, EODHD, BNR; returns daily since 2000 (EUR/RON since 2005)

Interpreting the table

- Returns: $d \approx 0$ for the S&P 500 (-0.01) and EUR/RON (-0.01); small but positive for the BET (0.12)
  - absolute returns: 0.53, 0.39 and 0.39: long memory in volatility (Section 6)
- Romanian 12-month inflation: $d \approx 1.00$, against 0.31 for the monthly rate: the 12-month rate sums 12 overlapping monthly changes
  - the filter $1 + L + \cdots + L^{11}$ multiplies the spectrum by $\sin^2(6\lambda)/\sin^2(\lambda/2)$: 144 at $\lambda = 0$, 0 at $\lambda = 2\pi/12 \approx 0.52$, inside the band used ($\lambda_m = 0.89$); this steep fall looks like a pole at 0 and pushes $\hat{d}$ towards 1; model the monthly rate
- US unemployment: $d \approx 0.91$, non-stationary; mean reversion ($d < 1$) is not established: [0.80, 1.01] contains 1 (and above $d = 3/4$ the SE is not valid); US inflation 1985–2019: 0.10, against 0.55 on 1947–2026

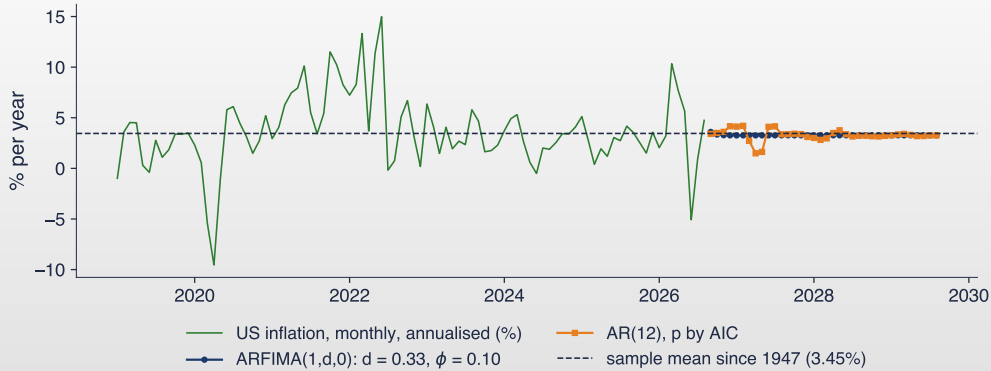
Recap: Estimating d

- R/S and DFA: slopes of log–log plots, graphical, biased in small samples
- GPH and local Whittle: regressions or likelihoods on the m lowest frequencies; report several m
- Exact ML and Whittle: efficient if the ARMA part is right; short memory can masquerade as d
- Romanian monthly inflation: ARFIMA$(0, d, 0)$ with $\hat{d} = 0.31$ beats ARMA by AIC and BIC

Forecasts from the AR(∞) form

- An invertible ARFIMA can be written as $\pi(L)(x_t - \mu) = \varepsilon_t$, with $\pi(L) = \phi(L)(1-L)^d/\theta(L) = \sum_k \pi_k L^k$
  - π_k: the AR(∞) weights (for ARFIMA$(0, d, 0)$, the weights of $(1-L)^d$); $\hat{x}_{T+h}$: the forecast made at T for $T+h$
  - one step: $\hat{x}_{T+1} = \mu - \sum_{k=1}^{T} \pi_k (x_{T+1-k} - \mu)$; h steps: the same recursion, with forecasts in place of unknown values
- Example, ARFIMA$(0, 0.4, 0)$: $-\pi_k = 0.4,\ 0.12,\ 0.064,\ 0.042, \ldots$
  - $\hat{x}_{T+1} - \mu = 0.4(x_T - \mu) + 0.12(x_{T-1} - \mu) + 0.064(x_{T-2} - \mu) + \ldots$: the whole history counts
  - an AR(1) uses only x_T; after a long period above the mean the ARFIMA forecast stays higher, for longer
- Forecasts revert to μ hyperbolically; the forecast error variance $\sigma^2 \sum_{j<h} \psi_j^2$ (ψ_j: the MA(∞) weights) grows slowly towards $\gamma(0)$ Ray (1993)

US inflation: ARFIMA and AR forecasts



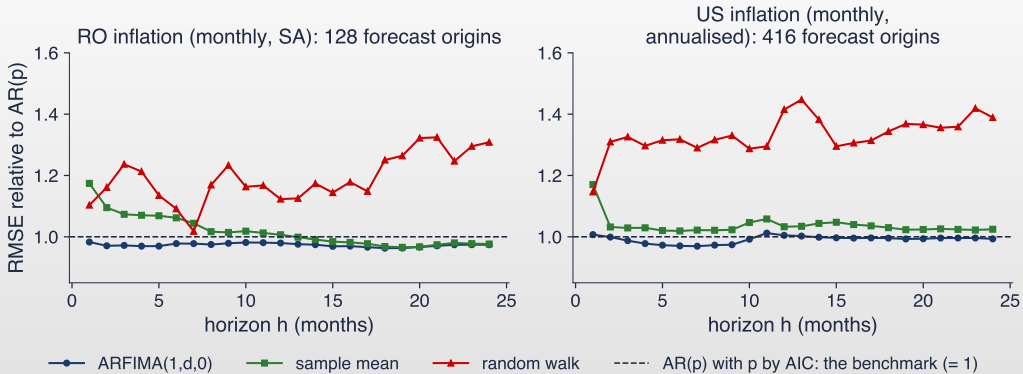
- US monthly CPI inflation, annualised, $T = 954$ values up to August 2026; ARFIMA(1,d,0) estimated by Whittle; AR(p) with p by AIC; 36 months ahead. FRED has no October 2025 CPI (not published): the November 2025 value is the change since September, annualised over its two months

TSA_ch8_forecasting

Interpreting the two forecasts

- ARFIMA: $\hat{d} = 0.33$, $\hat{\phi} = 0.10$; AR: $p = 12$ lags, the largest allowed
  - the AR needs a year of lags to imitate what d does with one parameter
- Last 12 months (12-month inflation) 3.30%; the forecasts: 3.60% and 3.40% next month, 3.26% and 3.37% in a year, 3.27% and 3.25% in three years
  - both approach the long-run mean 3.45%; the paths differ little because the AR(12) already carries a long memory
- Whether the long-run mean of 1947–2026 is the right anchor today is a question about regimes, not about d

Does long memory improve forecasts?



- Pseudo out-of-sample, expanding window: Romania 128 origins (January 2014–August 2024), seasonal adjustment inside each window; United States 416 origins (January 1990–August 2024), errors matched by calendar month (no October 2025 CPI). RMSE relative to AR(p); below 1: better

TSA_ch8_forecasting

Interpreting the forecast comparison

- Romania: ARFIMA beats the AR at every horizon, by 1.7–3.7% (ratio 0.983 at 1 month, 0.980 at 12, 0.974 at 24)
  - RMSE at 12 months: ARFIMA 0.53, AR 0.54, mean 0.54, random walk 0.61 (pp per month)
- United States: ratios between 0.969 and 1.012: practically a tie with a rich AR(p)
  - the random walk is far worse at all horizons beyond one month
- Long memory gives modest, steady gains; the large errors come from regime changes (2008, 2021–2022) that no linear model foresees; test the gains with Diebold–Mariano (Chapter 4)

Recap: Forecasting with ARFIMA

- Forecasts from the AR(∞) weights π_k: the whole history counts
- Hyperbolic return to the mean; slowly growing forecast uncertainty
- Inflation: small gains over AR in Romania, a tie in the United States

Case study: Ding, Granger and Engle (1993)

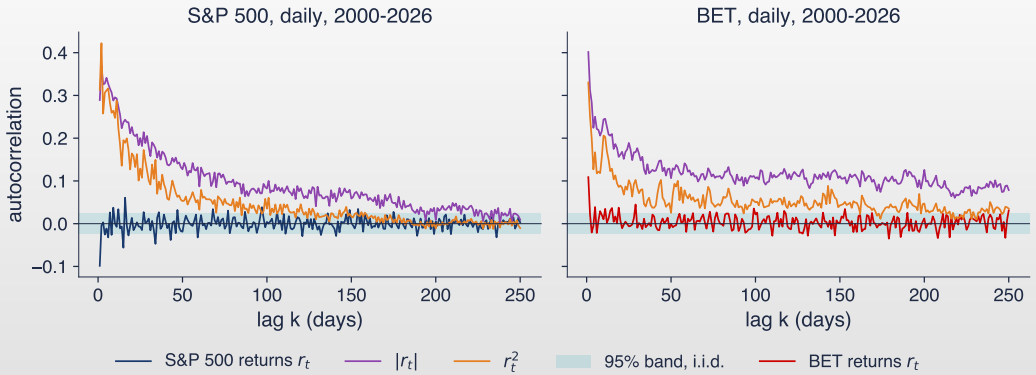


Robert F. Engle, Nobel Prize 2003 with Clive Granger

Photo: Jon Demske (2022); CC BY-SA 4.0; Wikimedia Commons

- Ding, Granger and Engle (1993): daily S&P 500 returns, 1928–1991
  - returns are almost uncorrelated, but $|r_t|^\delta$ is autocorrelated for thousands of lags
  - $\delta > 0$: a power; the autocorrelation is strongest for $\delta \approx 1$
  - the "long memory property" of stock market returns: memory lives in the size, not in the sign
- A GARCH(1,1) (Chapter 5) implies autocorrelations of r_t^2 that fall like $(\alpha + \beta)^k$ (α, β: the ARCH and GARCH parameters): exponentially
- This led to FIGARCH Baillie, Bollerslev and Mikkelsen (1996) and, with high-frequency data, to realised volatility models Andersen et al. (2003), Corsi (2009)

Returns, absolute and squared returns



- Daily log returns of the S&P 500 ($T = 6\,714$) and of the BET ($T = 6\,670$), 2000–2026; ACF up to lag 250 (one trading year); shaded: $\pm 1.96/\sqrt{T}$

TSA_ch8_volatility_memory

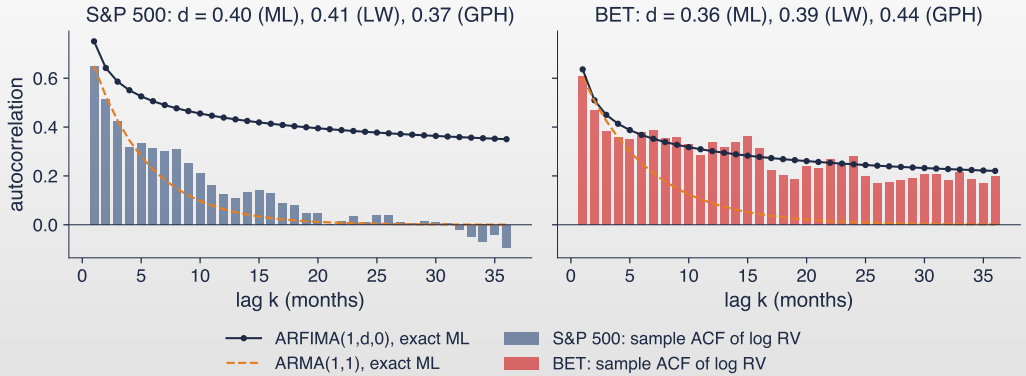
Interpreting the volatility ACF

- S&P 500: $|r_t|$ has $\hat{\rho} = 0.29$ at lag 1 and still 0.08 at lag 100; 227 of 250 lags above the band
  - BET: 0.40 at lag 1, 0.08 at lag 250, 250 of 250 above the band
- $|r_t|$ is more persistent than r_t^2, as in Ding, Granger and Engle (1993); squares are dominated by a few crash days
  - local Whittle for $|r_t|$: 0.53 (S&P 500), 0.39 (BET); after a random shuffle of the days: 0.00 and -0.02
- The shuffle keeps the distribution and destroys the order: the memory is in the timing of calm and turbulent days, not in fat tails

Realised volatility

- **Realised variance** of month m Andersen et al. (2003)
  - $RV_m = \sum_{t \in m} r_t^2$: the sum of the squared daily returns of the month
  - realised volatility $\sqrt{RV_m}$ measures the volatility of month m almost without a model; with intraday data the same idea gives a daily RV
- We model $\log \sqrt{RV_m}$: the logarithm makes the distribution close to the Normal distribution and removes the positivity constraint
  - the stylised fact of Andersen et al. (2003): log RV is well described by a long-memory model with $d \approx 0.4$
- Unlike $|r_t|$, $\log \sqrt{RV_m}$ is a precise measure: the noise of single days averages out within the month

Monthly realised volatility: ARFIMA or ARMA?



⊡ Log monthly realised volatility from daily returns, 321 months (2000–2026); lines: ACF of ARFIMA(1,d,0) and ARMA(1,1), both by exact ML

TSA_ch8_volatility_memory

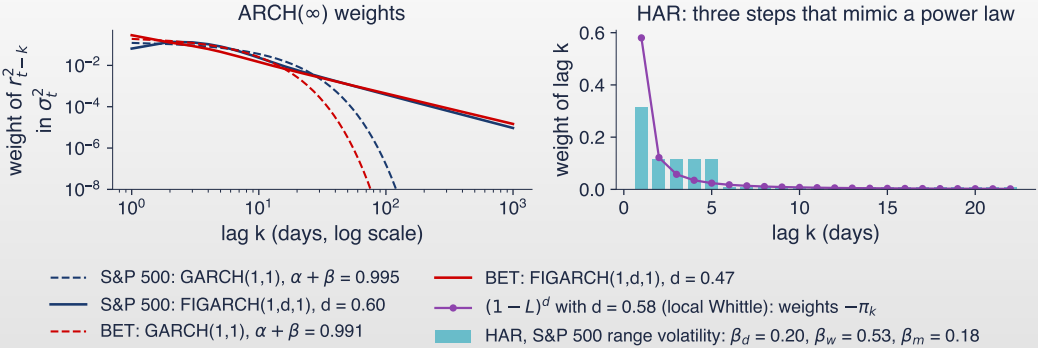
Interpreting realised volatility

- All estimators agree: d between 0.36 and 0.44 for both markets (ML 0.40 and 0.36)
  - BET: ARFIMA fits the slow tail (lag 24: sample 0.28) and wins on BIC (343.9 against 355.4 for ARMA(1,1))
- S&P 500: the sample ACF dies out after about 20 months; ARMA(1,1) with $\phi = 0.81$ fits as well (BIC 294.4 against 295.9)
  - with 27 years of data, long memory and a persistent ARMA are hard to tell apart
- The sample ACF of a long-memory series is biased downwards (the sample mean absorbs part of the slow component): judge models by likelihood, not by eye

FIGARCH and HAR

- **FIGARCH**$(1, d, 1)$ Baillie, Bollerslev and Mikkelsen (1996): write GARCH(1,1) as an ARMA(1,1) in ε_t^2 and apply $(1-L)^d$:
  - σ_t^2: the conditional variance; ε_t: the return shock; $\omega > 0$, β, ϕ: GARCH-type parameters (Chapter 5)
  - $\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \left[1 - \beta L - (1-\phi L)(1-L)^d\right]\varepsilon_t^2$, that is $\sigma_t^2 = \omega^* + \sum_{k\geq 1} \lambda_k \varepsilon_{t-k}^2$
  - $\omega^* = \omega/(1-\beta)$; the ARCH($\infty$) weights λ_k decay like k^{-1-d}; $d = 0$: GARCH, $d = 1$: IGARCH
- **HAR** (heterogeneous autoregressive model) Corsi (2009):
  $RV_{t+1} = c + \beta_d RV_t + \beta_w \overline{RV}_{t-4:t} + \beta_m \overline{RV}_{t-21:t} + u_{t+1}$
  - $\overline{RV}_{t-4:t}$, $\overline{RV}_{t-21:t}$: averages over the last 5 and 22 days (weekly, monthly); β_d, β_w, β_m: their weights; estimated by OLS
  - the idea: market participants with daily, weekly and monthly horizons
  - not a long-memory model, but three steps that imitate a power law over the horizons that matter
- Both answer the failure of GARCH from Chapter 5: volatility shocks die out too fast in GARCH

FIGARCH against GARCH; HAR against $(1-L)^d$



- Left: ARCH(∞) weights of GARCH(1,1) and FIGARCH(1,d,1), Student t errors, daily returns 2000–2026 (`arch`)
- Right: HAR for the daily range-based volatility of the S&P 500 (Parkinson estimator from high and low prices)

TSA_ch8_volatility_memory

Interpreting FIGARCH and HAR

- FIGARCH: $\hat{d} = 0.60$ (S&P 500) and 0.47 (BET); BIC falls by 34 and 71 against GARCH
  - weight of a shock 100 days ago in today's variance: GARCH $1.6 \cdot 10^{-7}$, FIGARCH $3.9 \cdot 10^{-4}$ (S&P 500): three orders of magnitude more
- GARCH reaches persistence by $\alpha + \beta = 0.995$, near 1: the near-IGARCH result of Chapter 5 is a symptom of long memory
  - HAR: $\hat{\beta}_d = 0.20$, $\hat{\beta}_w = 0.53$, $\hat{\beta}_m = 0.18$; the weekly component dominates
- The HAR steps (0.31 at lag 1, 0.11 at lags 2–5, 0.008 at lags 6–22) decay roughly like the weights $-\pi_k$ of $(1 - L)^d$, $d = 0.58$ from local Whittle (0.58 at lag 1, 0.12 at lag 2, 0.018 at lag 6, 0.002 at lag 22): less weight at lag 1, a step function

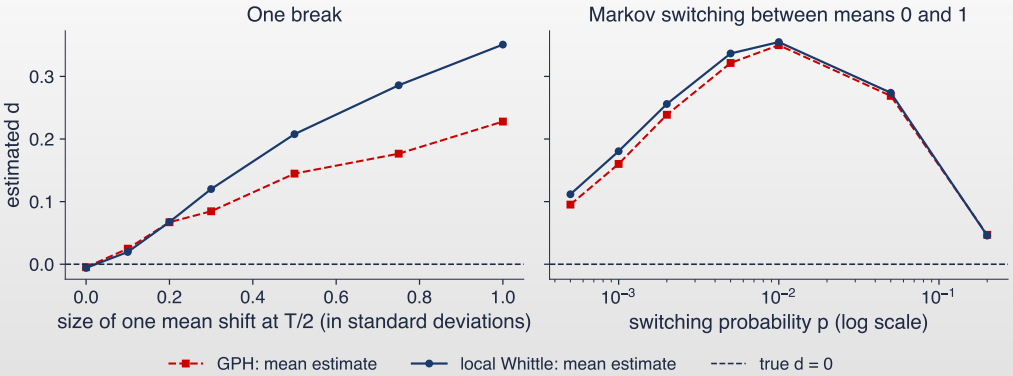
Recap: Long memory in volatility

- Returns: $d \approx 0$; absolute returns and realised volatility: $d \approx 0.4$
- The shuffle test: memory in the order of the days, not in the distribution
- FIGARCH: hyperbolic ARCH(∞) weights; HAR: a simple OLS approximation

Structural breaks, regime switching and spurious long memory

- A level shift adds a slow, step-like component: low frequencies gain power and the sample ACF decays slowly
  - the same mechanism that makes a broken series look like a unit root (Perron, Chapter 3)
- Diebold and Inoue (2001): a Markov-switching mean with rare switches is "observationally equivalent" to long memory
  - the fewer the switches in a sample of length T, the closer the series looks to $I(d)$
  - regime-switching models are the topic of Chapter 10
- Granger and Hyung (2004): occasional breaks explain much of the long memory of S&P 500 absolute returns

Apparent long memory in short-memory processes



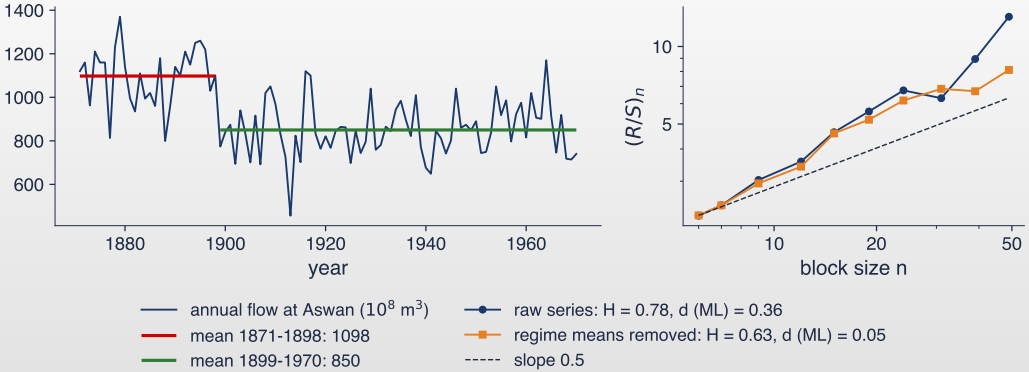
- 200 samples of $T = 1000$, true $d = 0$. Left: $N(0,1)$ noise with one mean shift at $T/2$; right: $N(0,1)$ noise plus a mean that switches between 0 and 1 with probability p each period

TSA_ch8_spurious_memory

Interpreting the spurious memory

- One shift of half a standard deviation already gives $\hat{d} = 0.21$ (local Whittle); a shift of one standard deviation gives 0.35
  - the break is hard to see by eye in noisy data, yet the estimator reports moderate long memory
- Regime switching: with $p = 0.01$ (about 10 switches) $\hat{d} = 0.35$; with $p = 0.2$ (frequent switches) $\hat{d} = 0.05$
  - rare regimes look like memory, frequent ones average out into short memory
- A significant $\hat{d}$ is consistent with long memory and with breaks: other evidence is needed to decide

The Nile in 1898: memory or a break?



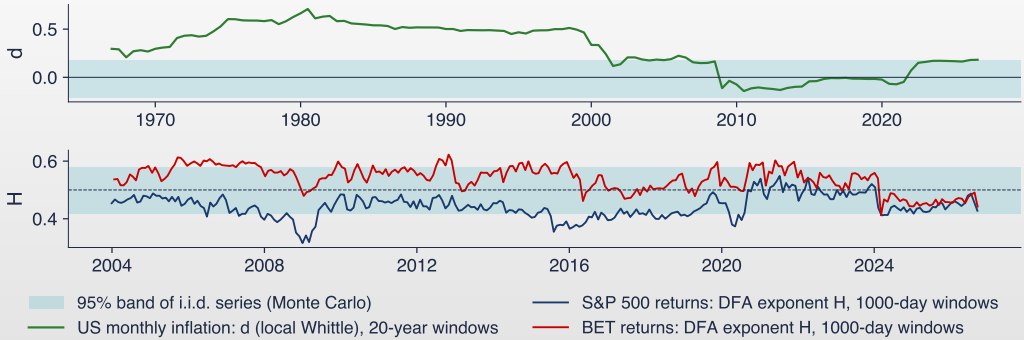
⊡ Left: the flow with the means before and after 1898, the change point of Cobb (1978) (the first Aswan dam was completed in 1902); right: R/S plot of the raw series and of the series with the two regime means removed

TSA_ch8_spurious_memory

Interpreting the Nile break

- The mean falls from 1098 to 850 (-23%) after 1898
  - raw series: $\hat{d} = 0.36$ (exact ML, SE 0.07), local Whittle 0.40, R/S $H = 0.78$
- Regime means removed: $\hat{d} = 0.05$ (SE 0.10), local Whittle -0.17; $\hat{\rho}(1)$ falls from 0.50 to 0.16
  - for these 100 years one break explains almost all of the memory
- Hurst's own evidence came from much longer records (the Roda gauge, from the 7th century); the lesson is to test for breaks before reading d

The stability of memory over time



- Top: local Whittle $\hat{d}$ of US monthly inflation in rolling 20-year windows (dated at the window end)
- Bottom: DFA exponent of daily S&P 500 and BET returns in 1000-day windows moved by 21 days; shaded: 95% Monte Carlo bands for i.i.d. series Weron (2002)

TSA_ch8_spurious_memory

Interpreting the rolling estimates

- US inflation: $\hat{d}$ peaks at 0.71 in the window ending July 1980 (the Great Inflation and the Volcker disinflation) and falls to -0.14 (July 2010)
  - in the calm years of the Great Moderation (the Fed announced an explicit 2% target only in 2012) d is inside the band of no memory $[-0.209, 0.182]$; the 2021–2022 surge lifts it to 0.1820, at the upper edge of the band (0.1823) but not above it
- Returns: the S&P 500 exponent stays mostly below 0.5 (minimum 0.32, December 2008); the BET exceeds the band in 24% of the windows
  - a less liquid market shows some persistence in returns, fading in recent years
- The full-sample US estimate (0.55) mixes regimes: memory that comes and goes with monetary regimes is a sign of breaks, not of a constant d

Checks against spurious long memory

- Plot the series and test for breaks (Chapter 3: Zivot–Andrews, Bai–Perron); re-estimate d on subsamples and after removing regime means
- Report $\hat{d}$ for several bandwidths m; a value that collapses when m changes is fragile
- Compare ARFIMA with ARMA and with regime-switching models by likelihood and out-of-sample forecasts (Chapter 10)
- Use the shuffle test and Monte Carlo bands built for the same sample length
- Avoid overlapping data (12-month inflation, rolling sums): they create artificial low-frequency power

Recap: Spurious long memory

- Breaks and rare regime switches produce $\hat{d} > 0$ in short-memory series
- The Nile: one break in 1898 explains most of the memory of 1871–1970
- US inflation: d changes with the monetary regime

Fractional cointegration

- Chapter 7: two $I(1)$ series are cointegrated if a combination $y_t - \beta x_t$ is $I(0)$
  - **fractional cointegration**: $x_t, y_t \sim I(d)$ and $y_t - \beta x_t \sim I(d-b)$ with $0 < b \leq d$
  - the equilibrium error may itself have long memory: deviations from equilibrium are corrected, but slowly
- Example: purchasing power parity Cheung and Lai (1993): the real exchange rate reverts to parity with d between 0 and 1
  - the Engle–Granger and Johansen tests of Chapter 7 assume $b = 1$ and may miss such slow adjustment
- Beyond this course; a natural project: estimate d of the residual of a cointegrating regression with local Whittle

Possible contribution of AI

- **Code**: a first draft of fractional differencing, of the GPH and local Whittle estimators, of an ARFIMA likelihood
- **Explanation**: a second explanation of the periodogram, of the bandwidth trade-off, of the link $H = d + 1/2$
- **Exploration**: d for many countries' inflation or many assets' volatility, with subsamples and bandwidths
- Example prompt
  - `Write Python code that downloads Romanian monthly HICP (Eurostat prc_hicp_minr), computes the monthly inflation rate, removes the monthly means, estimates d with GPH and local Whittle for bandwidths T^0.5 to T^0.8, fits ARFIMA(p,d,q) by exact Gaussian maximum likelihood and compares it with ARMA(1,1) by BIC.`

Checks you must run

- The library: `statsmodels` has no ARFIMA class; an AI answer that calls one has invented it
- The sign and the scale: GPH regresses on $-\log(4\sin^2(\lambda/2))$, so the slope is d, not $-2d$; R/S and DFA give H, not d
- The range: a stationary ARFIMA likelihood needs $d < 0.5$; difference first if $\hat{d}$ is near or above 0.5
- Simulate a series with known d and check that the code recovers it
- Breaks, seasonality and overlapping data before any claim of long memory
- Every cited reference: it must exist; check the DOI

Key takeaways

- Long memory: hyperbolic ACF decay $\rho(k) \sim Ck^{2d-1}$ and a spectral pole at zero; ARMA memory is exponential
- $(1-L)^d$ with real d; ARFIMA(p, d, q) separates long-run memory (d) from short-run dynamics (ARMA)
- Stationary for $d < 1/2$, invertible for $d > -1/2$, mean-reverting for $d < 1$; $H = d + 1/2$
- Estimate d with several methods and bandwidths; prefer likelihood-based comparisons with ARMA
- Inflation and volatility show long memory; returns do not; breaks and regimes can fake it

Key formulas

Quantity	Formula
ARFIMA(p,d,q)	$\phi(L)(1-L)^d(x_t-\mu)=\theta(L)\varepsilon_t$
Weights of $(1-L)^d$	$\pi_0=1, \quad \pi_k=\pi_{k-1}(k-1-d)/k$
ACF of ARFIMA$(0,d,0)$	$\rho(1)=d/(1-d), \quad \rho(k)=\rho(k-1)(k-1+d)/(k-d)\sim Ck^{2d-1}$
Spectrum near 0	$f(\lambda)\sim G\lambda^{-2d}$
GPH	$\log I(\lambda_j)=c+d\,[-\log(4\sin^2(\lambda_j/2))]+u_j, \quad \mathrm{SE}=\pi/\sqrt{24m}$
Local Whittle	$\min_d \log\big(\frac{1}{m}\sum_j \lambda_j^{2d} I(\lambda_j)\big)-\frac{2d}{m}\sum_j \log\lambda_j, \quad \mathrm{SE}=1/(2\sqrt{m})$
R/S, DFA	$E(R/S)_n\approx cn^H, \quad F(n)\approx cn^H, \quad H=d+1/2$
FIGARCH	$\sigma_t^2=\omega+\beta\sigma_{t-1}^2+[1-\beta L-(1-\phi L)(1-L)^d]\varepsilon_t^2$

Self-assessment

- **Question**: what are π_1 and π_2 of $(1-L)^{0.2}$?
  - **Answer**: $\pi_1 = -0.2$, $\pi_2 = -0.2 \cdot 0.8/2 = -0.08$
- **Question**: is ARFIMA$(0, 0.7, 0)$ stationary?
  - **Answer**: no ($d \geq 0.5$), but it is mean-reverting; its first difference is ARFIMA$(0, -0.3, 0)$, stationary and anti-persistent
- **Question**: GPH gives $\hat{d} = 0.35$ on a series with a large level shift. Is this evidence of long memory?
  - **Answer**: not by itself: a break produces the same low-frequency power; re-estimate after removing the shift and on subsamples
- Next: Chapter 9, machine learning for time series

References (1/3)

- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), 579–625. doi:10.1111/1468-0262.00418
- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1), 5–59. doi:10.1016/0304-4076(95)01732-1
- Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1), 3–30. doi:10.1016/S0304-4076(95)01749-6
- Beran, J., Feng, Y., Ghosh, S., & Kulik, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer. doi:10.1007/978-3-642-35512-7
- Brockwell, P. J., & Davis, R. A. (1991). *Time Series: Theory and Methods* (2nd ed.). Springer. doi:10.1007/978-1-4419-0320-4
- Cheung, Y.-W., & Lai, K. S. (1993). A fractional cointegration analysis of purchasing power parity. *Journal of Business & Economic Statistics*, 11(1), 103–112. doi:10.1080/07350015.1993.10509936
- Cobb, G. W. (1978). The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika*, 65(2), 243–251. doi:10.1093/biomet/65.2.243
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174–196. doi:10.1093/jjfinec/nbp001
- Diebold, F. X., & Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1), 131–159. doi:10.1016/S0304-4076(01)00073-2
- Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1), 83–106. doi:10.1016/0927-5398(93)90006-D

References (2/3)

- Fox, R., & Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, 14(2), 517–532. doi:10.1214/aos/1176349936
- Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4), 221–238. doi:10.1111/j.1467-9892.1983.tb00371.x
- Granger, C. W. J., & Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11(3), 399–421. doi:10.1016/j.jempfin.2003.03.001
- Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1), 15–29. doi:10.1111/j.1467-9892.1980.tb00297.x
- Graves, T., Gramacy, R., Watkins, N., & Franzke, C. (2017). A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951–1980. *Entropy*, 19(9), 437. doi:10.3390/e19090437
- Hassler, U., & Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics*, 13(1), 37–45. doi:10.1080/07350015.1995.10524577
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1), 165–176. doi:10.1093/biomet/68.1.165
- Huang, C., & Petukhina, A. (2022). *Applied Time Series Analysis and Forecasting with Python*. Springer. doi:10.1007/978-3-031-13584-2
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116(1), 770–799. doi:10.1061/TACEAT.0006518

References (3/3)

- Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica*, 59(5), 1279–1313. doi:10.2307/2938368
- Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4), 422–437. doi:10.1137/1010093
- Mandelbrot, B. B., & Wallis, J. R. (1968). Noah, Joseph, and operational hydrology. *Water Resources Research*, 4(5), 909–918. doi:10.1029/WR004i005p00909
- Mandelbrot, B. B., & Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, 5(5), 967–988. doi:10.1029/WR005i005p00967
- Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49(2), 1685–1689. doi:10.1103/PhysRevE.49.1685
- Ray, B. K. (1993). Modeling long-memory processes for optimal long-range prediction. *Journal of Time Series Analysis*, 14(5), 511–525. doi:10.1111/j.1467-9892.1993.tb00161.x
- Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5), 1630–1661. doi:10.1214/aos/1176324317
- Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53(1–3), 165–188. doi:10.1016/0304-4076(92)90084-5
- Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20(1), 87–127. doi:10.1111/1467-9892.00127
- Weron, R. (2002). Estimating long-range dependence: Finite sample properties and confidence intervals. *Physica A*, 312(1–2), 285–299. doi:10.1016/S0378-4371(02)00961-5