








Time Series Analysis

Chapter 1: Stochastic processes and stationarity



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

- **ACF**: Autocorrelation Function
- **AI**: Artificial Intelligence
- **AR**: AutoRegressive (model)
- **ARIMA**: AutoRegressive Integrated Moving Average (model)
- **ARMA**: AutoRegressive Moving Average (model)
- **BET**: Bucharest Exchange Trading (index)
- **BNR**: Banca Națională a României (the National Bank of Romania)
- **BVB**: Bursa de Valori București (the Bucharest Stock Exchange)
- **EU**: European Union
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **GDP**: Gross Domestic Product
- **HICP**: Harmonised Index of Consumer Prices
- **MA**: Moving Average
- **PACF**: Partial AutoCorrelation Function
- **SARIMA**: Seasonal ARIMA (model)
- **VAR**: Vector AutoRegression

Today's question and route

- **Question**: we observe one path of a series, a single history; what can we learn from it about the mechanism that produced it?
  - the answer rests on two assumptions: **stationarity** and **ergodicity**
  - the main tool: the **autocorrelation function** (ACF)
- **Route** of the chapter
  - stochastic processes and their moments; strict and weak stationarity
  - white noise and the random walk; the lag operator, differencing, the Wold decomposition, ergodicity
  - sample ACF and PACF with confidence bands; the Box–Pierce and Ljung–Box tests
  - transformations (log, differencing, Box–Cox) and a first look at real series
- Seminar 1 comes before this lecture: its primer gives the definitions; here we derive and explain them

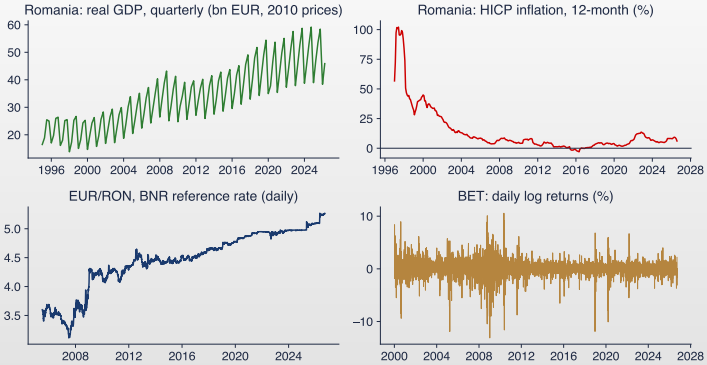
Learning outcomes

- Define a stochastic process and its mean, autocovariance and autocorrelation functions
- Distinguish strict from weak stationarity, and check weak stationarity for simple processes
- Derive the moments of white noise, of a moving average and of a random walk
- Compute and read the sample ACF and PACF with their $\pm 1.96/\sqrt{T}$ bands
- Test for autocorrelation with the Box–Pierce and Ljung–Box statistics
- Choose a transformation (log, difference, Box–Cox) that makes a real series closer to stationary

Reading and tools

- Textbook: Huang and Petukhina (2022), Ch. 1–2 (concepts, exploratory analysis)
  - companion, free online: Hyndman and Athanasopoulos (2021), Ch. 2 (time series graphics) and Ch. 3 (transformations)
- Theory: Brockwell and Davis (2016), Ch. 1; Hamilton (1994), Ch. 3
- Python Quantlets of this chapter: Quantlets/Ch_01
  - ACF, PACF and portmanteau tests with `statsmodels`
- Lecture notebook: open in Google Colab
- Video course: Applied Time Series Analysis with Python

Four series of this course



- Romanian real GDP, 1995Q1–2026Q2 (Eurostat); HICP inflation (Eurostat); EUR/RON, the BNR reference rate; BET daily log returns since 2000
- Four frequencies (quarterly, monthly, daily) and four kinds of behaviour

TSA_ch1_real_series

Interpreting the four series

- **GDP**: from 16.5 to 45.9 bn EUR per quarter; a trend and a seasonal pattern whose swing grows with the level
  - the mean of the series changes with time
- **Inflation**: 102% in June 1997, −3.0% in May 2016, 6.1% in August 2026
  - a slow, persistent movement: this month resembles the previous one
- **EUR/RON**: from 3.60 to 5.26 lei; it wanders without returning to a fixed level
  - a candidate for a random walk (Section 3)
- **BET returns**: they oscillate around a constant level close to 0; the extremes: −13.1% on 7 January 2009, +10.6% on 10 May 2010
  - but the amplitude changes: standard deviation 3.62% in September 2008–March 2009, 0.66% in 2017

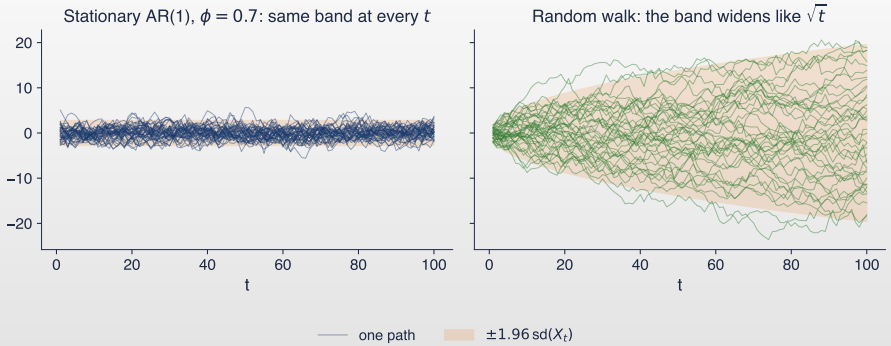
A time series as one realisation

- **Time series**: observations $x_1, x_2, \ldots, x_T$ ordered in time, at equal intervals
  - T = sample size; t = time index (day, month, quarter)
- The key idea: each x_t is the observed value of a **random variable** X_t
  - the BET return of tomorrow is unknown today: it has a distribution of possible values
  - history ran only once: we see one value from each distribution
- Consequence: we need assumptions that link the distributions at different dates
  - otherwise T observations would estimate T different means

Stochastic process

- **Definition**: a **stochastic process** is a family of random variables $\{X_t : t \in \mathcal{T}\}$ defined on the same probability space $(\Omega, \mathcal{F}, P)$
  - $\mathcal{T} = \mathbb{Z}$ or $\{1, \ldots, T\}$: **discrete time**, the case of this course
  - Ω: the set of possible "histories" ω; P: their probabilities
- Two ways to look at $X_t(\omega)$
  - fix t: X_t is a random variable (all values possible at date t)
  - fix ω: $t \mapsto X_t(\omega)$ is a **path** (trajectory, realisation): what we observe
- The process is described by its **finite-dimensional distributions**: the joint distribution of $(X_{t_1}, \ldots, X_{t_k})$ for any dates $t_1, \ldots, t_k$

An ensemble of paths



- Left: 40 paths of $X_t = 0.7X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$ i.i.d. (independent and identically distributed); right: 40 paths of $X_t = X_{t-1} + \varepsilon_t$, $X_0 = 0$
- The shaded band: ± 1.96 standard deviations of X_t at each t

TSA_ch1_processes

Interpreting the ensemble

- A vertical slice at date t shows the **distribution of** X_t; one path shows **one history**
  - the moments of the process (mean, variance) are averages across paths, at a fixed t
- Left: the same band at every t: $\mathrm{Var}(X_t) = 1/(1 - 0.7^2) = 1.96$, band ± 2.74
  - the distribution of X_t does not depend on t: a first picture of **stationarity**
- Right: the band widens; across 5000 simulated paths, $\mathrm{Var}(X_{100}) = 100.8$ (theory: 100)
  - the distribution changes with t: the random walk is **not stationary**

Mean, autocovariance and autocorrelation functions

- For a process with $E[X_t^2] < \infty$ for every t:
  - **mean function**: $\mu_t = E[X_t]$
  - **autocovariance function**: $\gamma(t,s) = \mathrm{Cov}(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)]$
  - **autocorrelation function**: $\rho(t,s) = \gamma(t,s)/\sqrt{\gamma(t,t)\,\gamma(s,s)}$
- "Auto": the covariance of the series with itself at two dates
  - $\gamma(t,t) = \mathrm{Var}(X_t)$; $\rho(t,t) = 1$; $|\rho(t,s)| \leq 1$ (Cauchy–Schwarz inequality)
- **Example**: $X_t = a + bt + \varepsilon_t$, ε_t i.i.d. with mean 0 and variance σ^2
  - $\mu_t = a + bt$ depends on t; $\gamma(t,t) = \sigma^2$, $\gamma(t,s) = 0$ for $t \neq s$

Recap: Time series and stochastic processes

- A time series is one path of a stochastic process $\{X_t\}$
- Moments are averages across paths at fixed dates: μ_t, $\gamma(t,s)$, $\rho(t,s)$
- With one path, we must assume these functions do not change with time

Strict stationarity

- **Definition**: $\{X_t\}$ is **strictly stationary** if, for every $k \geq 1$, all dates $t_1, \ldots, t_k$ and every shift h:
  - $(X_{t_1}, \ldots, X_{t_k}) \overset{d}{=} (X_{t_1+h}, \ldots, X_{t_k+h})$, where $\overset{d}{=}$ means equal joint distributions
- Consequences
  - $k = 1$: all X_t have the same distribution
  - $k = 2$: the joint distribution of (X_t, X_{t+h}) depends only on the lag h
- Example: an i.i.d. sequence is strictly stationary
  - the definition concerns entire distributions; the moments need not even exist (an i.i.d. Cauchy sequence)
- Hard to check from data: we would need all joint distributions

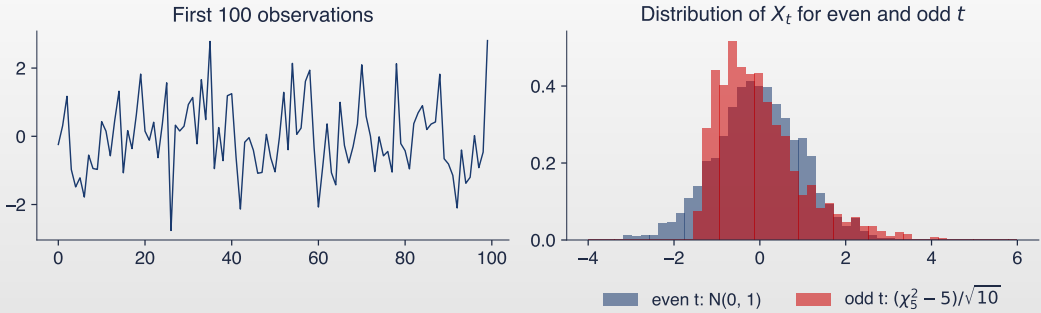
Weak stationarity

- **Definition** Khintchine (1934): $\{X_t\}$ is **weakly stationary** (covariance stationary, second-order stationary) if
  - (i) $E[X_t^2] < \infty$ for every t
  - (ii) $E[X_t] = \mu$, the same for every t
  - (iii) $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ depends only on the lag h, not on t
- Then the functions have one argument, the **lag** h:
  - $\gamma(h) = \text{Cov}(X_t, X_{t+h})$, $\gamma(0) = \text{Var}(X_t)$
  - $\rho(h) = \gamma(h)/\gamma(0)$: the **ACF** (autocorrelation function)
- In this course, "stationary" means weakly stationary unless stated otherwise

Strict and weak stationarity compared

- **Strict + finite variance $\Rightarrow$ weak**
  - equal distributions give equal means; equal joint distributions of (X_t, X_{t+h}) give equal covariances
- **Weak $\not\Rightarrow$ strict**
  - weak stationarity fixes only the first two moments; the shape of the distribution may change (next slide)
- **Strict $\not\Rightarrow$ weak** if the variance is infinite
  - i.i.d. Student-t with 2 degrees of freedom: strictly stationary, no finite variance
- **Gaussian processes**: all finite-dimensional distributions are multivariate Normal
  - a Normal distribution is fixed by its means and covariances, so for them weak $\Leftrightarrow$ strict

Weakly but not strictly stationary



- Independent X_t: $N(0,1)$ for even t, $(\chi^2_5 - 5)/\sqrt{10}$ for odd t; both have mean 0 and variance 1
- Sample of 4000: variance 0.99 (even) and 1.00 (odd); skewness -0.02 and 1.30 (theory: 0 and 1.26)

TSA_ch1_processes

Interpreting the counterexample

- Weakly stationary: $E[X_t] = 0$, $\gamma(0) = 1$, $\gamma(h) = 0$ for $h \neq 0$, at every t
  - the ACF cannot tell this process from Gaussian white noise
- Not strictly stationary: the distribution of X_t alternates between symmetric and right-skewed
  - $P(X_t < -1.6)$: about 5% for even t, 0 for odd t ($X_t \geq -5/\sqrt{10} = -1.58$)
- Lesson: second moments summarise dependence, but not the whole distribution; tail risk needs more (Chapter 5)

Properties of the autocovariance function

- For a weakly stationary process:
  - **symmetry**: $\gamma(-h) = \gamma(h)$, since $\mathrm{Cov}(X_t, X_{t-h}) = \mathrm{Cov}(X_{t-h}, X_t)$
  - **bound**: $|\gamma(h)| \leq \gamma(0)$, so $|\rho(h)| \leq 1$
  - **non-negative definiteness**: $\sum_{i,j} a_i a_j \gamma(i-j) \geq 0$ for all real $a_1, \ldots, a_n$
- Proof of the third property
  - $0 \leq \mathrm{Var}\left(\sum_i a_i X_i\right) = \sum_{i,j} a_i a_j \mathrm{Cov}(X_i, X_j) = \sum_{i,j} a_i a_j \gamma(i-j)$
- Consequence: not every sequence is an ACF
  - $\rho(1) = 0.9$, $\rho(2) = -0.9$ is impossible: with $a = (1, -1, 1)$, $\mathrm{Var}(X_1 - X_2 + X_3) = \gamma(0)(3 - 4\rho(1) + 2\rho(2)) = -2.4\,\gamma(0) < 0$

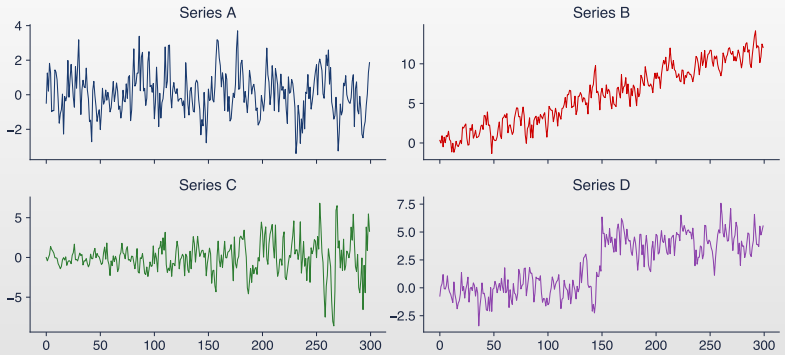
Worked example: a moving average of order 1

- $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, $\{\varepsilon_t\}$ uncorrelated, mean 0, variance σ^2: the **MA(1)** (moving average) process
  - θ: the weight of the previous shock; ε_t: the shock of period t
  - mean: $E[X_t] = 0$
  - $\gamma(0) = E[(\varepsilon_t + \theta\varepsilon_{t-1})^2] = \sigma^2(1+\theta^2)$ (the cross term has mean 0)
  - $\gamma(1) = E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t+1} + \theta\varepsilon_t)] = \theta\sigma^2$
  - $\gamma(h) = 0$ for $|h| \geq 2$: no common shock
- None of these depends on t: the MA(1) is weakly stationary for every θ
  - $\rho(1) = \theta/(1+\theta^2)$; with $\theta = 0.6$, $\sigma^2 = 1$: $\gamma(0) = 1.36$, $\rho(1) = 0.441$
  - $|\rho(1)| \leq 0.5$ for any θ (maximum at $\theta = 1$)

Worked example: an autoregression of order 1

- $X_t = \phi X_{t-1} + \varepsilon_t$, $|\phi| < 1$: the **AR(1)** (autoregressive) process; assume it is stationary and solve for the moments
  - ϕ: the weight of the previous value; ε_t: uncorrelated shocks with mean 0 and variance σ^2
  - mean: $\mu = \phi\mu + 0$, so $\mu = 0$
  - variance: $\gamma(0) = \phi^2\gamma(0) + \sigma^2$, so $\gamma(0) = \sigma^2/(1-\phi^2)$
  - multiply by X_{t-h} and take expectations: $\gamma(h) = \phi\gamma(h-1)$, so $\rho(h) = \phi^{|h|}$
- Numbers: $\phi = 0.8$, $\sigma^2 = 1$
  - $\gamma(0) = 2.78$; $\rho(1) = 0.8$, $\rho(5) = 0.328$, $\rho(10) = 0.107$: geometric decay
- With $|\phi| \geq 1$ the variance formula fails: no stationary solution of this form (Chapters 2–3)

Question for the room: which series are stationary?



- Four simulated series, 300 observations each; the same AR(1) noise with $\phi = 0.5$ underlies all of them
- Which of the conditions (ii) and (iii) of weak stationarity fails for each series?

TSA_ch1_processes

Answer: only series A is stationary

- **A**: stationary AR(1): constant mean and variance, the ACF depends only on the lag
- **B**: deterministic trend, $\mu_t = 0.04\,t$: condition (ii) fails
  - a **trend-stationary** series: stationary around a line, after subtracting the trend
- **C**: the standard deviation grows 7.5 times along the sample: condition (iii) fails at $h = 0$
- **D**: a level shift of $+4$ at $t = 150$: condition (ii) fails
  - a structural break; in real data: a change of regime or of policy (the Nile, Section 8)
- In practice, a plot of the series is the first check; formal tests come in Chapter 3

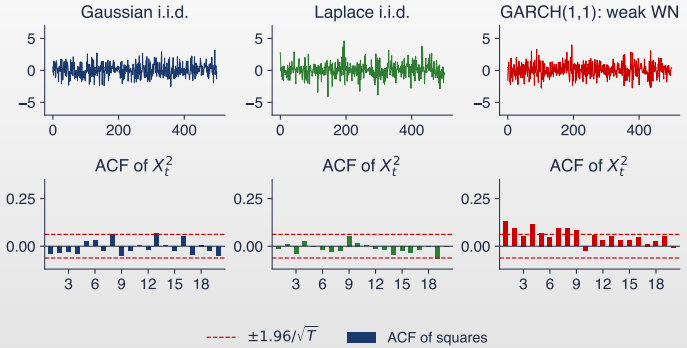
Recap: Stationarity

- Strict: all joint distributions are shift-invariant; weak: constant mean, γ depends only on the lag
- Strict + finite variance $\Rightarrow$ weak; for Gaussian processes they coincide
- MA(1) is always stationary; AR(1) is stationary for $|\phi| < 1$, with $\rho(h) = \phi^{|h|}$
- Trends, changing variance and breaks violate stationarity

White noise

- **Definition**: $\{\varepsilon_t\}$ is **white noise**, $\varepsilon_t \sim \text{WN}(0, \sigma^2)$, if
  - $E[\varepsilon_t] = 0$, $\text{Var}(\varepsilon_t) = \sigma^2$ and $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$
  - so $\rho(0) = 1$ and $\rho(h) = 0$ for every $h \neq 0$
- Three versions, from weakest to strongest
  - **weak** white noise: only uncorrelated; non-linear dependence is allowed (for example, in the squares)
  - **i.i.d.** (strong) white noise: independent and identically distributed
  - **Gaussian** white noise: i.i.d. $N(0, \sigma^2)$; for jointly Normal variables, uncorrelated means independent
- White noise is the building block of every model in this course: ARMA, ARIMA, GARCH, VAR

Three kinds of white noise



- Top: 500 of 1000 simulated values, variance 1 in all three; bottom: the ACF of the squares X_t^2
- GARCH(1,1) (Chapter 5): $\sigma_t^2 = 0.05 + 0.10\,\varepsilon_{t-1}^2 + 0.85\,\sigma_{t-1}^2$, $\varepsilon_t = \sigma_t z_t$; σ_t^2: the variance of day t given the past; z_t: i.i.d. $N(0,1)$

TSA_ch1_processes

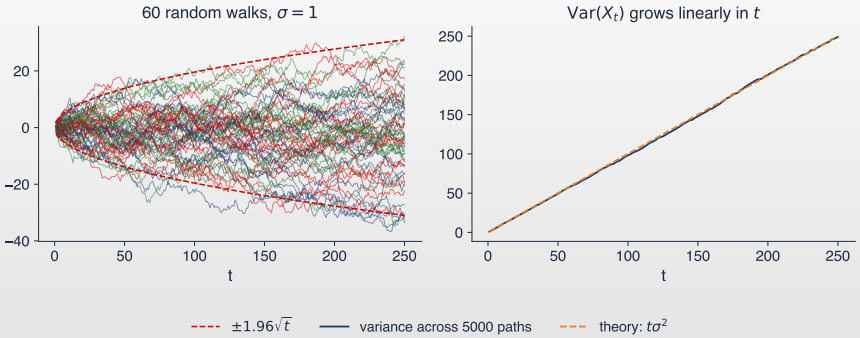
Interpreting the three white noises

- All three are white noise by construction; Ljung–Box $Q^*(10)$ (Section 6) on X_t: p = 0.278 (Gaussian), 0.986 (Laplace), 0.049 (GARCH)
  - the GARCH series is rejected at 5% although it is uncorrelated: with changing variance the usual test rejects too often
- Laplace: heavier tails (kurtosis 5.2 against 3.0), but still i.i.d.: the squares are uncorrelated (p = 0.680)
  - heavy tails alone do not create dependence
- GARCH: the squares are strongly correlated, $Q^*(10) = 73.2$, $p < 0.001$
  - a weak white noise: unpredictable sign and level, predictable size (volatility clustering)
  - daily returns behave like this (BET, Section 5)

The random walk

- **Definition**: $X_t = X_{t-1} + \varepsilon_t$, $t \geq 1$, $X_0 = 0$, $\varepsilon_t \sim \text{WN}(0, \sigma^2)$; by substitution, $X_t = \sum_{i=1}^{t} \varepsilon_i$
  - every shock stays in the level forever: $\partial X_{t+h} / \partial \varepsilon_t = 1$ for all $h \geq 0$
- Moments (uncorrelated shocks are enough)
  - $E[X_t] = 0$; $\text{Var}(X_t) = \sum_{i=1}^{t} \text{Var}(\varepsilon_i) = t\sigma^2$
  - for $t \leq s$: $\text{Cov}(X_t, X_s) = \text{Cov}\big(\sum_{i \leq t} \varepsilon_i, \sum_{j \leq s} \varepsilon_j\big) = t\sigma^2$
  - $\rho(t, s) = t/\sqrt{ts} = \sqrt{t/s}$
- **Not stationary**: the variance grows with t, and the correlation depends on t, not only on $s - t$
  - example: $\rho(X_{100}, X_{110}) = \sqrt{100/110} = 0.953$, but $\rho(X_{10}, X_{20}) = \sqrt{0.5} = 0.707$

Random walks and their variance



- Left: 60 paths with $\sigma = 1$ and the band $\pm 1.96\sqrt{t}$; right: the variance of X_t across 5000 paths and the line $t\sigma^2$
- At $t = 250$: simulated variance 249.1 (theory: 250); share of paths outside the band: 5.0%

TSA_ch1_processes

Interpreting the random walks

- The spread grows like $\sqrt{t}$: uncertainty about the level accumulates
  - a forecast of the level 100 days ahead has variance $100\sigma^2$
- Paths look like they have trends and cycles, although nothing but chance drives them
  - the same lesson as Slutzky (1937): sums of random shocks can look like economic cycles
- Simulated $\rho(X_{100}, X_{110}) = 0.953$ against 0.953 in theory
  - levels far apart in time are still highly correlated: very slow ACF decay

The random walk with drift

- $X_t = c + X_{t-1} + \varepsilon_t$, $X_0 = 0$: $X_t = ct + \sum_{i=1}^{t} \varepsilon_i$; c is the **drift**
  - $E[X_t] = ct$: a linear trend in the mean; $\mathrm{Var}(X_t) = t\sigma^2$ as before
- The first difference is stationary: $\Delta X_t = X_t - X_{t-1} = c + \varepsilon_t$
  - a **difference-stationary** series, with a **stochastic trend**
- Compare with the trend-stationary $Y_t = a + bt + u_t$, u_t stationary; a: intercept, b: slope of the trend
  - both have a linear mean; in Y_t shocks fade, in X_t they last forever
  - telling the two apart is the unit-root problem of Chapter 3
- Log prices of stocks and indices are often close to a random walk with a small positive drift

The Bucharest Stock Exchange

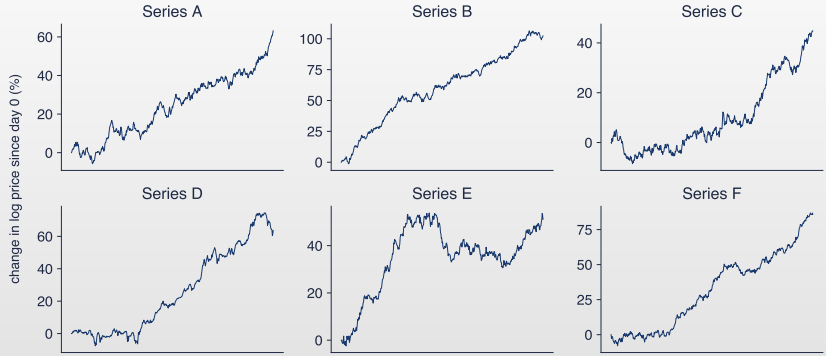


The Bucharest Stock Exchange (BVB), 2024

Photo: Corina Chitu (2024); CC BY-SA 4.0; Wikimedia Commons

- The BET index: the reference index of the BVB (Bucharest Stock Exchange), computed since 19 September 1997
- Daily closes since 2000: 6 670 log returns $r_t = 100\,\Delta \ln P_t$ (P_t: the closing level on day t); standard deviation 1.40% per day
- Is the BET log price a random walk? A first, visual test on the next slide

Question for the room: which one is real?



- One panel: the BET log price over its last 500 trading days; five panels: random walks with the same drift and the same daily standard deviation
- Which panel is the BET?

TSA_ch1_processes

Answer: series D

- Series D is the BET, 11 September 2024 to 18 September 2026: mean daily change 0.13%, standard deviation 0.99%
  - by eye, it cannot be told apart from the random walks
- The data are not exactly a random walk: the first-order autocorrelation of the daily changes is 0.12; Ljung–Box $Q^*(10)$ p-value 0.018
  - small departures, invisible in the level, show up in the ACF of the differences
- Lesson: study the changes, not the level; the tools are the ACF and the portmanteau tests (Sections 5–6)

Recap: White noise and the random walk

- White noise: mean 0, constant variance, no autocorrelation; weak, i.i.d. or Gaussian
- Uncorrelated is not independent: the squares of a weak white noise can be correlated
- Random walk: $\mathrm{Var}(X_t) = t\sigma^2$, $\rho(t,s) = \sqrt{t/s}$: not stationary; its difference is white noise
- Trend-stationary and difference-stationary series look alike but react differently to shocks

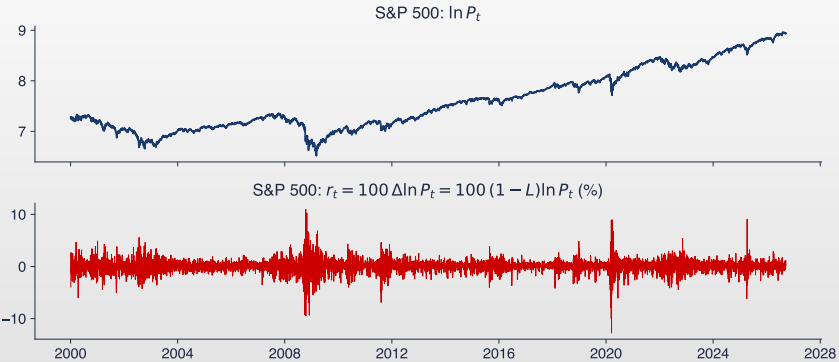
The lag operator

- **Definition**: $LX_t = X_{t-1}$; powers: $L^k X_t = X_{t-k}$, $L^0 = 1$ (also written B, the backshift operator)
  - a constant is unchanged: $Lc = c$
- **Lag polynomials**: $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$
  - p: the degree (the largest lag); $\phi_1, \ldots, \phi_p$: real coefficients; $\phi(L)X_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}$
  - AR(1): $(1 - \phi L)X_t = \varepsilon_t$; MA(1): $X_t = (1 + \theta L)\varepsilon_t$
  - they multiply like ordinary polynomials: $(1 - L)(1 + L) = 1 - L^2$
- **Inversion**: for $|\phi| < 1$, $(1 - \phi L)^{-1} = 1 + \phi L + \phi^2 L^2 + \ldots$ (a geometric series)
  - so the AR(1) is $X_t = \sum_{j \geq 0} \phi^j \varepsilon_{t-j}$: an infinite moving average
- The language of Chapters 2–4: ARMA, ARIMA and SARIMA models are written with lag polynomials

Differencing

- **First difference**: $\Delta X_t = (1-L)X_t = X_t - X_{t-1}$; **second**: $\Delta^2 X_t = (1-L)^2 X_t = X_t - 2X_{t-1} + X_{t-2}$
  - **seasonal difference** with period s: $\Delta_s X_t = (1-L^s)X_t = X_t - X_{t-s}$ ($s = 4$ for quarters, $s = 12$ for months)
- **Worked example**: $x = (10, 12, 15, 14, 18)$
  - $\Delta x = (2, 3, -1, 4)$; $\Delta^2 x = (1, -4, 5)$; each difference loses one observation
- What differencing removes
  - a linear trend: $\Delta(a + bt) = b$; a quadratic trend needs Δ^2
  - a random walk: $\Delta X_t = \varepsilon_t$; a seasonal pattern that repeats exactly: Δ_s
- **Integrated process**: $X_t \sim I(d)$ if $\Delta^d X_t$ is stationary and $\Delta^{d-1} X_t$ is not
  - white noise is $I(0)$, the random walk is $I(1)$; testing d: Chapter 3

S&P 500: the level and its difference



- Top: the log of the S&P 500 index since 2000; bottom: the daily log returns, 6 714 days to 18 September 2026

TSA_ch1_transformations

Interpreting the S&P 500 transformation

- Log price: sample ACF 0.999 at lag 1 and still 0.973 at lag 50
  - the behaviour of a random walk: the level remembers everything
- Log returns: mean 0.025% per day, standard deviation 1.21%; ACF at lag 1: -0.099
  - a constant level: one difference removed the stochastic trend
- Not yet white noise in the strict sense: quiet and turbulent periods alternate (2008, 2020)
  - the returns are close to a weak white noise; their variance is the subject of Chapter 5

The Wold decomposition

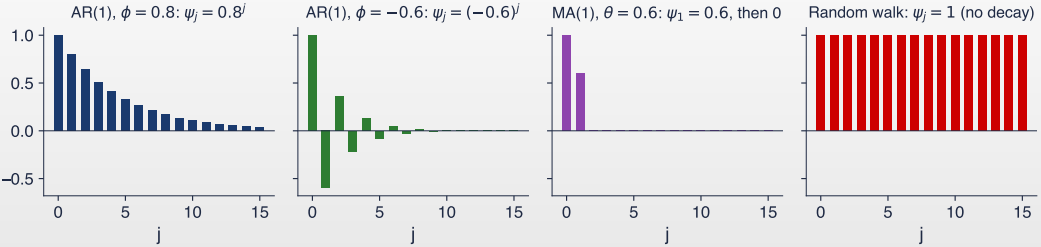
- **Theorem** Wold (1938): every weakly stationary process with mean 0 can be written as
  - $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \eta_t$, with $\psi_0 = 1$, $\sum_j \psi_j^2 < \infty$
  - ε_t: white noise, the error of the best linear forecast of X_t from its past (the **innovation**)
  - η_t: a deterministic part, perfectly predictable from the distant past (for example, a fixed cycle)
- Meaning
  - any stationary series is a weighted sum of past shocks: MA(∞)
  - the weights ψ_j = the effect of a shock after j periods (impulse response)
  - ARMA models (Chapter 2) approximate the infinite ψ_j with few parameters



Herman Wold, Uppsala, 1969

Photo: Uppsala-Bild (1969); CC BY 4.0; Wikimedia Commons

Wold weights of four processes



- ψ_j, $j = 0, \ldots, 15$: AR(1) with $\phi = 0.8$ and $\phi = -0.6$, MA(1) with $\theta = 0.6$, and the random walk
- AR(1), $\phi = 0.8$: $\psi_{\mathbf{5}} = 0.328$, $\psi_{\mathbf{10}} = 0.107$; $\mathrm{Var}(X_t) = \sigma^{\mathbf{2}} \sum_j \psi_j^{\mathbf{2}} = 2.78\,\sigma^{\mathbf{2}}$

TSA_ch1_wold_ergodicity

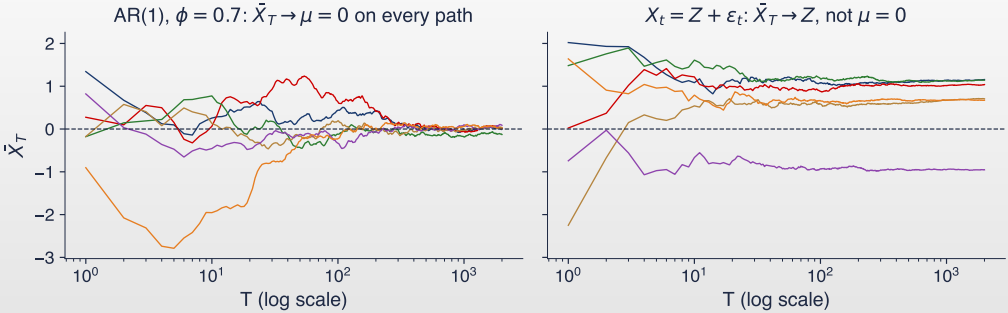
Interpreting the Wold weights

- Stationary processes: $\psi_j \to 0$, so shocks fade and the variance $\sigma^2 \sum_j \psi_j^2$ is finite
  - $\phi < 0$: the effect alternates in sign; MA(1): the effect stops after one period
- Random walk: $\psi_j = 1$ for all j, $\sum_j \psi_j^2 = \infty$: no Wold form, no stationarity
  - the economic meaning: a permanent shock (a technology shock to GDP, news in a price)
- The ACF is fixed by the weights: $\gamma(h) = \sigma^2 \sum_j \psi_j \psi_{j+h}$

Ergodicity

- Stationarity says the moments are constant; we still need to estimate them from **one** path
  - ensemble mean: $\mu = E[X_t]$ (across paths); time mean: $\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t$ (along one path)
- **Definition**: a stationary process is **ergodic for the mean** if $\bar{X}_T \to \mu$ (in mean square) as $T \to \infty$
  - a sufficient condition: $\gamma(h) \to 0$ as $h \to \infty$ (correlation dies out)
  - if $\sum_h |\gamma(h)| < \infty$, then $T\,\mathrm{Var}(\bar{X}_T) \to \sum_{h=-\infty}^{\infty} \gamma(h)$
- **Counterexample**: $X_t = Z + \varepsilon_t$, $Z \sim N(0,1)$ drawn once, ε_t white noise
  - stationary, with $\gamma(h) = 1$ for every $h \neq 0$; $\bar{X}_T \to Z$, not $\mu = 0$

Time averages: ergodic and non-ergodic



- Running mean $\bar{X}_T$ along 6 paths, T up to 2 000: AR(1) with $\phi = 0.7$ (left) and $X_t = Z + \varepsilon_t$ (right)
- Final values: between -0.13 and 0.09 on the left; between -0.95 and 1.16 on the right, close to the drawn Z

TSA_ch1_wold_ergodicity

Interpreting ergodicity

- Left: every path "learns" the true mean 0; one long history is as good as many histories
  - convergence is slow when the correlation is strong: $\text{Var}(\bar{X}_T) \approx \gamma(0)(1+\phi)/((1-\phi)T)$
- Right: each path converges to its own Z; no amount of data from one path reveals $\mu = 0$
  - an economy with a permanent, unobserved "type" behaves like this
- Ergodicity cannot be tested from one path; it is assumed whenever we estimate moments from a single series

Recap: The lag operator, differencing and Wold

- $L^k X_t = X_{t-k}$; $\Delta = 1 - L$; $\Delta_s = 1 - L^s$; $I(d)$: stationary after d differences
- Wold: a stationary process is an MA(∞) of its innovations, with $\sum \psi_j^2 < \infty$
- Ergodicity lets time averages estimate ensemble moments; it needs correlations that die out

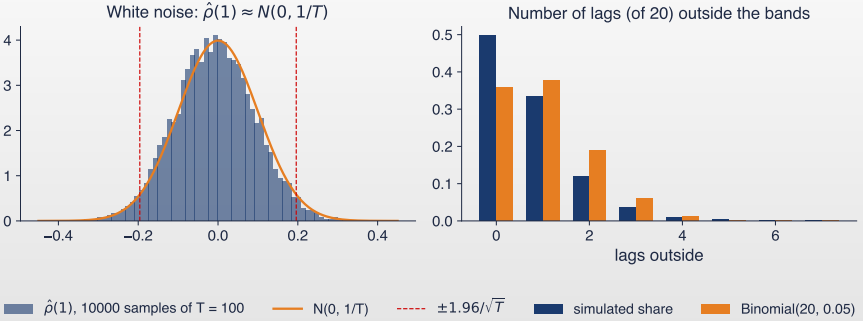
Sample autocovariance and autocorrelation

- From $x_1, \ldots, x_T$, with the sample mean $\bar{x} = \frac{1}{T}\sum_t x_t$:
  - $\hat{\gamma}(h) = \frac{1}{T}\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x})$, $0 \le h < T$
  - $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$: the **sample ACF**; the plot of $\hat{\rho}(h)$ against h is the **correlogram**
- Why divide by T and not by $T - h$?
  - the sequence $\hat{\gamma}(h)$ is then non-negative definite, like a true autocovariance Brockwell and Davis (2016)
  - the price: a small bias towards 0 at large lags; use $h \le T/4$
- **Worked example**: $x = (3, 5, 4, 6, 7)$, $\bar{x} = 5$, deviations $(-2, 0, -1, 1, 2)$
  - $\hat{\gamma}(0) = 10/5 = 2.0$; $\hat{\gamma}(1) = (0 + 0 - 1 + 2)/5 = 0.2$; $\hat{\gamma}(2) = (2 + 0 - 2)/5 = 0.0$
  - $\hat{\rho}(1) = 0.10$, $\hat{\rho}(2) = 0.00$; with $T = 5$ the band is ± 0.88: far too few data

Confidence bands for the sample ACF

- **Large-sample result** of Bartlett (1946): if X_t is i.i.d. with finite variance, for large T
  - $\hat{\rho}(1), \ldots, \hat{\rho}(m)$ are approximately independent $N(0, 1/T)$
  - 95% band: $\pm 1.96/\sqrt{T}$; a bar outside it rejects $\rho(h) = 0$ at 5%, **for that lag alone**
- For an MA(q), at lags $h > q$: $\mathrm{Var}(\hat{\rho}(h)) \approx \frac{1}{T}\left(1 + 2\sum_{k=1}^{q} \rho(k)^2\right)$
  - the wider bands of `statsmodels` `plot_acf` (Chapter 2)
- Limitations of the band
  - with 20 lags and no correlation, about one bar is expected outside the band by chance
  - the band assumes i.i.d. data; for a weak white noise (GARCH returns) the true band is wider

Sampling distribution of the sample ACF



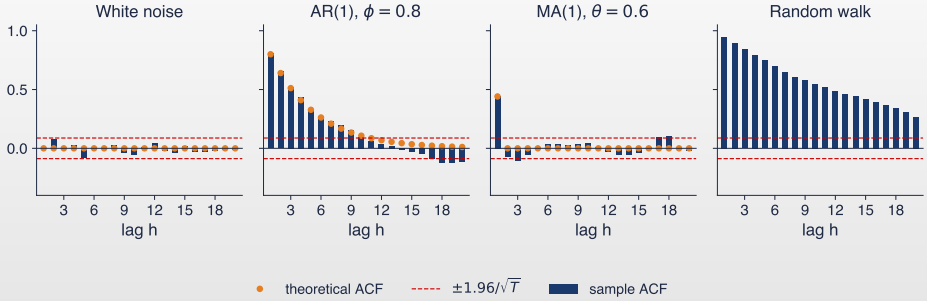
- 10000 samples of Gaussian white noise with $T = 100$: $\hat{\rho}(1)$ (left) and the number of lags, out of 20, outside ± 0.196 (right)

TSA_ch1_acf_pacf

Interpreting the simulation

- $\hat{\rho}(1)$: mean -0.008 (theory $-1/T = -0.01$), standard deviation 0.098 (theory $1/\sqrt{T} = 0.1$)
  - Bartlett's approximation is already good at $T = 100$
- At least one of 20 bars outside the band in 50% of the samples, although there is no correlation at all
  - independent tests would give $1 - 0.95^{20} = 64\%$; at distant lags $\mathrm{Var}(\hat{\rho}(h)) < 1/T$, so slightly fewer bars cross
- Do not over-read single bars: judge the pattern, or use a joint test (Section 6)

Theoretical and sample ACF of four processes



- $T = 500$ simulated values; bars: sample ACF; dots: theoretical ACF; band ± 0.088

TSA_ch1_acf_pacf

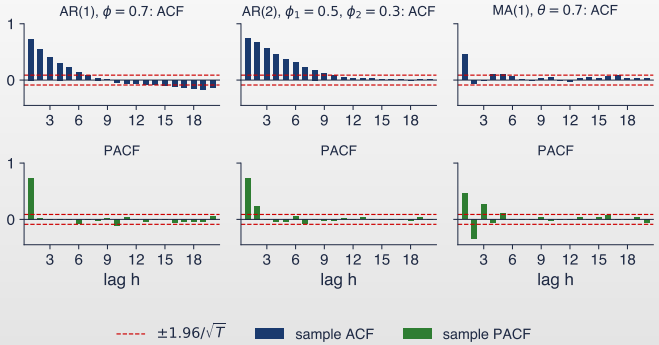
Interpreting the four correlograms

- White noise: all bars small; the largest, $\hat{\rho}(2) = 0.08$, is noise
- AR(1), $\phi = 0.8$: geometric decay, $\hat{\rho}(1) = 0.81$, $\hat{\rho}(2) = 0.65$ (theory 0.80 and 0.64)
- MA(1), $\theta = 0.6$: one spike, $\hat{\rho}(1) = 0.44$ (theory 0.44), then nothing: the ACF **cuts off** after lag 1
- Random walk: $\hat{\rho}(1) = 0.95$, still 0.55 at lag 10: the slow, almost linear decay of a non-stationary series
  - the sample ACF of a random walk has no theoretical counterpart: $\rho(h)$ is not defined

Partial autocorrelation

- **Definition**: the **PACF** at lag h, ϕ_{hh}, is the last coefficient of the best linear predictor of X_t from $X_{t-1}, \ldots, X_{t-h}$:
  - $X_t = \phi_{h1} X_{t-1} + \cdots + \phi_{hh} X_{t-h} + e_t$; $\phi_{h1}, \ldots, \phi_{hh}$: the coefficients of the regression on h lags; e_t: the prediction error
  - the correlation of X_t and X_{t-h} after removing the linear effect of the lags in between
- First two values
  - $\phi_{11} = \rho(1)$; $\phi_{22} = \dfrac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}$
  - the general case: the Durbin–Levinson recursion; the sample PACF uses $\hat{\rho}(h)$ in place of $\rho(h)$
- **Worked example**
  - AR(1): $\rho(2) = \rho(1)^2$, so $\phi_{22} = 0$: once X_{t-1} is known, X_{t-2} adds nothing
  - AR(2) with $\phi_1 = 0.5$, $\phi_2 = 0.3$: $\rho(1) = 0.714$, $\rho(2) = 0.657$, $\phi_{22} = (0.657 - 0.510)/(1 - 0.510) = 0.300 = \phi_2$
- Same band as the ACF under white noise: $\pm 1.96/\sqrt{T}$

ACF and PACF of three processes



- $T = 500$ simulated values of AR(1) ($\phi = 0.7$), AR(2) ($\phi_1 = 0.5$, $\phi_2 = 0.3$) and MA(1) ($\theta = 0.7$); top: sample ACF; bottom: sample PACF

TSA_ch1_acf_pacf

Interpreting the ACF and PACF

- AR(1): the ACF decays; the PACF has one spike, $\hat{\phi}_{11} = 0.73$, then $\hat{\phi}_{22} = 0.03$
  - AR(2): two PACF spikes, $\hat{\phi}_{22} = 0.24$ (theory 0.3), then $\hat{\phi}_{33} = 0.01$
- MA(1): the mirror image; one ACF spike ($\hat{\rho}(1) = 0.46$), a PACF that alternates and decays (-0.34, 0.26, ...)
- The identification rule of Chapter 2 Huang and Petukhina (2022):
  - PACF cuts off after p: AR(p); ACF cuts off after q: MA(q); both decay: ARMA

Real data: three markets

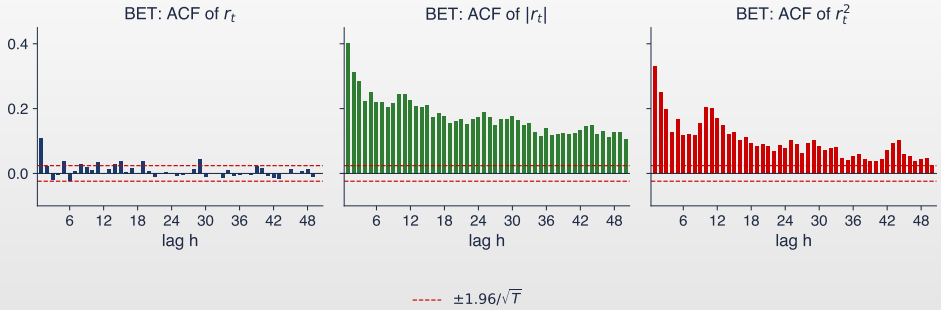


The National Bank of Romania (BNR), publisher of the EUR/RON reference rate

Photo: Ștefan Jurcă (2015); CC BY 2.0; Wikimedia Commons

- Next: the ACF of BET daily log returns since 2000, of their absolute values and of their squares
- The same analysis for EUR/RON (the BNR reference rate) and the S&P 500: Seminar 1, task B2
- Question: are BET returns white noise? Of which kind?

BET: ACF of returns, absolute and squared returns



- Daily log returns, 6 670 days, lags 1–50; band ± 0.024

TSA_ch1_acf_pacf

Interpreting the BET correlograms

- Returns: $\hat{\rho}(1) = 0.109$, clearly outside the band; $Q^*(10) = 108$
  - statistically significant but small: yesterday's return explains about $\hat{\rho}(1)^2 = 1.2\%$ of today's variance
  - a typical sign of a less liquid market (non-synchronous trading)
- Absolute returns: $\hat{\rho}(1) = 0.40$ and still 0.11 at lag 50; squares: $\hat{\rho}(1) = 0.33$
  - large moves follow large moves: volatility clustering, a long memory in the size of returns
- Conclusion: BET returns are close to a **weak** white noise; the mean is hardly predictable, the variance is (Chapter 5)

Recap: Sample ACF and PACF

- $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$, with divisor T; use lags up to about $T/4$
- Under i.i.d. data: $\hat{\rho}(h) \approx N(0, 1/T)$, band $\pm 1.96/\sqrt{T}$; 1 bar in 20 crosses by chance
- PACF: direct effect of lag h; AR(p) cuts off in the PACF, MA(q) in the ACF
- Returns: small linear, strong non-linear dependence

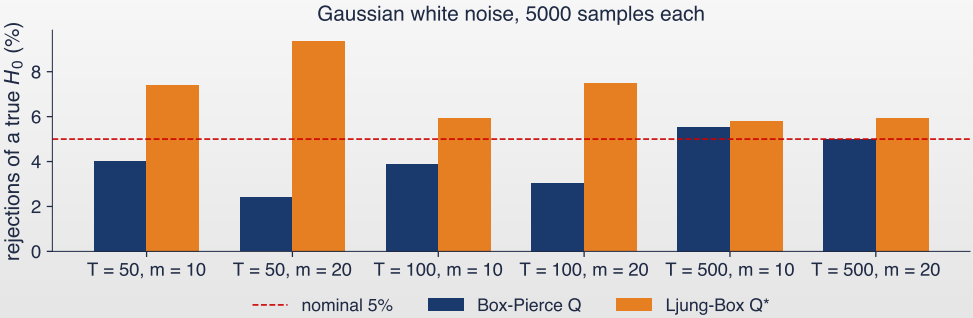
Testing many autocorrelations at once

- H_0: $\rho(1) = \cdots = \rho(m) = 0$ (the series is white noise up to lag m); H_1: at least one $\rho(h) \neq 0$
  - a **portmanteau** test: one statistic "carries" m autocorrelations
- The **Box–Pierce statistic**, from Box and Pierce (1970): $Q(m) = T \sum_{h=1}^{m} \hat{\rho}(h)^2$
  - under H_0, each $\sqrt{T}\hat{\rho}(h) \approx N(0,1)$, independent: the sum of m squares is $\approx \chi^2(m)$
- The **Ljung–Box statistic**, from Ljung and Box (1978): $Q^*(m) = T(T+2) \sum_{h=1}^{m} \dfrac{\hat{\rho}(h)^2}{T-h}$
  - the weight $(T+2)/(T-h)$ corrects the small variance of $\hat{\rho}(h)$ at distant lags; same $\chi^2(m)$ limit
- Reject H_0 at 5% if $Q^* > \chi^2_{0.95}(m)$; for example $\chi^2_{0.95}(10) = 18.31$
  - $\chi^2(m)$: the chi-square distribution with m degrees of freedom; $\chi^2_{0.95}(m)$: its 95% quantile
  - p-value: the probability under H_0 of a statistic at least as large as the observed one; reject if it is below 0.05
  - on residuals of an ARMA(p,q): $\chi^2(m-p-q)$ (Chapter 2)

Worked example: the Ljung–Box statistic

- $T = 100$; $\hat{\rho}(1) = 0.25$, $\hat{\rho}(2) = 0.12$, $\hat{\rho}(3) = -0.08$; band ± 0.196
  - only $\hat{\rho}(1)$ is outside the band
- Box–Pierce: $Q(3) = 100\,(0.0625 + 0.0144 + 0.0064) = 8.33$
  - Ljung–Box: $Q^*(3) = 100 \cdot 102\left(\frac{0.0625}{99} + \frac{0.0144}{98} + \frac{0.0064}{97}\right) = 8.61$
- Critical value $\chi^2_{0.95}(3) = 7.81$: reject H_0; p-value 0.035
  - the series is not white noise; the dependence is concentrated at lag 1
- Choice of m: too small misses distant lags; too large dilutes the evidence
  - common choices: $m = 10$ for non-seasonal data, $m = 2s$ for seasonal data with period s Hyndman and Athanasopoulos (2021)

Box–Pierce and Ljung–Box in small samples



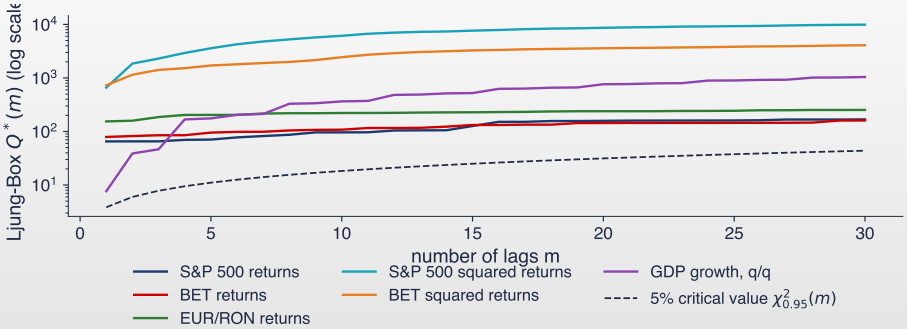
⊡ Share of 5000 Gaussian white-noise samples in which each test rejects a true H_0 at the 5% level

TSA_ch1_portmanteau

Interpreting the size simulation

- $T = 50$, $m = 20$: Box–Pierce rejects in 2.4% of samples, Ljung–Box in 9.4%
  - Box–Pierce is too lenient: it misses dependence; Ljung–Box errs the other way when m is large relative to T
- $T = 500$, $m = 20$: 5.0% and 5.9%: both close to 5%
- Practice: use Ljung–Box (the default in software), and keep m well below T (for example $m \leq T/5$)

Ljung–Box statistics of real series



- $Q^*(m)$ for $m = 1, \ldots, 30$ (log scale) against the 5% critical value $\chi^2_{0.95}(m)$ (dashed)

TSA_ch1_portmanteau

Interpreting the Ljung–Box statistics

Series	T	$\hat{\rho}(1)$	$\hat{\rho}(4)$	$Q^*(10)$	p
S&P 500, daily returns	6 714	−0.10	−0.03	96.3	< 0.001
S&P 500, squared returns	6 714	0.31	0.31	6132.8	< 0.001
BET, daily returns	6 670	0.11	−0.01	108.0	< 0.001
BET, squared returns	6 670	0.33	0.13	2440.0	< 0.001
EUR/RON, daily returns	5 352	0.17	−0.06	221.7	< 0.001
GDP Romania, q/q growth	125	−0.24	0.96	364.8	< 0.001
GDP Romania, y/y growth	122	0.63	0.03	75.0	< 0.001
HICP Romania, monthly inflation	260	0.40	0.12	126.7	< 0.001
Nile, yearly flow	100	0.50	0.24	88.1	< 0.001
Sunspots, yearly	309	0.82	−0.28	627.4	< 0.001

- Every series rejects white noise; for the squared returns the statistic is tens of times larger than for the returns
- Rejection says only that **some** correlation exists; the ACF shows where (GDP: lag 4, the season)

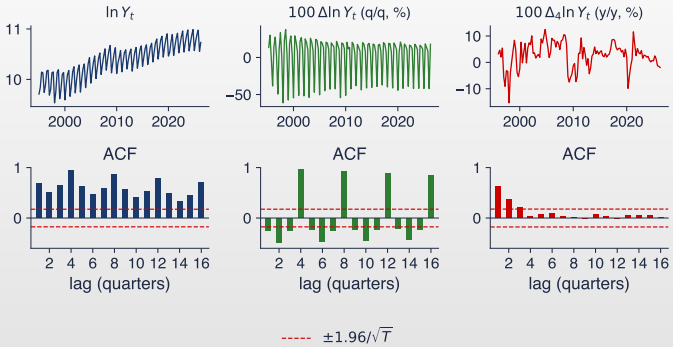
Recap: Portmanteau tests

- Box–Pierce $Q(m)$ and Ljung–Box $Q^*(m)$ sum the squared sample autocorrelations; both $\approx \chi^2(m)$ under white noise
- Ljung–Box has better size in small samples; keep m well below T
- A rejection is a signal to model, not a model; under GARCH effects the test of returns rejects too often

Reasons to transform a series

- Goals
  - stabilise the variance (seasonal swings that grow with the level)
  - turn exponential growth into a linear trend
  - remove a stochastic trend or a seasonal pattern (differencing)
- **Logarithm**: $\ln Y_t$; its difference is the growth rate in continuous terms
  - $100\,\Delta \ln Y_t \approx$ percentage change for small changes: $\ln 1.03 = 0.0296$, about 3%
  - not for large ones: $\ln 1.30 = 0.262$, $\ln 0.70 = -0.357$; log changes add up over time, percentages do not
- Order of operations for a trending, seasonal, positive series
  - first the log (variance), then the differences (trend, season); never the other way round

Romanian real GDP: log, quarterly and annual growth



- Quarterly, not seasonally adjusted, 2010 prices (Eurostat), 1995Q1–2026Q2; top: the series; bottom: their sample ACF

TSA_ch1_transformations

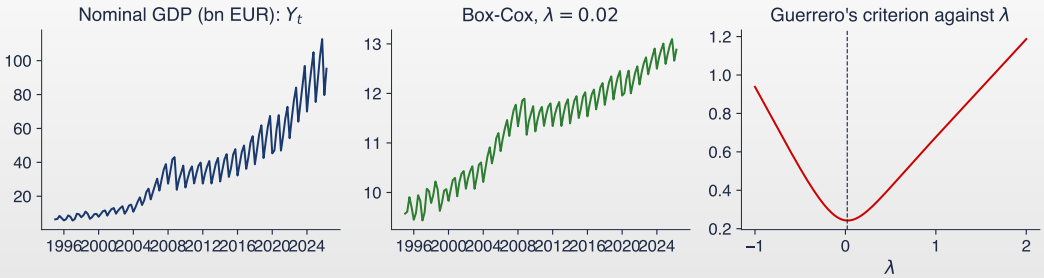
Interpreting the GDP transformations

- $\ln Y_t$: ACF 0.68 at lag 1 and 0.87 at lag 8, with peaks every 4 quarters: trend and season
- $\Delta \ln Y_t$: the trend is gone, the season dominates: $\hat{\rho}(4) = 0.96$, $\hat{\rho}(2) = -0.49$; standard deviation 27.5%
  - a quarter-on-quarter rate of unadjusted data mostly measures the season
- $\Delta_4 \ln Y_t$ (annual growth): mean 2.7%, the ACF dies out after 2–3 quarters ($\hat{\rho}(1) = 0.63$, $\hat{\rho}(4) = 0.03$)
  - close to stationary, with business-cycle persistence; 2020Q2: -9.9%; last quarter: -2.0%

The Box–Cox transformation

- **Definition** Box and Cox (1964), for $y > 0$: $w = \dfrac{y^\lambda - 1}{\lambda}$ for $\lambda \neq 0$, $w = \ln y$ for $\lambda = 0$
  - $\lambda = 1$: no change in shape; $\lambda = 0.5$: square root; $\lambda = 0$: log; $\lambda = -1$: inverse
  - continuous in λ: $(y^\lambda - 1)/\lambda \to \ln y$ as $\lambda \to 0$
- **Worked example**: $y = 100$ and $y = 400$
  - $\lambda = 0.5$: $w = 18$ and 38; $\lambda = 0$: $w = 4.605$ and 5.991
  - the smaller λ, the more large values are compressed
- Choosing λ: Guerrero (1993), as in Hyndman and Athanasopoulos (2021)
  - split the series into years; choose λ so that $\text{sd}_i/\text{mean}_i^{1-\lambda}$ is as constant as possible across years; sd_i, mean_i: the standard deviation and the mean of year i
- Forecasts are made for w and transformed back: $y = (\lambda w + 1)^{1/\lambda}$; the back-transformed value is the median, not the mean

Box–Cox for Romanian nominal GDP



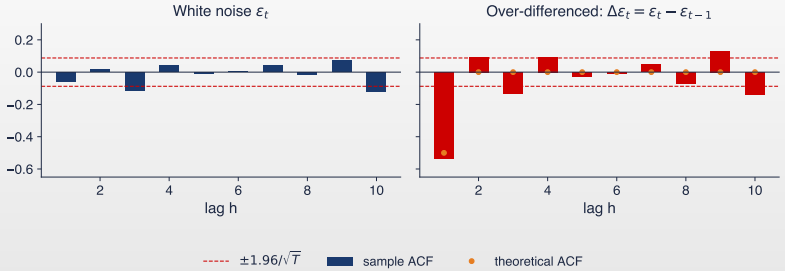
- Quarterly nominal GDP, not seasonally adjusted, from 6.4 to 95.2 bn EUR; Guerrero's criterion on yearly blocks of 4 quarters

TSA_ch1_transformations

Interpreting the Box–Cox choice

- Guerrero's $\hat{\lambda} = 0.02$: practically the logarithm (criterion 0.243 at $\hat{\lambda}$, 0.243 at $\lambda = 0$, 0.68 at $\lambda = 1$)
  - choose the round, interpretable value: $\lambda = 0$
- Why the log fits: the seasonal swing is a stable share of the level, about 39% in the first five years and 41% in the last five
  - multiplicative seasonality becomes additive after the log
- Box–Cox fixes the variance only; the trend and the season still need differencing

Over-differencing



⊡ Sample ACF of 500 values of white noise (left) and of its first difference (right)

TSA_ch1_transformations

Interpreting over-differencing

- $\Delta\varepsilon_t = \varepsilon_t - \varepsilon_{t-1}$ is an MA(1) with $\theta = -1$: $\rho(1) = -1/2$; here $\hat{\rho}(1) = -0.54$
  - the variance doubles: 0.95 becomes 2.01
- A strongly negative $\hat{\rho}(1)$, close to -0.5, after differencing: a sign that the series did not need it
  - the same happens to a trend-stationary series: $\Delta(a + bt + \varepsilon_t) = b + \varepsilon_t - \varepsilon_{t-1}$
- Difference as little as needed; Chapter 3 gives formal tests for the number of differences

Recap: Transformations

- Log first for positive series whose swings grow with the level; then differences
- $100\,\Delta \ln Y_t$: growth in %; Δ_4, Δ_{12}: annual growth of quarterly and monthly data
- Box–Cox with Guerrero's λ; for Romanian GDP $\hat{\lambda} \approx 0$, the log
- Over-differencing creates $\rho(1) \approx -0.5$ and inflates the variance

Case study: Yule (1927) and Slutzky (1937)

- **Question** of the 1920s: where do the regular cycles of sunspots and of business activity come from?
  - the dominant view: hidden periodic components (sums of sine waves), estimated with the periodogram
- Yule (1927): the sunspot number depends on its own last two values plus a random disturbance
  - $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$: the first autoregression; a damped pendulum hit by random shocks
  - a cycle without any sine wave: the shocks keep it alive, the dynamics give it its period
- Slutzky (1937): moving sums of purely random numbers (a lottery series) produce smooth, cycle-like waves
  - the moving average (MA) process, and a warning: cycles in data may be created by averaging
- Legacy: stationary processes built from white noise; Wold (1938) unified the two views (Section 4); Huang and Petukhina (2022), Ch. 3

The Nile and the sunspots

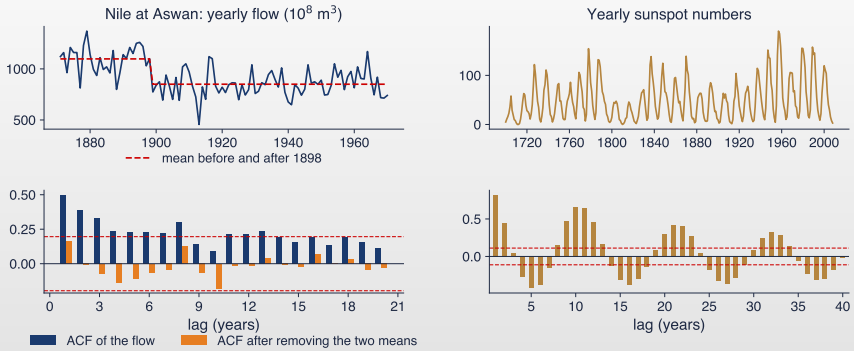


The Nilometer on Roda Island, Cairo: Nile levels have been recorded here for more than a thousand years

Photo: Berthold Werner (2010); CC BY-SA 3.0; Wikimedia Commons

- **Nile**: yearly flow at Aswan, 1871–1970, 100 observations; a drop around 1898 Cobb (1978)
- **Sunspots**: yearly numbers, 1700–2008, 309 observations; the series that inspired the autoregression Yule (1927)
- Both ship with `statsmodels`; both teach a lesson about reading an ACF

The Nile and the sunspots: series and ACF



⊡ Left: the Nile flow with its mean before and after 1898, and two ACFs; right: sunspot numbers and their ACF up to lag 40

TSA_ch1_textbook_series

Interpreting the two classic series

- Nile: $\hat{\rho}(1) = 0.50$ and a slow decay; $Q^*(10) = 88.1$ ($p < 0.001$)
  - after removing the two means (1098 and 850): $\hat{\rho}(1) = 0.16$, $Q^*(10) = 13.0$ ($p = 0.225$)
  - a slowly decaying ACF can come from a **break**, not from persistence or a unit root
- Sunspots: a damped wave, minimum -0.43 at lag 5, peak 0.66 at lag 10
  - the solar cycle of about 11 years; a stationary series with a cycle, modelled by Yule (1927) with an AR(2)
- Read an ACF together with the plot of the series and with what you know about how the data were produced

Possible contribution of AI

- **Code**: a first draft of a script that downloads a series, plots it with its ACF and PACF, and runs Ljung–Box for several m
- **Explanation**: a second explanation of a definition, or of why a correlogram looks as it does
- **Exploration**: the same diagnostics on many series at once (all BVB stocks, all EU inflation rates)
- Example prompt
  - `Write Python code that loads the Romanian monthly HICP from Eurostat, computes 100 times the monthly log difference since 2005, plots the series with its ACF and PACF up to lag 36, and reports the Ljung-Box statistic for m = 12 and m = 24.`

Checks you must run

- The divisor of $\hat{\gamma}(h)$ (T, not $T - h$) and whether lag 0 is included in the plot
- Which bands are drawn: $\pm 1.96/\sqrt{T}$ or Bartlett's MA bands; they answer different questions
- That Ljung–Box is applied to the stationary transformation, not to the level; and with $m - p - q$ degrees of freedom on residuals
- That "significant" is not read as "large" or "profitable"
- Dates, frequencies and units of the data (seasonally adjusted or not, index base, percent or decimals)
- Every cited reference: it must exist; check the DOI

Key takeaways

- A time series is one path of a stochastic process; inference from one path needs stationarity and ergodicity
- Weak stationarity: constant mean and an autocovariance that depends only on the lag
- White noise (weak, i.i.d., Gaussian) is the building block; the random walk is its cumulative sum and is not stationary
- Sample ACF and PACF with $\pm 1.96/\sqrt{T}$ bands; Ljung–Box tests many lags jointly
- Real series need transformations: log or Box–Cox for the variance, differences for trend and season
- Returns: nearly uncorrelated, but their squares are strongly correlated

Key formulas

Quantity	Formula
Autocovariance, ACF	$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}), \quad \rho(h) = \gamma(h)/\gamma(0)$
MA(1)	$\gamma(0) = \sigma^2(1+\theta^2), \quad \rho(1) = \theta/(1+\theta^2), \quad \rho(h) = 0,\ \lvert h\rvert \geq 2$
AR(1)	$\gamma(0) = \sigma^2/(1-\phi^2), \quad \rho(h) = \phi^{\lvert h\rvert}$
Random walk	$\mathrm{Var}(X_t) = t\sigma^2, \quad \mathrm{Cov}(X_t, X_s) = \min(t,s)\,\sigma^2$
Wold	$X_t = \sum_{j\geq 0}\psi_j \varepsilon_{t-j} + \eta_t, \quad \sum_j \psi_j^2 < \infty$
Sample ACF	$\hat\rho(h) = \sum_{t=1}^{T-h}(x_t-\bar x)(x_{t+h}-\bar x)/\sum_{t=1}^{T}(x_t-\bar x)^2, \quad \pm 1.96/\sqrt{T}$
PACF	$\phi_{11} = \rho(1), \quad \phi_{22} = (\rho(2)-\rho(1)^2)/(1-\rho(1)^2)$
Ljung–Box	$Q^*(m) = T(T+2)\sum_{h=1}^{m}\hat\rho(h)^2/(T-h) \sim \chi^2(m)$
Box–Cox	$w = (y^\lambda - 1)/\lambda, \quad w = \ln y\ (\lambda = 0)$

Self-assessment

- **Question**: $X_t = \varepsilon_t - 0.5\varepsilon_{t-1}$, $\sigma^2 = 4$. What are $\gamma(0)$ and $\rho(1)$?
  - **Answer**: $\gamma(0) = 4 \times 1.25 = 5$; $\rho(1) = -0.5/1.25 = -0.4$
- **Question**: a correlogram of 40 lags of a series with $T = 400$ shows two bars just outside ± 0.098. Is the series correlated?
  - **Answer**: not necessarily: about 2 of 40 bars cross by chance; run Ljung–Box
- **Question**: after differencing, $\hat{\rho}(1) = -0.48$ and nothing else is significant. What does this suggest?
  - **Answer**: over-differencing: the original series was probably already stationary
- Next: Chapter 2, ARMA models: identification with the ACF and PACF, estimation, diagnostics and forecasts

References (1/2)

- Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society*, 8(1), 27–41. doi:10.2307/2983611
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, 26(2), 211–243. doi:10.1111/j.2517-6161.1964.tb00553.x
- Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332), 1509–1526. doi:10.1080/01621459.1970.10481180
- Brockwell, P. J., & Davis, R. A. (1991). *Time Series: Theory and Methods* (2nd ed.). Springer. doi:10.1007/978-1-4419-0320-4
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (3rd ed.). Springer. doi:10.1007/978-3-319-29854-2
- 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
- Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, 12(1), 37–48. doi:10.1002/for.3980120104
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press. doi:10.2307/j.ctv14jx6sm

References (2/2)

- Huang, C., & Petukhina, A. (2022). *Applied Time Series Analysis and Forecasting with Python*. Springer. doi:10.1007/978-3-031-13584-2
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. otexts.com/fpp3
- Khintchine, A. (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, 109, 604–615. doi:10.1007/BF01449156
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303. doi:10.1093/biomet/65.2.297
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications* (4th ed.). Springer. doi:10.1007/978-3-319-52452-8
- Slutzky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica*, 5(2), 105–146. doi:10.2307/1907241
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksell. archive.org
- Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society A*, 226, 267–298. doi:10.1098/rsta.1927.0007