








Time Series Analysis

Chapter 3: Unit roots and ARIMA models



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)
- **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)
- **CUSUM**: CUmulative SUM (control chart)
- **DF**: Dickey–Fuller (test)
- **DF-GLS**: Dickey–Fuller test with GLS detrending
- **DS**: Difference-Stationary (series)
- **DW**: Durbin–Watson (statistic)
- **EU**: European Union
- **FRED**: Federal Reserve Economic Data

Acronyms used in this material (2/2)

- **GDP**: Gross Domestic Product
- **GLS**: Generalised Least Squares
- **GNP**: Gross National Product
- **HICP**: Harmonised Index of Consumer Prices
- **KPSS**: Kwiatkowski–Phillips–Schmidt–Shin (test)
- **MA**: Moving Average
- **MAIC**: Modified Akaike Information Criterion
- **OLS**: Ordinary Least Squares
- **PACF**: Partial AutoCorrelation Function
- **PP**: Phillips–Perron (test)
- **RMSE**: Root Mean Squared Error
- **SARIMA**: Seasonal ARIMA (model)
- **SE**: Standard Error
- **TBATS**: Trigonometric seasonality, Box–Cox, ARMA errors, Trend, Seasonal components (model)
- **TS**: Trend-Stationary (series)
- **US**: United States
- **VAR**: Vector AutoRegression
- **ZA**: Zivot–Andrews (test)

Today's question and route

- **Question**: when a series drifts upwards, is the shock of today temporary or permanent, and how should we forecast it?
  - the answer decides whether we detrend or difference, and how wide our forecast intervals are
- **Route** of the chapter
  - deterministic and stochastic trends; integrated processes $I(d)$; spurious regression
  - unit-root tests: Dickey–Fuller, ADF, Phillips–Perron; the KPSS stationarity test; structural breaks
  - ARIMA(p, d, q) models: identification, estimation, diagnostics, forecasts with growing intervals
- We build on Chapter 1 (stationarity, random walk, differencing) and Chapter 2 (ARMA models); Seminar 3 comes before this lecture

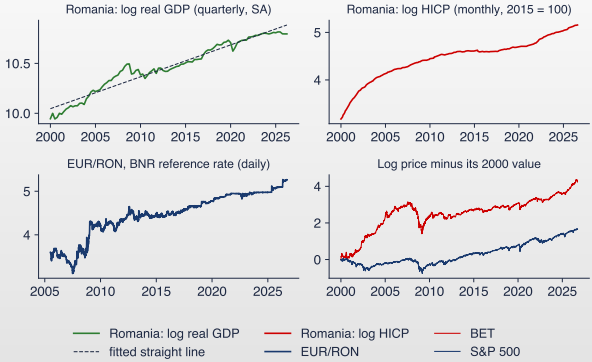
Learning outcomes

- Distinguish trend-stationary from difference-stationary series, and explain why the distinction matters
- Recognise a spurious regression and explain why it happens
- Run and interpret the ADF, Phillips–Perron and KPSS tests, with the right deterministic terms and lags
- Explain how a structural break can mimic a unit root
- Identify, estimate, check and forecast an ARIMA(p, d, q) model, and explain why its intervals grow with the horizon
- Use automatic ARIMA selection critically

Reading and tools

- Textbook: Huang and Petukhina (2022), Ch. 4–5 (ARIMA models and non-stationarity)
  - companion, free online: Hyndman and Athanasopoulos (2021), Ch. 9 (ARIMA models)
- Theory: Hamilton (1994), Ch. 15–17 (trends, unit roots); Box, Jenkins and Reinsel (2008) for the ARIMA methodology
- Python Quantlets of this chapter: Quantlets/Ch_03
  - `adfuller`, `kpss`, `zivot_andrews` and `ARIMA` from `statsmodels`; the Phillips–Perron test written out in a few lines
- Lecture notebook: open in Google Colab
- Video course: Applied Time Series Analysis with Python

Four series that drift



- Romanian real GDP (Eurostat, seasonally adjusted) and HICP (Eurostat); EUR/RON, the BNR reference rate; BET and S&P 500 log prices, daily since 2000
- None of them has a constant mean: Chapter 1 called them non-stationary

TSA_ch3_trends

Interpreting the four series

- GDP grows by 0.80% per quarter on average (about 3.2% per year), but it strays from the straight line by up to 18%
  - the 2009 recession: did GDP come back to the old line, or did it start a new one?
- Consumer prices: $\times 7.2$ since 2000; EUR/RON: from 3.60 to 5.26 lei; BET: $+426$ log points, S&P 500: $+166$ log points (to 18 September 2026)
- Two models can produce such pictures
  - a fixed line plus stationary noise (**deterministic trend**)
  - a random walk with drift (**stochastic trend**)

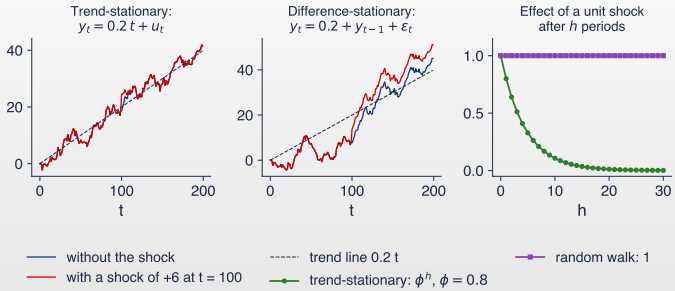
Trend-stationary series

- **Definition**: $y_t = \alpha + \beta t + u_t$, with u_t a stationary process (for example a stationary AR(1), Chapter 2)
  - α: the intercept; β: the slope, the average change per period; t: the time index
  - y_t is **trend-stationary** (TS): stationary after subtracting the line $\alpha + \beta t$
- Properties
  - $E[y_t] = \alpha + \beta t$; $\text{Var}(y_t) = \text{Var}(u_t)$, constant
  - a shock moves y_t away from the line only temporarily: with $u_t = \phi u_{t-1} + \varepsilon_t$, its effect after h periods is ϕ^h
  - long-run forecasts return to the line; their intervals have a bounded width
- Right treatment: estimate the trend by OLS (ordinary least squares) and model the residuals as ARMA

Difference-stationary series

- **Definition**: $y_t = \beta + y_{t-1} + u_t$, u_t stationary; by substitution, $y_t = y_0 + \beta t + \sum_{i=1}^{t} u_i$
  - β: the drift, the average change per period; y_0: the starting value
  - y_t is **difference-stationary** (DS): $\Delta y_t = \beta + u_t$ is stationary
  - $\sum_{i \leq t} u_i$ is the **stochastic trend**; βt is the drift
- Properties
  - the same mean $y_0 + \beta t$ as a TS series, but $\text{Var}(y_t)$ grows with t (Chapter 1: $t\sigma^2$ for a random walk)
  - every shock stays in the level forever: there is no line to return to
  - forecast intervals widen without limit as the horizon grows
- Right treatment: difference the series and model Δy_t as ARMA: the ARIMA model of this chapter

Trend-stationary and difference-stationary series: the effect of a shock



- Left and centre: the same shocks drive a TS series (u_t AR(1), $\phi = 0.8$) and a random walk with drift 0.2; the red paths add one shock of $+6$ at $t = 100$
- Right: the effect of a unit shock after h periods

TSA_ch3_trends

Interpreting the two trends

- TS series: the shock dies out; its effect is $0.8^{10} = 0.107$ after 10 periods and 0.012 after 20
  - at $t = 200$ the two red and blue paths coincide
- Random walk: the whole path shifts up by 6 and stays there
  - a recession in a DS economy is a permanent loss of output, not a dip below a trend
- By eye, the two series without the shock look alike: we need a formal test (Sections 3–4)

Integrated processes

- **Definition**: y_t is **integrated of order** d, $y_t \sim I(d)$, if $\Delta^d y_t$ is stationary and $\Delta^{d-1} y_t$ is not
  - $I(0)$: stationary (with an invertible ARMA representation, Chapter 2); $I(1)$: stationary after one difference; $I(2)$: after two
  - $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = y_t - 2y_{t-1} + y_{t-2}$
- Typical orders in economics and finance
  - log prices of stocks, exchange rates, log GDP: $I(1)$; returns and growth rates: $I(0)$
  - price levels in high-inflation periods: possibly $I(2)$, so that inflation itself is $I(1)$
- Rules
  - $I(0) + I(1) = I(1)$: the stochastic trend dominates; a TS series is $I(0)$ after removing the trend, not after differencing

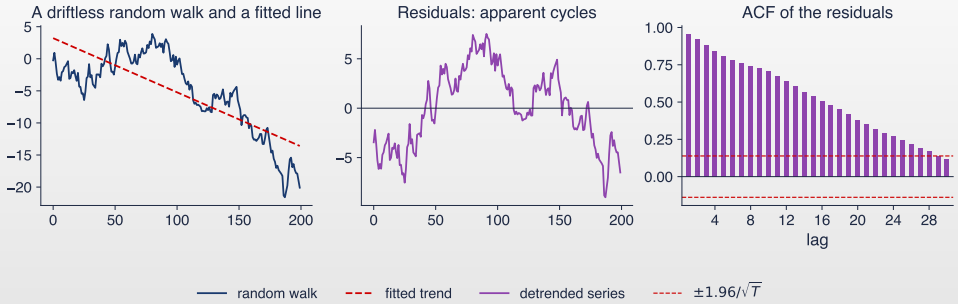
The unit root

- Write an AR(p) with the lag operator L (Chapter 1): $\phi(L)y_t = \varepsilon_t$, $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$
  - Chapter 2: stationary if all roots of $\phi(z) = 0$ lie outside the unit circle, $|z| > 1$
- **Unit root**: one root equals $z = 1$, so $\phi(z) = (1 - z)\,\phi^*(z)$ with the roots of ϕ^* outside the circle
  - then $\phi^*(L)\,\Delta y_t = \varepsilon_t$: Δy_t is a stationary AR($p-1$), and $y_t \sim I(1)$
  - AR(1): $\phi(z) = 1 - \phi z$ has the root $1/\phi$, a unit root when $\phi = 1$: the random walk
- **Example**: $y_t = 1.5y_{t-1} - 0.5y_{t-2} + \varepsilon_t$; $1 - 1.5z + 0.5z^2 = (1 - z)(1 - 0.5z)$
  - roots 1 and 2: $\Delta y_t = 0.5\,\Delta y_{t-1} + \varepsilon_t$, an AR(1) in differences

Case study: Nelson and Plosser (1982)

- **Question**: are US macroeconomic series stationary around a trend, as the business-cycle models of the 1970s assumed?
  - Nelson and Plosser (1982): 14 long annual series (real GNP, employment, prices, wages, interest rates, stock prices), some since 1860
- **Method**: the Dickey–Fuller test of Section 3, with a constant and a trend
  - the unit root was rejected for one series only, the unemployment rate
- **Consequence**: shocks to output look permanent, not transitory
  - a large debate on the nature of recessions; and a warning against detrending by a straight line (next slide)
  - the answer was later challenged by structural breaks (Section 5)

Detrending a random walk



- Left: a driftless random walk ($T = 200$) and the OLS line $\hat{a} + \hat{b}t$; centre: the residuals; right: their ACF

TSA_ch3_trends

Interpreting the detrended random walk

- The line fits: $R^2 = 0.61$, $t = -17.7$ for the slope, although the true slope is 0
  - over 2 000 random walks: $|t| > 1.96$ in 90% of cases, median $R^2 = 0.44$
- The residuals look like long cycles: $\hat{\rho}(1) = 0.96$ and a slow decay
  - Nelson and Kang (1981): detrending a DS series creates **spurious cycles**; a business cycle may be an artefact of the method
- The reverse error (differencing a TS series) is over-differencing (Chapter 1 and Section 6)

Recap: Deterministic and stochastic trends

- TS: $y_t = \alpha + \beta t + u_t$, shocks fade; DS: $\Delta y_t = \beta + u_t$, shocks are permanent
- $I(d)$: stationary after d differences; a unit root in $\phi(z)$ means $I(1)$
- The wrong treatment creates artefacts: spurious cycles (detrending a DS series) or over-differencing (differencing a TS series)

1926: nonsense correlations

- Yule (1926) asked: *Why do we sometimes get nonsense-correlations between time-series?*
  - his example: in England and Wales, 1866–1911, the share of Church of England marriages and the mortality rate had a correlation of about 0.95
  - no causal link: both series simply declined over the period
- His explanation: for series that wander like sums of random terms, the sample correlation does not settle near 0
  - large values occur often by chance; the usual standard error does not apply
- Fifty years later, the same problem reappeared in econometric models estimated on levels

1974: Granger and Newbold

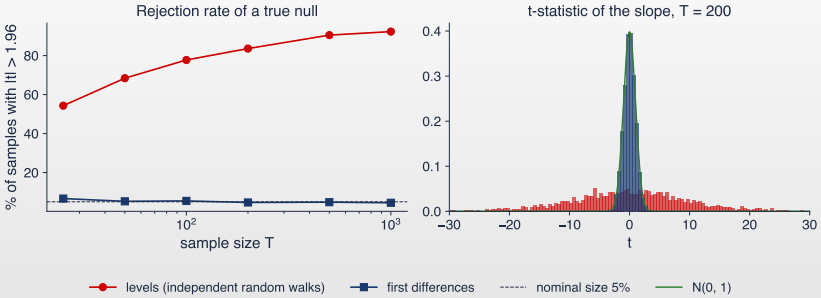


Clive Granger (1934–2009), Nobel Prize in Economics 2003

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

- Granger and Newbold (1974): regress one random walk on another, **independent** random walk
  - $y_t = a + bx_t + e_t$ with y_t, x_t independent $I(1)$ series
  - the t-test of $b = 0$ rejects far too often; R^2 is high; the Durbin–Watson statistic is close to 0
- Their rule of thumb: be suspicious when $R^2 >$ DW
  - DW (Durbin–Watson) $\approx 2(1 - \hat{\rho}_1)$ of the residuals: near 0 means $I(1)$-like residuals
- Granger and Newbold were then at the University of Nottingham; Granger later received the Nobel Prize for cointegration (Chapter 7)

Spurious regression by simulation



- 2 000 pairs of independent Gaussian random walks for each T; OLS of y on a constant and x, in levels and in first differences
- Left: share of samples with $|t| > 1.96$; right: the distribution of t for $T = 200$

TSA_ch3_spurious_regression

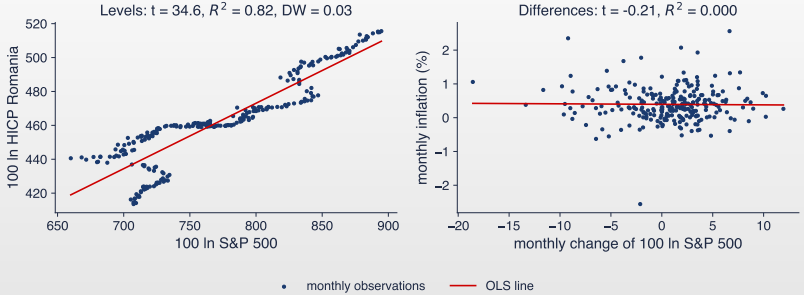
Interpreting the simulation

- Levels: a true null $b = 0$ is rejected in 68% of samples with $T = 50$ and in 92% with $T = 1000$
  - more data makes it worse: the median $|t|$ grows from 3.2 ($T = 50$) to 14.2 ($T = 1000$)
  - median R^2 about 0.17 for every T; median DW 0.08 at $T = 200$
- Differences: 4.7% at $T = 200$, close to the nominal 5%
  - the t-test works again because Δy_t and Δx_t are stationary

Why the t-test fails

- Under $b = 0$ the error $e_t = y_t - a$ is itself a random walk: the OLS assumptions (stationary, uncorrelated errors) fail
  - the usual standard error is far too small
- Phillips (1986) derived the large-sample behaviour
  - $\hat{b}$ does not converge to 0: it converges to a random variable
  - $t_{\hat{b}}$ grows like $\sqrt{T}$, so the rejection rate tends to 100%
  - R^2 has a non-degenerate limit distribution; DW $\to 0$
- Exception: if $y_t - bx_t$ is stationary for some b, the series are **cointegrated** and the levels regression is meaningful Engle and Granger (1987) (Chapter 7)

A spurious regression on real data



- Romanian HICP (Eurostat) against the S&P 500 (month-end close), 260 months since January 2005, both as 100 ln
- Left: levels; right: monthly changes

TSA_ch3_spurious_regression

Interpreting the HICP regression

- Levels: slope 0.39, $t = 34.6$, $R^2 = 0.82$, DW $= 0.03$
  - "US stocks explain Romanian consumer prices": $R^2 >$ DW, a textbook spurious regression
- Differences: $t = -0.21$ (p $= 0.83$), $R^2 = 0.0002$
  - monthly inflation and monthly stock returns are unrelated
- Both levels contain a trend; the regression only says that both went up

Recap: Spurious regression

- Regressing one $I(1)$ series on an unrelated $I(1)$ series gives large t, high R^2 and DW near 0
- More observations do not help: t grows like $\sqrt{T}$
- Remedies: test the order of integration first; regress differences; or test for cointegration (Chapter 7)

Testing $\phi = 1$

- AR(1): $y_t = \phi y_{t-1} + \varepsilon_t$; subtract y_{t-1}: $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$, with $\gamma = \phi - 1$
  - H_0: $\gamma = 0$ (unit root, $I(1)$); H_1: $\gamma < 0$ (stationary); a one-sided, left-tail test
  - statistic: $\tau = \hat{\gamma}/\mathrm{SE}(\hat{\gamma})$, the usual OLS t-ratio; SE: the standard error; a very negative τ speaks against the unit root
- Under H_0, τ is **not** Student or Normal Dickey and Fuller (1979)
  - y_{t-1} is a random walk, so the regressor is not stationary; $\hat{\phi}$ converges at rate T, not $\sqrt{T}$ (**superconsistency**)
  - the limit of τ is a functional of Brownian motion, skewed to the left Hamilton (1994), Ch. 17
  - critical values come from simulation: Fuller (1996); today from response surfaces MacKinnon (2010)

Dickey and Fuller at Iowa State University



Snedecor Hall, Iowa State University: the department of statistics and its Statistical Laboratory (1933)

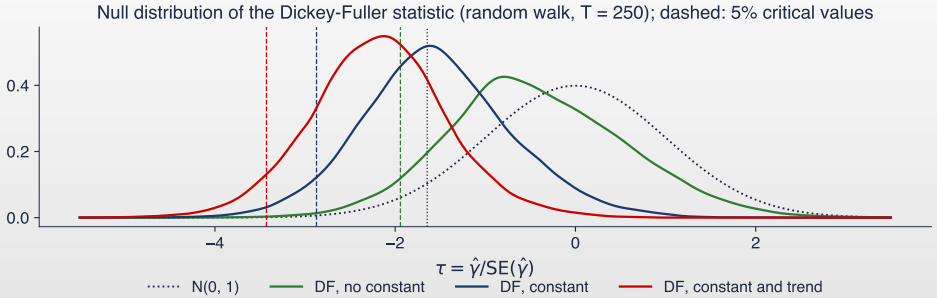
Photo: Maitra (2023); CC BY-SA 4.0; Wikimedia Commons

- **Wayne A. Fuller**: statistician at Iowa State, author of *Introduction to Statistical Time Series* Fuller (1996)
- **David A. Dickey**: his doctoral student at Iowa State (1976), later at North Carolina State University
- Dickey and Fuller (1979): the distribution of $\hat{\phi}$ and τ under a unit root; Dickey and Fuller (1981): joint tests of the unit root and the deterministic terms
- Said and Dickey (1984): the augmented test for ARMA errors (this section)

Three Dickey–Fuller regressions

- **No constant**: $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$
  - H_1: stationary with mean 0; rarely realistic
- **Constant**: $\Delta y_t = c + \gamma y_{t-1} + \varepsilon_t$
  - H_0: random walk (with drift if $c \neq 0$); H_1: stationary around a non-zero mean
  - for series without a trend: interest rates, inflation, exchange rates, returns
- **Constant and trend**: $\Delta y_t = c + bt + \gamma y_{t-1} + \varepsilon_t$
  - c: the constant; bt: a linear trend with slope b
  - H_0: random walk with drift; H_1: trend-stationary: exactly the TS against DS question of Section 1
  - for trending series: log GDP, log prices, log price indices
- Each case has its own distribution of τ and its own critical values

The Dickey–Fuller distributions



- 20 000 random walks with $T = 250$: the density of τ in the three regressions, against $N(0, 1)$; dashed lines: 5% critical values MacKinnon (2010)

TSA_ch3_dickey_fuller

Interpreting the three distributions

- All three lie to the left of $N(0,1)$; the more deterministic terms, the further left
  - simulated 5% quantiles: -1.95 (no constant), -2.86 (constant) and -3.42 (constant and trend); tabulated: -1.94, -2.87 and -3.43
- With the Normal critical value -1.645, a true unit root is "rejected" in 46% of samples (constant) and 77% (constant and trend)
  - the wrong table turns random walks into "stationary" series

Critical values of the Dickey–Fuller test

	no constant			constant			constant and trend		
T	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	−2.61	−1.95	−1.61	−3.57	−2.92	−2.60	−4.15	−3.50	−3.18
100	−2.59	−1.94	−1.61	−3.50	−2.89	−2.58	−4.05	−3.46	−3.15
250	−2.57	−1.94	−1.62	−3.46	−2.87	−2.57	−4.00	−3.43	−3.14
∞	−2.57	−1.94	−1.62	−3.43	−2.86	−2.57	−3.96	−3.41	−3.13

- Source: the response surfaces of MacKinnon (2010), as used by `statsmodels`; p-values from MacKinnon (1994)
- Decision: reject the unit root at 5% if τ is **below** (more negative than) the 5% value
- The values depend only weakly on T; they apply unchanged to the ADF test

Worked example: Romanian real GDP

- $y_t = 100 \ln$ real GDP, 2000Q1–2026Q2; regression with constant and trend, no lags ($T = 105$ differences)
  - OLS: $\hat{\gamma} = -0.0961$, SE($\hat{\gamma}$) $= 0.0416$, so $\hat{\phi} = 1 + \hat{\gamma} = 0.904$
- $\tau = -0.0961/0.0416 = -2.31$; 5% critical value with constant and trend: -3.45
  - $-2.31 > -3.45$: do not reject the unit root (p = 0.43)
  - with the Normal value -1.645 we would have rejected it: the wrong table gives the wrong answer
- Reading: "not rejected" is not "proved"
  - if $\phi = 0.904$ were true, a shock would halve in $\ln 0.5 / \ln 0.904 \approx 6.9$ quarters; with 106 observations the test can hardly tell this from $\phi = 1$

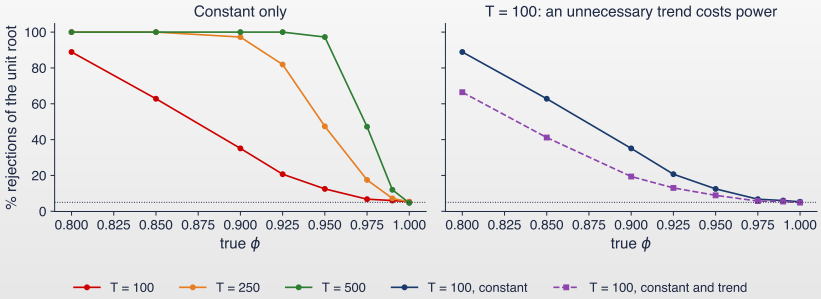
The augmented Dickey–Fuller test

- If Δy_t is autocorrelated, ε_t in the DF regression is not white noise and the size of the test is wrong
  - the remedy: add lagged differences
- **ADF** (augmented Dickey–Fuller) regression:
  - $\Delta y_t = c + bt + \gamma y_{t-1} + \sum_{j=1}^{k} \delta_j \Delta y_{t-j} + \varepsilon_t$; same τ, same critical values
  - δ_j: the coefficients of the past differences; k: the number of lagged differences
  - an AR(p) in levels becomes exactly this regression with $k = p - 1$ and $\gamma = -\phi(1)$
- Said and Dickey (1984): with ARMA errors the test stays valid if k grows slowly with T
  - the lags approximate the MA part by a long AR
- The ADF test is the default unit-root test in software: `adfuller` in `statsmodels`

Choosing the number of lags k

- Maximum: $k_{\max} = 12\,(T/100)^{1/4}$ Schwert (1989); for $T = 105$, $k_{\max} = 13$
  - then choose $k \le k_{\max}$ by an information criterion, AIC or BIC (Chapter 2), on the same sample for every k
- Too few lags: autocorrelated errors, wrong size (with a negative MA part, far too many rejections)
  - too many lags: lower power
- Ng and Perron (2001): a modified AIC (MAIC) gives a better size; GLS detrending Elliott, Rothenberg and Stock (1996) gives more power
  - MAIC: modified Akaike information criterion; GLS: generalised least squares; both appear in the DF-GLS test
- Check the result: Ljung–Box on the ADF residuals Ljung and Box (1978); report k together with τ

Size and power of the Dickey–Fuller test



- 2 000 AR(1) paths $y_t = \phi y_{t-1} + \varepsilon_t$ for each ϕ; share of rejections at 5%; at $\phi = 1$ this is the size

TSA_ch3_dickey_fuller

Interpreting the power curves

- Size: 5.30%, 5.20% and 4.65% at $\phi = 1$ (with a constant): the test holds its 5% level
- Power at $\phi = 0.95$: 12% with $T = 100$, 47% with $T = 250$, 97% with $T = 500$
  - a quarterly series of 25 years ($T = 100$) almost never distinguishes $\phi = 0.95$ from a unit root
  - what matters is the time span, not the frequency: daily data over two years do not help
- An unnecessary trend costs power: at $T = 100$ and $\phi = 0.9$, 35% with a constant, 19% with constant and trend

Choosing the deterministic terms

- The alternative must be a plausible description of the data
  - trending series (log GDP, log prices): constant and trend
  - series that fluctuate around a level (interest rates, inflation, exchange rates, returns): constant
- Omitting a needed trend: a TS series looks like a unit root; the test almost never rejects
  - adding an unneeded trend: lower power (previous chart)
- Look at the plot first, then choose; Dickey and Fuller (1981) give joint F-type tests of the unit root and the trend

How many differences: $I(1)$ or $I(2)$?

- Dickey and Pantula (1987): test from the highest plausible order downwards
  - step 1: ADF on Δy_t; if it does not reject, y_t may be $I(2)$
  - step 2: if it rejects, ADF on y_t; reject: $I(0)$; do not reject: $I(1)$
- Why downwards: if $y_t \sim I(2)$, the test on y_t assumes at most one unit root and is not valid
- Example: Romanian consumer prices since 2005 (the table in Section 4)
  - $\ln P$: unit root; inflation π: ADF p $= 0.32$, KPSS 0.404; change of inflation: clearly stationary
  - so $\ln P$ is $I(1)$ or $I(2)$, depending on how persistent inflation is: Section 7 returns to this

Recap: The Dickey–Fuller test

- $\Delta y_t = c + bt + \gamma y_{t-1} + \sum_j \delta_j \Delta y_{t-j} + \varepsilon_t$; H_0: $\gamma = 0$; τ against Dickey–Fuller critical values
- Deterministic terms follow the plot; lags by AIC or BIC up to $12(T/100)^{1/4}$
- Low power near $\phi = 1$: not rejecting is weak evidence for a unit root
- Order of integration: test from the top (Δy_t first)

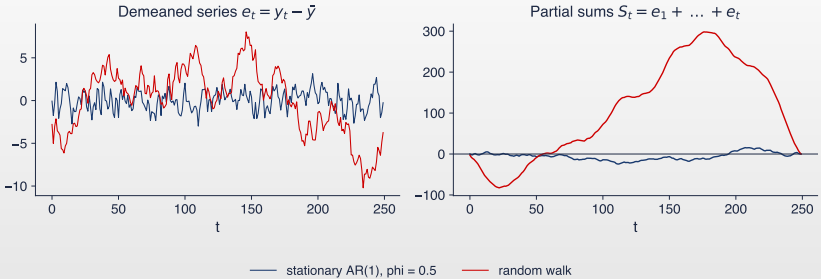
The Phillips–Perron test

- Phillips and Perron (1988): keep the simple DF regression (no lagged differences) and correct τ instead
  - $Z_\tau = \sqrt{\hat{\gamma}_0 / \hat{\lambda}^2}\,\tau - \dfrac{(\hat{\lambda}^2 - \hat{\gamma}_0)\, T\, \mathrm{SE}(\hat{\gamma})}{2\hat{\lambda}\, s}$
  - $\hat{\gamma}_0$: variance of the residuals; s^2: their OLS variance; $\hat{\lambda}^2$: their **long-run variance**
- Long-run variance (Newey–West, Bartlett weights): $\hat{\lambda}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{L}(1 - \frac{j}{L+1})\hat{\gamma}_j$
  - $\hat{\gamma}_j$: autocovariances of the residuals; L: the number of autocovariances included (the bandwidth)
  - if the residuals are white noise, $\hat{\lambda}^2 = \hat{\gamma}_0$ and $Z_\tau = \tau$: the correction only matters when they are autocorrelated
- Same H_0 and the same critical values as ADF; robust to heteroskedasticity; no lag choice, but a bandwidth L
  - weak point: large size distortions with a negative MA component Schwert (1989)

The KPSS test: stationarity as the null

- Kwiatkowski et al. (1992) (KPSS: Kwiatkowski–Phillips–Schmidt–Shin): $y_t = \xi t + r_t + \varepsilon_t$, $r_t = r_{t-1} + u_t$, $u_t \sim \text{WN}(0, \sigma_u^2)$
  - ξt: a deterministic trend; r_t: a random walk, the moving level; ε_t: stationary noise
  - H_0: $\sigma_u^2 = 0$ (the random walk is a constant: y_t stationary around a level or a trend); H_1: $\sigma_u^2 > 0$, a unit root
- Statistic: regress y_t on a constant (and a trend); residuals e_t; partial sums $S_t = e_1 + \cdots + e_t$
  - $\eta = \dfrac{1}{T^2 \hat{\lambda}^2} \sum_{t=1}^{T} S_t^2$, with $\hat{\lambda}^2$ the long-run variance of e_t
  - reject stationarity for **large** η: 5% critical values 0.463 (level) and 0.146 (trend); 1%: 0.739 and 0.216
- A test of stationarity, not of a unit root: it reverses the burden of proof

KPSS: why partial sums



- A stationary AR(1) ($\phi = 0.5$) and a random walk, $T = 250$: the demeaned series and their partial sums S_t
- KPSS (level): 0.16 for the AR(1), 0.73 for the random walk; 5% critical value 0.463

TSA_ch3_unit_root_tests

Interpreting the partial sums

- Stationary series: deviations cancel out, S_t stays close to 0
  - $\sum S_t^2$ grows like T^2, so η stays bounded
- Random walk: long runs on one side of the mean, S_t makes large excursions
  - $\sum S_t^2$ grows like T^4: η grows with T and the test rejects
- The same idea as the CUSUM charts of quality control: cumulated deviations reveal a drifting level

ADF and KPSS together

	KPSS does not reject	**KPSS rejects**
ADF rejects	$I(0)$: both agree	conflict: break, long memory, outliers?
ADF does not reject	inconclusive: the data cannot decide	$I(1)$: both agree

- The two tests have opposite null hypotheses: agreement is stronger evidence than either test alone
- Inconclusive is common with short samples: low power on both sides
  - then decide by the purpose: for forecasting, $d = 1$ is the cautious choice (wider intervals)
- Conflict: look for breaks (Section 5) or fractional integration (Chapter 8)

Unit-root tests on real series

Series	det.	T	ADF (p)	k	PP	KPSS	ADF+KPSS
Romania, real GDP, $100 \ln Y$	c, t	106	-2.31 (0.428)	0	-2.19	0.198	$I(1)$
its growth Δ	c	105	-10.77 (<0.001)	0	-10.88	0.263	$I(0)$
Romania, HICP, $100 \ln P$	c, t	260	-0.96 (0.949)	6	-0.88	0.314	$I(1)$
Romania, 12-month inflation π	c	260	-1.94 (0.315)	15	-2.31	0.404	inconclusive
its change $\Delta\pi$	c	259	-4.83 (<0.001)	14	-11.31	0.105	$I(0)$
US, real GDP, $100 \ln Y$	c, t	318	-1.54 (0.816)	1	-1.20	0.581	$I(1)$
its growth Δ	c	317	-15.51 (<0.001)	0	-15.44	0.456	$I(0)$
EUR/RON, $100 \ln S$	c	5 353	-1.22 (0.664)	7	-1.15	10.255	$I(1)$
daily returns	c	5 352	-27.42 (<0.001)	6	-60.82	0.055	$I(0)$
BET, $100 \ln P$	c, t	6 671	-1.90 (0.653)	19	-2.35	1.143	$I(1)$
daily returns	c	6 670	-16.50 (<0.001)	18	-74.55	0.360	$I(0)$
S&P 500, $100 \ln P$	c, t	6 715	-2.08 (0.557)	34	-2.36	2.281	$I(1)$
daily returns	c	6 714	-14.26 (<0.001)	35	-91.41	0.449	$I(0)$
Nile, yearly flow	c	100	-4.05 (0.001)	1	-6.38	0.869	conflict

- det.: deterministic terms (c: constant; t: trend); k: ADF lags (AIC); 5% critical values: ADF and PP about -2.86 (c), -3.41 (c, t); KPSS 0.463 (c), 0.146 (c, t)

TSA_ch3_unit_root_tests

Interpreting the unit-root table

- Log prices, the exchange rate, log GDP and log HICP: $I(1)$; their differences: $I(0)$
  - Romanian GDP: ADF p $= 0.43$ and KPSS $0.198 > 0.146$: a unit root, as Nelson and Plosser (1982) found for the US
- Romanian 12-month inflation: neither test rejects (KPSS $0.404 < 0.463$): inconclusive
  - a very persistent series, possibly with breaks; Sections 5 and 7
- Nile: both reject: a conflict
  - Chapter 1 showed a level shift in 1898: the next section explains the conflict
- Daily returns: PP gives far more negative values than ADF (BET: -74.6 against -16.5), but both reject; the verdict is the same

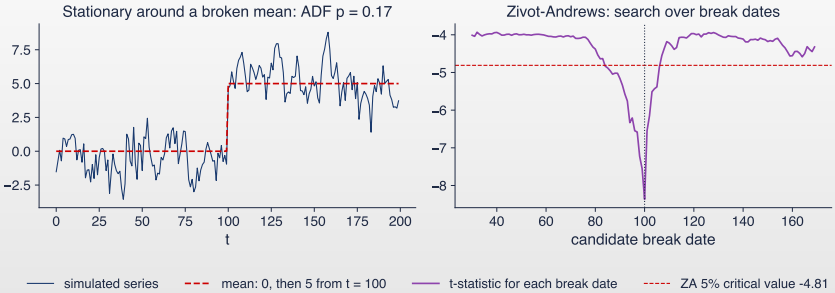
Recap: Phillips–Perron and KPSS

- PP: DF regression plus a long-run-variance correction; same H_0 and critical values as ADF
- KPSS: H_0 stationarity; large partial sums reject it
- Read ADF and KPSS together: agree, inconclusive, or conflict

Perron (1989): a break can look like a unit root

- Perron (1989): a stationary series around a mean or a trend that **shifts once** looks persistent
  - the DF regression explains the shift by $\hat{\phi}$ close to 1; the test loses power
- He re-examined Nelson and Plosser with two known breaks: the 1929 crash (level) and the 1973 oil shock (slope)
  - with dummy variables for the break, the unit root was rejected for most series
  - reading: most shocks are transitory; only a few rare events are permanent
- The critique: the break date was chosen after looking at the data

A level shift misleads the ADF test



- Left: AR(1) noise ($\phi = 0.6$) around a mean that jumps from 0 to 5 at $t = 100$; right: the t-statistic on y_{t-1} for every candidate break date, with a level dummy

TSA_ch3_structural_breaks

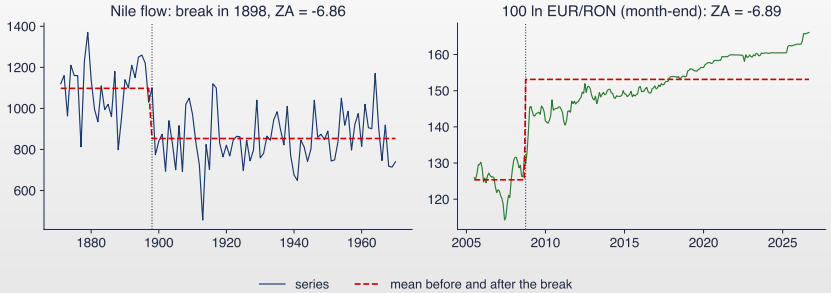
Interpreting the simulated break

- ADF with a constant: p = 0.17: the stationary series "has a unit root"
  - over 2 000 such series, the DF test rejects in 72% of cases; without the break: 100%
- With the break modelled, the Zivot–Andrews statistic falls to -7.47, far below the 5% value -4.81
  - the curve on the right reaches its minimum, -8.37, at the true date $t = 100$

Zivot and Andrews (1992): an unknown break date

- Zivot and Andrews (1992): let the data choose the break date T_B
  - for each T_B in the central 70% of the sample, run the ADF regression with a break dummy:
  - $\Delta y_t = c + \theta DU_t + bt + \gamma y_{t-1} + \sum_j \delta_j \Delta y_{t-j} + \varepsilon_t$, $DU_t = 1$ for $t > T_B$
  - DU_t: a dummy variable, 0 before and 1 after the break; θ: the size of the level shift
  - the statistic is the **minimum** t-ratio of γ over all dates
- Three versions: break in the level, in the slope, in both
  - critical values more negative than ADF: about -4.81 (level) and -5.07 (both) at 5%
- H_0: unit root without a break; H_1: stationary with one break
  - several breaks: Bai and Perron (1998); a rejection does not prove that the break is the right story

Zivot–Andrews on two real series



- Left: the Nile flow (statsmodels data set, 1871–1970); right: 100 ln EUR/RON, month-end BNR reference rate
- Dotted line: the break date chosen by the test (break in the level)

TSA_ch3_structural_breaks

Interpreting the two breaks

- Nile: ZA $= -6.86$, break in 1898, the mean falls from 1098 to 853
  - the date matches the history: the first Aswan dam (built 1898–1902) and a drier climate; the ADF–KPSS conflict is explained
- EUR/RON (month-end): ADF p $= 0.55$, but ZA $= -6.89$, break in October 2008: the 2008 depreciation (mean 3.50 lei before, 4.62 after)
  - a managed exchange rate: long calm periods and rare adjustments, not a pure random walk
  - but after 2008 the rate still drifts up: "stationary around one break" is also a simplification

Recap: Structural breaks

- A one-time shift in the mean or trend biases unit-root tests towards non-rejection
- Zivot–Andrews searches the break date; its critical values are more negative
- Unit root or break: check the dates against known events

The ARIMA model

- **Definition**: $y_t \sim$ ARIMA(p, d, q) (autoregressive integrated moving average) if $\Delta^d y_t$ is a stationary, invertible ARMA(p, q):
  - $\phi(L)\,(1 - L)^d\, y_t = c + \theta(L)\,\varepsilon_t$, $\varepsilon_t \sim \text{WN}(0, \sigma^2)$
  - p: AR order; d: number of differences (unit roots); q: MA order; $\phi(L)$, $\theta(L)$ as in Chapter 2
- The AR polynomial of the levels is $\phi(z)(1 - z)^d$: d roots exactly on the unit circle
  - "integrated": y_t is a d-fold cumulative sum of an ARMA process
- Box, Jenkins and Reinsel (2008) made the model popular; in practice $d \leq 2$, and $d = 1$ most often

Special cases and the constant

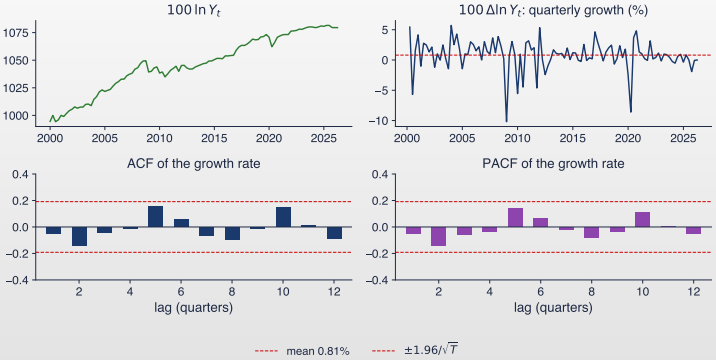
Model	Equation and meaning
ARIMA(0,1,0)	$\Delta y_t = c + \varepsilon_t$: random walk (with drift c)
ARIMA(1,1,0)	$\Delta y_t = c + \phi \Delta y_{t-\mathbf{1}} + \varepsilon_t$: growth with momentum
ARIMA(0,1,1)	$\Delta y_t = \varepsilon_t + \theta \varepsilon_{t-\mathbf{1}}$: simple exponential smoothing with $\alpha = 1 + \theta$ (Chapter 0)
ARIMA(0,2,2)	$\Delta^{\mathbf{2}} y_t = \varepsilon_t + \theta_{\mathbf{1}} \varepsilon_{t-\mathbf{1}} + \theta_{\mathbf{2}} \varepsilon_{t-\mathbf{2}}$: the model behind Holt's linear method

- The constant c changes the long-run forecast Hyndman and Athanasopoulos (2021), Ch. 9.7
  - $d = 0$: forecasts go to the mean; $d = 1$: $c \neq 0$ gives a straight line (drift), $c = 0$ a flat line
  - $d = 2$: $c \neq 0$ gives a quadratic trend, rarely sensible; use $c = 0$

The Box–Jenkins cycle with unit roots

- **1. Transform**: logarithm or Box–Cox if the variance grows with the level (Chapter 1)
- **2. Choose** d: plot, ACF, ADF and KPSS; test Δy_t first; stop at the first stationary difference
- **3. Choose** p, q: ACF and PACF of $\Delta^d y_t$ (Chapter 2), then compare candidates by AICc on the same d
- **4. Estimate** by maximum likelihood; decide on the constant
- **5. Check** the residuals: ACF, Ljung–Box with $m - p - q$ degrees of freedom, outliers
- **6. Forecast** the levels with intervals; go back to step 2 or 3 if the checks fail

Romanian GDP: identification



- Real GDP, seasonally adjusted (Eurostat), 2000Q1–2026Q2, $T = 106$; top: 100 ln Y_t and the quarterly growth; bottom: ACF and PACF of the growth

TSA_ch3_arima_gdp

Interpreting the identification plots

- $d = 1$: the level is $I(1)$ (Section 4); the growth rate is stationary, mean 0.81% per quarter (about 3.2% per year)
  - two large falls: -10.2% in 2009Q1 (global financial crisis) and -8.6% in 2020Q2 (pandemic)
- p and q: no ACF or PACF value outside ± 0.19; $\hat{\rho}(1) = -0.05$, $\hat{\rho}(2) = -0.14$
  - Ljung–Box $Q^*(8) = 7.3$ (p = 0.51): the growth rate looks like white noise around its mean
- The plots suggest ARIMA(0,1,0) with drift; the next slide checks the alternatives

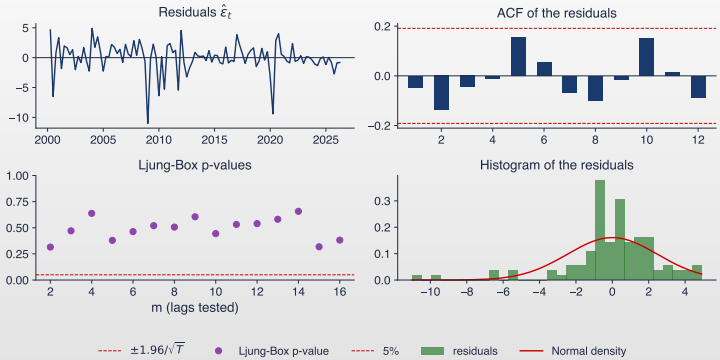
Model choice by AICc

	$q = 0$	$q = 1$	$q = 2$
$p = 0$	491.9	493.7	493.6
$p = 1$	493.8	494.9	495.8
$p = 2$	493.7	495.7	495.4†

- ARIMA$(p, 1, q)$ with drift for $100 \ln Y_t$, estimated by maximum likelihood; AICc: AIC with the small-sample correction (Chapter 2); †: the optimiser did not converge
- Lowest AICc: ARIMA(0,1,0); the next model is 1.7 points worse
  - differences below 2 points are not meaningful: keep the simplest model
- Estimate: drift $\hat{c} = 0.81$ (SE 0.30, $t = 2.72$), $\hat{\sigma} = 2.47$: $100\,\Delta \ln Y_t = 0.81 + \hat{\varepsilon}_t$

TSA_ch3_arima_gdp

Romanian GDP: residual checks



- Residuals of ARIMA(0,1,0) with drift: time plot, ACF, Ljung–Box p-values for $m = 2, \ldots, 16$, histogram with a Normal density

TSA_ch3_arima_gdp

Interpreting the residual checks

- No autocorrelation left: $Q^*(8) = 7.3$ (p $= 0.51$), $Q^*(12)$ p $= 0.54$; all p-values above 5%
- Not Normal: kurtosis 7.9, Jarque–Bera 145; the largest residual is -11.0 in 2009Q1
  - two crises dominate the tails; the Normal intervals of the next slides are too narrow in a crisis and slightly too wide in calm times
  - options: dummy variables for 2009 and 2020, or a bootstrap of the residuals for the intervals

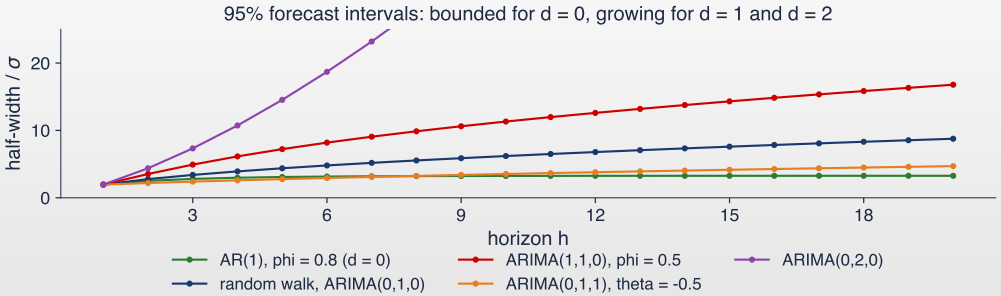
Forecasting with ARIMA

- **Point forecasts**: write the model for the levels and iterate it, with $\hat{\varepsilon}_{T+j} = 0$ for $j \geq 1$
  - ARIMA(1,1,0): $y_t = (1+\phi)y_{t-1} - \phi y_{t-2} + \varepsilon_t$
- **Forecast error**: $y_{T+h} - \hat{y}_{T+h} = \sum_{j=0}^{h-1} \psi_j \varepsilon_{T+h-j}$, with $\psi(z) = \theta(z)/[\phi(z)(1-z)^d]$
  - variance $\sigma^2 \sum_{j=0}^{h-1} \psi_j^2$; 95% interval $\hat{y}_{T+h} \pm 1.96\,\sigma\sqrt{\sum_{j<h} \psi_j^2}$
  - for $d \geq 1$ the ψ_j do not decay to 0: the variance grows without bound
- Two closed forms
  - random walk: $\psi_j = 1$, variance $h\sigma^2$: the interval grows like $\sqrt{h}$
  - ARIMA(0,1,1): $\psi_j = 1 + \theta$ for $j \geq 1$, variance $\sigma^2[1 + (h-1)(1+\theta)^2]$

Worked example: ARIMA(1,1,0) forecasts

- $\Delta y_t = 0.6\,\Delta y_{t-1} + \varepsilon_t$, $\sigma = 2$; last data: $y_T = 108$, $\Delta y_T = 5$
- **Point forecasts**: forecast the change, then add it
  - $\widehat{\Delta y}_{T+1} = 0.6 \cdot 5 = 3$, $\hat{y}_{T+1} = 111.00$; $\widehat{\Delta y}_{T+2} = 1.8$, $\hat{y}_{T+2} = 112.80$; $\hat{y}_{T+3} = 113.88$
- ψ **weights**: $\psi_0 = 1$, $\psi_1 = 1 + \phi = 1.60$, $\psi_2 = 1 + \phi + \phi^2 = 1.96$
  - variances: 4, $4(1 + 1.60^2) = 14.24$, $4(1 + 1.60^2 + 1.96^2) = 29.61$
  - 95% half-widths: 3.92, 7.40, 10.66: the interval almost triples in three steps
- The momentum $\phi = 0.6$ makes the uncertainty grow faster than for a random walk ($\sqrt{3} = 1.73$ times in three steps)

How fast the intervals grow



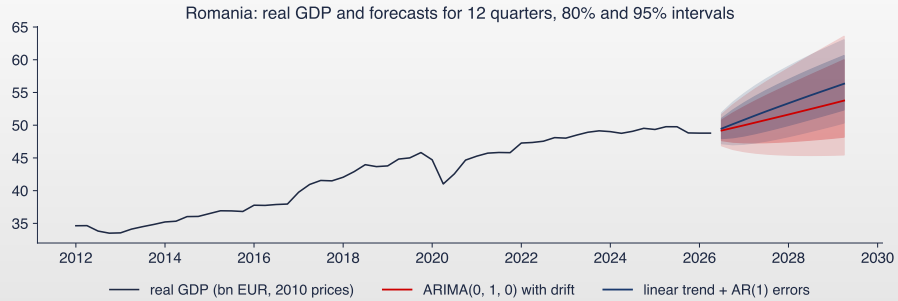
⊡ Half-width of the 95% forecast interval, in units of σ, for horizons 1 to 20, from the ψ weights of five models

TSA_ch3_forecast_intervals

Interpreting the interval widths

- $d = 0$, AR(1) with $\phi = 0.8$: the width levels off at $1.96\sigma/\sqrt{1-\phi^2}$ (3.27σ at $h = 20$)
- $d = 1$: growth like $\sqrt{h}$; at $h = 20$: 8.77σ (random walk), 16.78σ (ARIMA(1,1,0), $\phi = 0.5$), 4.70σ (ARIMA(0,1,1), $\theta = -0.5$)
  - a negative MA part offsets part of each shock: narrower intervals
- $d = 2$: growth like $h^{3/2}$ (105.00σ at $h = 20$): over-differencing makes long-run forecasts useless

Romanian GDP: two forecasts



- ARIMA(0,1,0) with drift (red) and a linear trend with AR(1) errors (blue), both estimated on 2000Q1–2026Q2; forecasts to 2029Q2, transformed back to bn EUR (2010 prices)

TSA_ch3_arima_gdp

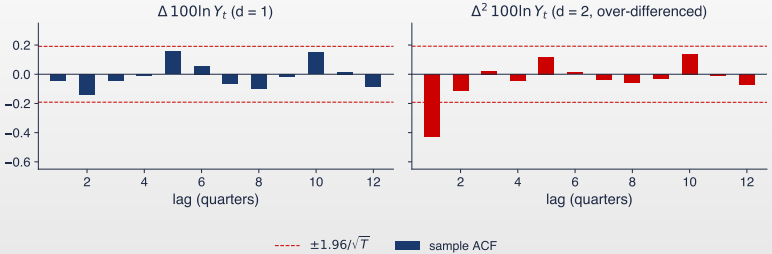
Interpreting the GDP forecasts

- The random walk with drift grows by 0.81% per quarter from the last value (48.8 bn EUR, 2010 prices, in 2026Q2): 53.8 bn after 12 quarters
  - 95% half-width (in % of GDP): 4.8 after 1 quarter, 9.7 after 4, 16.8 after 12, i.e. $\times 3.46 = \sqrt{12}$
- The TS model pulls GDP back towards its old line ($\hat{\phi} = 0.92$): 56.3 bn after 12 quarters, half-width 11.2%
  - more optimistic and more confident, because it assumes the recent slowdown is temporary
- The tests favour DS at 5% (ADF p = 0.428; KPSS 0.198 $>$ 0.146), but are inconclusive at 1% (KPSS $<$ 0.216): the choice of d is a forecasting decision with consequences

Over-differencing

- Differencing an $I(0)$ series creates an MA unit root: $\Delta\varepsilon_t = \varepsilon_t - \varepsilon_{t-1}$, $\theta = -1$ (Chapter 1)
  - the result is not invertible: no AR(∞) representation, poor estimation Plosser and Schwert (1977)
- Symptoms
  - $\hat{\rho}(1)$ of the differenced series close to -0.5
  - the variance increases after differencing instead of decreasing
  - an estimated MA coefficient close to -1
- Rule: difference only as long as the variance goes down and the tests ask for it

Over-differencing Romanian GDP



- ACF of the quarterly growth $\Delta\, 100 \ln Y_t$ $(d = 1)$ and of its difference $\Delta^2\, 100 \ln Y_t$ $(d = 2)$

TSA_ch3_arima_gdp

Interpreting over-differencing

- $d = 1$: $\hat{\rho}(1) = -0.05$, variance 6.16; $d = 2$: $\hat{\rho}(1) = -0.43$, variance 12.80: all three symptoms
- An MA(1) fitted to $\Delta^2 100 \ln Y_t$ gives $\hat{\theta} = -0.987$ (SE 0.080): practically -1, the extra difference is cancelled by the MA part
- Correct choice: $d = 1$

Recap: ARIMA models

- ARIMA(p, d, q): an ARMA(p, q) for $\Delta^d y_t$; the constant decides the long-run slope
- Romanian GDP since 2000: ARIMA(0,1,0) with drift, 0.81% per quarter; residuals uncorrelated but heavy-tailed
- Forecast variance $\sigma^2 \sum_{j<h} \psi_j^2$: bounded for $d = 0$, growing for $d \geq 1$
- Over-differencing: $\hat{\rho}(1) \approx -0.5$, larger variance, MA coefficient near -1

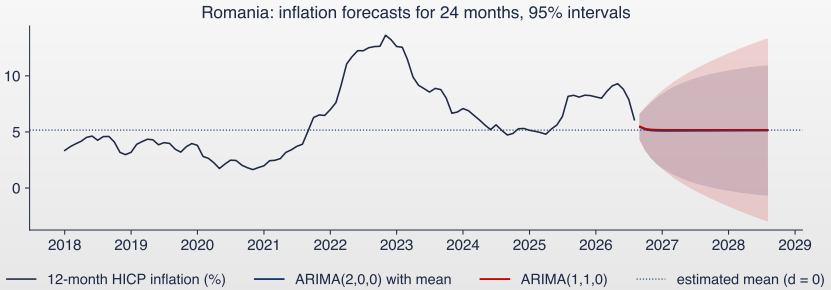
Automatic ARIMA

- Hyndman and Khandakar (2008) (`auto.arima` in R; `pmdarima` and `statsforecast` in Python):
  - d: repeated KPSS tests at 5%: difference while KPSS rejects
  - p, q: a stepwise search from a few starting models, by AICc, on the chosen d; a constant or drift if $d \leq 1$
  - rejects models with roots too close to the unit circle and fits that fail
- Useful for many series at once and as a starting point
  - evaluated by its authors on the 3 003 series of the M3 forecasting competition, it was competitive with the best methods there
- But each step embeds a choice that you must be able to defend (next slide)

Automatic ARIMA: checks

- **AICc is not comparable across** d: with $d = 1$ the likelihood is computed for Δy_t, a different series
  - choose d by tests and judgement, then compare p, q
- **Near-cancelling roots**: Romanian GDP from 1995Q1: the lowest AICc, 607.5 (ARIMA(0,1,0): 611.1), belongs to ARIMA(2,1,2), whose fit did not converge
  - MA roots of modulus 1.0002, AR roots 1.014: AR and MA almost cancel; $\hat{\theta}_2 = 1.000$ with SE 6.96
  - a model fitted to the noise of the 1990s data, not a description of the economy
- **Always**: plot the data, check the convergence and the roots, run the residual tests, compare with a simple benchmark (random walk, Chapter 0)

Romanian inflation: $d = 0$ or $d = 1$?



- 12-month HICP inflation (Eurostat), January 2005–August 2026; best model by AICc for $d = 0$ (with a mean) and for $d = 1$, $p \leq 3$, $q \leq 2$; forecasts for 24 months

TSA_ch3_arima_in_practice

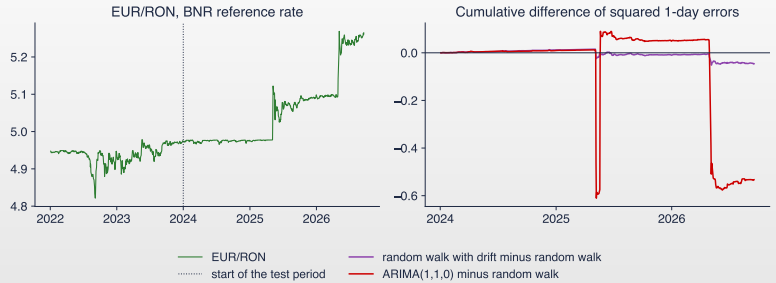
Interpreting the two inflation models

- $d = 0$: ARIMA(2,0,0), $\hat{\phi}_1 = 1.32$, $\hat{\phi}_2 = -0.34$, sum 0.976: stationary, but barely; mean 5.2%
  - $d = 1$: ARIMA(1,1,0), $\hat{\phi} = 0.33$: almost the same short-run dynamics
- After 24 months: 5.1% in $[-0.6, 10.9]$ against 5.2% in $[-2.9, 13.3]$
  - similar points, different intervals: d decides how much uncertainty we report
- KPSS (0.404) keeps $d = 0$, as automatic ARIMA would; ADF ($p = 0.32$) suggests $d = 1$
  - both models leave autocorrelated residuals: 12-month rates overlap and inherit the season; Chapter 4 models monthly inflation

Case study: Meese and Rogoff (1983)

- **Question**: can exchange-rate models forecast better than a random walk?
  - Meese and Rogoff (1983): monetary models of the 1970s, time series models (ARIMA, VAR) and the random walk; dollar rates, horizons of 1 to 12 months
- **Finding**: out of sample, none beat the random walk in RMSE (root mean squared error)
  - even with the actual future values of the fundamentals plugged in
- **Legacy**: the "Meese–Rogoff puzzle" and the random walk as the benchmark for every exchange-rate forecast
  - next chart: the same comparison for EUR/RON

EUR/RON: ARIMA against the random walk



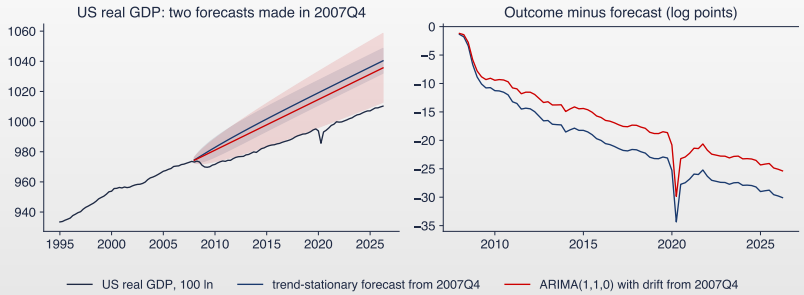
- Parameters estimated on 4 673 daily changes until December 2023; 679 one-day-ahead forecasts from January 2024; right: the cumulated difference of squared errors (below 0: the model beats the random walk)

TSA_ch3_arima_in_practice

Interpreting the EUR/RON comparison

- RMSE of the daily change (100 ln): random walk 0.1330, with drift 0.1327, ARIMA(1,1,0) 0.1300 ($\hat{\phi} = 0.169$)
  - a gain of 2.2%; over 20 days (33 non-overlapping windows): 0.667 against 0.638, a gain of 4.3%
- The gain comes from a few days: two jumps of the rate (from 4.9746 to 5.2644 lei over the period) continued on the next day
  - on calm days the random walk is as good; a formal comparison needs the Diebold–Mariano test (Chapter 4)
- Lesson of Meese and Rogoff: report the random walk next to any exchange-rate forecast

A classic series: US real GDP after 2007



- FRED series GDPC1 since 1947; a TS model (linear trend, AR(2) errors) and a DS model (ARIMA(1,1,0) with drift) estimated up to 2007Q4 and forecast to 2026Q2; 95% intervals

TSA_ch3_us_gdp

Interpreting the US GDP forecasts

- The TS model expected a return to the pre-2008 line; the gap kept widening: -30 log points at the end (half-width of its interval: 8)
  - the 2008–2009 loss was never recovered: evidence for a permanent shock, as the DS view of Nelson and Plosser (1982) implies
- The DS model also misses (-25), but its interval (half-width 23) at least warns of large uncertainty
  - both assume the same drift as before 2008; growth after 2008 was slower: a break in the drift (Section 5)
- Unit-root tests on the full series: ADF p $= 0.82$, KPSS 0.58: $I(1)$

Recap: ARIMA in practice

- Automatic ARIMA: KPSS for d, AICc for p, q; check convergence, roots and residuals
- AICc compares models with the same d only
- Exchange rates: the random walk is a hard benchmark Meese and Rogoff (1983)
- TS or DS changes the long-run forecast and its interval: US GDP after 2008

Possible contribution of AI

- **Code**: a first draft of a script that runs ADF, PP and KPSS on many series and tabulates the verdicts
- **Explanation**: a second explanation of why the Dickey–Fuller distribution is not Normal, or of a ZA output
- **Exploration**: unit roots and breaks in all EU inflation rates, or ARIMA benchmarks for many Romanian series
- Example prompt
  - `Write Python code that loads Romanian quarterly real GDP (Eurostat namq_10_gdp, seasonally adjusted), runs ADF with constant and trend and KPSS on 100*log GDP and on its difference, fits ARIMA(p,1,q) with drift for p, q <= 2, and plots 8-quarter forecasts with 95% intervals.`

Checks you must run

- The direction of each test: ADF and PP have the unit root as H_0, KPSS has stationarity as H_0
- The deterministic terms, the number of lags and the critical values actually used (Dickey–Fuller, not Normal)
- That "not rejected" is not reported as "proved"
- That regressions between trending series are not presented as causal evidence
- Convergence warnings, roots near the unit circle, residual tests, and AICc compared only for the same d
- Every cited reference: it must exist; check the DOI

Key takeaways

- Trending series are TS (shocks fade) or DS (shocks are permanent); the treatment differs: detrend or difference
- Regressions between unrelated $I(1)$ series are spurious: high t and R^2, DW near 0
- ADF and PP test H_0: unit root with Dickey–Fuller critical values; KPSS tests H_0: stationarity; use them together
- The tests have low power near $\phi = 1$ and are misled by breaks; Zivot–Andrews allows one break
- ARIMA(p, d, q) = ARMA on $\Delta^d y_t$; its forecast intervals grow with the horizon when $d \geq 1$
- Automatic selection is a starting point: check d, convergence, roots, residuals and a random-walk benchmark

Key formulas

Quantity	Formula
TS and DS	$y_t = \alpha + \beta t + u_t, \quad \Delta y_t = \beta + u_t, \quad u_t \sim I(0)$
ADF	$\Delta y_t = c + bt + \gamma y_{t-1} + \sum_{j=1}^{k} \delta_j \Delta y_{t-j} + \varepsilon_t, \quad H_0$: $\gamma = 0, \quad \tau = \hat{\gamma}/\mathrm{SE}(\hat{\gamma})$
Lags	$k_{\max} = 12\,(T/100)^{1/4}$, then AIC or BIC
KPSS	$\eta = \sum_{t=1}^{T} S_t^2/(T^2\hat{\lambda}^2), \quad S_t = \sum_{s \le t} e_s, \quad$ 5%: 0.463 (c), 0.146 (c, t)
ARIMA(p, d, q)	$\phi(L)(1-L)^d y_t = c + \theta(L)\varepsilon_t$
Forecast variance	$\sigma^2 \sum_{j=0}^{h-1} \psi_j^2, \quad \psi(z) = \theta(z)/[\phi(z)(1-z)^d]$; random walk: $h\sigma^2$
ARIMA(0,1,1)	$\sigma^2[1 + (h-1)(1+\theta)^2], \quad \alpha_{\mathrm{SES}} = 1 + \theta$

Self-assessment

- **Question**: an ADF test with constant and trend gives $\tau = -2.9$ with $T = 200$. Do you reject the unit root at 5%?
  - **Answer**: no: $-2.9 > -3.43$; with the Normal table you would wrongly reject
- **Question**: ADF does not reject and KPSS does not reject. What do you conclude?
  - **Answer**: inconclusive: the sample is too short to decide; report both models, or choose d by the purpose
- **Question**: for ARIMA(0,1,1) with $\theta = -0.6$ and $\sigma = 1$, what is the variance of the 5-step forecast error?
  - **Answer**: $1 + 4 \cdot 0.4^2 = 1.64$
- Next: Chapter 4, seasonality: seasonal differences, SARIMA, TBATS and Prophet

References (1/3)

- Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1), 47–78. doi:10.2307/2998540
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control* (4th ed.). Wiley. doi:10.1002/9781118619193
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (3rd ed.). Springer. doi:10.1007/978-3-319-29854-2
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427–431. doi:10.1080/01621459.1979.10482531
- Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4), 1057–1072. doi:10.2307/1912517
- Dickey, D. A., & Pantula, S. G. (1987). Determining the order of differencing in autoregressive processes. *Journal of Business & Economic Statistics*, 5(4), 455–461. doi:10.1080/07350015.1987.10509614
- Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4), 813–836. doi:10.2307/2171846
- Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2), 251–276. doi:10.2307/1913236
- Fuller, W. A. (1996). *Introduction to Statistical Time Series* (2nd ed.). Wiley. doi:10.1002/9780470316917
- Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2), 111–120. doi:10.1016/0304-4076(74)90034-7

References (2/3)

- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press. doi:10.2307/j.ctv14jx6sm
- 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
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1–22. doi:10.18637/jss.v027.i03
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1–3), 159–178. doi:10.1016/0304-4076(92)90104-Y
- 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
- MacKinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business & Economic Statistics*, 12(2), 167–176. doi:10.1080/07350015.1994.10510005
- MacKinnon, J. G. (2010). Critical values for cointegration tests. Queen's Economics Department Working Paper No. 1227, Queen's University. ideas.repec.org/p/qed/wpaper/1227
- Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14(1–2), 3–24. doi:10.1016/0022-1996(83)90017-X
- Nelson, C. R., & Kang, H. (1981). Spurious periodicity in inappropriately detrended time series. *Econometrica*, 49(3), 741–751. doi:10.2307/1911520

References (3/3)

- Nelson, C. R., & Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, 10(2), 139–162. doi:10.1016/0304-3932(82)90012-5
- Ng, S., & Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6), 1519–1554. doi:10.1111/1468-0262.00256
- Perron, P. (1989). The Great Crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6), 1361–1401. doi:10.2307/1913712
- Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33(3), 311–340. doi:10.1016/0304-4076(86)90001-1
- Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–346. doi:10.1093/biomet/75.2.335
- Plosser, C. I., & Schwert, G. W. (1977). Estimation of a non-invertible moving average process: The case of overdifferencing. *Journal of Econometrics*, 6(2), 199–224. doi:10.1016/0304-4076(77)90015-x
- Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3), 599–607. doi:10.1093/biomet/71.3.599
- Schwert, G. W. (1989). Tests for unit roots: A Monte Carlo investigation. *Journal of Business & Economic Statistics*, 7(2), 147–159. doi:10.1080/07350015.1989.10509723
- Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89(1), 1–63. doi:10.2307/2341482
- Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the Great Crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, 10(3), 251–270. doi:10.1080/07350015.1992.10509904