



Time Series Analysis and Forecasting

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

Learning Objectives

By the end of this chapter, you will be able to:

1. Define stochastic processes and understand their properties
2. Distinguish between strict and weak (covariance) stationarity
3. Identify white noise and random walk processes
4. Compute and interpret ACF and PACF
5. Apply the lag operator and differencing
6. Conduct stationarity tests (ADF, KPSS)
7. Analyze financial time series data
8. Distinguish between unit root and trend-stationary processes

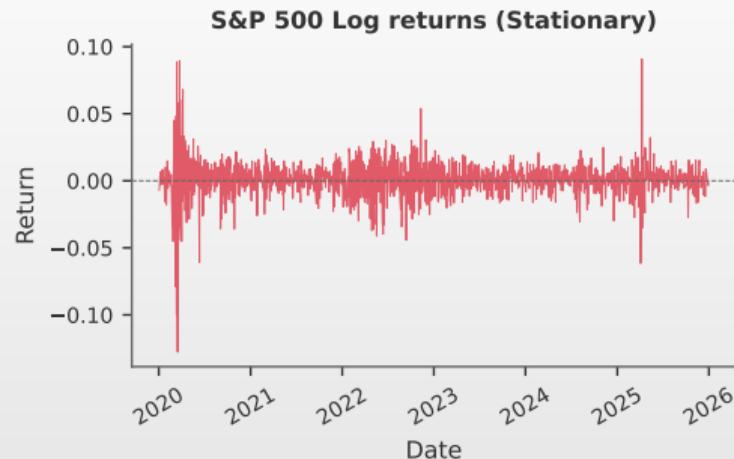
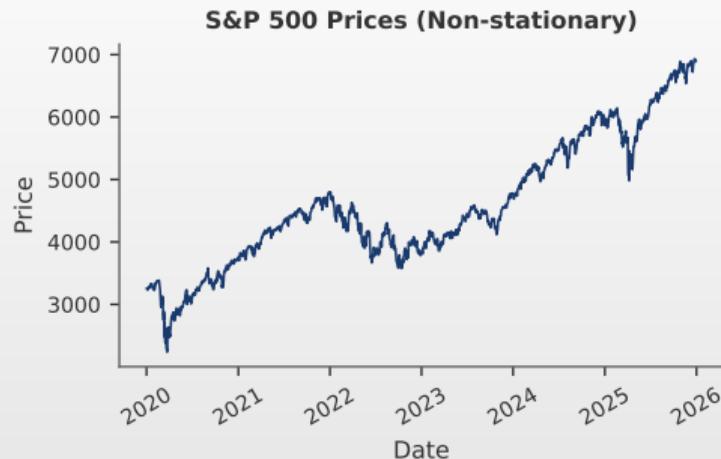


Outline

- Motivation
- Stochastic Processes
- Stationarity
- Lag Operator and Differencing
- White Noise and Random Walk
- Autocorrelation Functions
- Testing for Stationarity
- Financial Data Application
- Case Study: Stationarity Testing
- AI Use Case
- Summary
- Quiz



Examples: stationary vs. non-stationary series

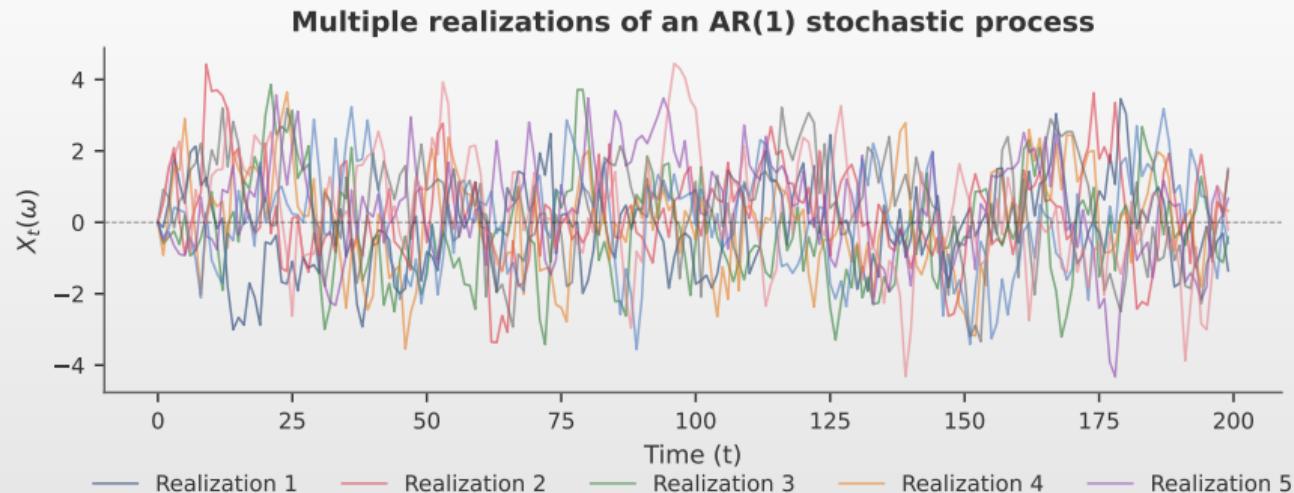


Observations

- Prices (left) are non-stationary: trend, the mean changes over time
- Returns (right) are stationary: mean ≈ 0 , approximately constant variance
- Log returns: $r_t = \ln P_t - \ln P_{t-1} \rightarrow$ non-stationary \rightarrow stationary



Stochastic process: visual illustration



Interpretation

- Each line is a **different realization** from the same underlying stochastic process
- We observe only **one realization**, yet aim to understand the properties of the process



Stochastic process: definition

Definition 1 (Stochastic Process)

- ◻ A **stochastic process** is a collection of random variables indexed by time
 - ▶ $\{X_t(\omega) : t \in \mathcal{T}, \omega \in \Omega\}$
 - ▶ Ω is the sample space of possible outcomes

Two Perspectives

- ◻ Fixed ω : A *realization* $\{X_t(\omega)\}_{t \in \mathcal{T}}$
- ◻ Fixed t : A *random variable* X_t

Key Insight

- ◻ A time series we observe is **one realization** of the underlying stochastic process

Moments of a stochastic process

The First Two Moments Characterize the Process

- **Mean Function:** $\mu_t = \mathbb{E}[X_t]$
- **Autocovariance (ACVF):** $\gamma(t, s) = \text{Cov}(X_t, X_s)$
 - ▶ $\gamma(t, s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)]$
- **Autocorrelation (ACF):**
 - ▶ $\rho(t, s) = \gamma(t, s) / \sqrt{\text{Var}(X_t) \cdot \text{Var}(X_s)}$

ACF Properties

- **Range:** $\rho(t, s) \in [-1, 1]$
- **Normalization:** $\rho(t, t) = 1$ (perfect correlation with itself)

Key Point

- **General:** μ_t and $\gamma(t, s)$ may depend on t
- **Stationary:** Removes this dependence



Why stationarity matters

Without Stationarity

- ◻ Mean, variance change over time
 - ▶ Estimates are inconsistent
- ◻ Past may not predict the future
- ◻ Standard methods fail
- ◻ Spurious correlations

With Stationarity

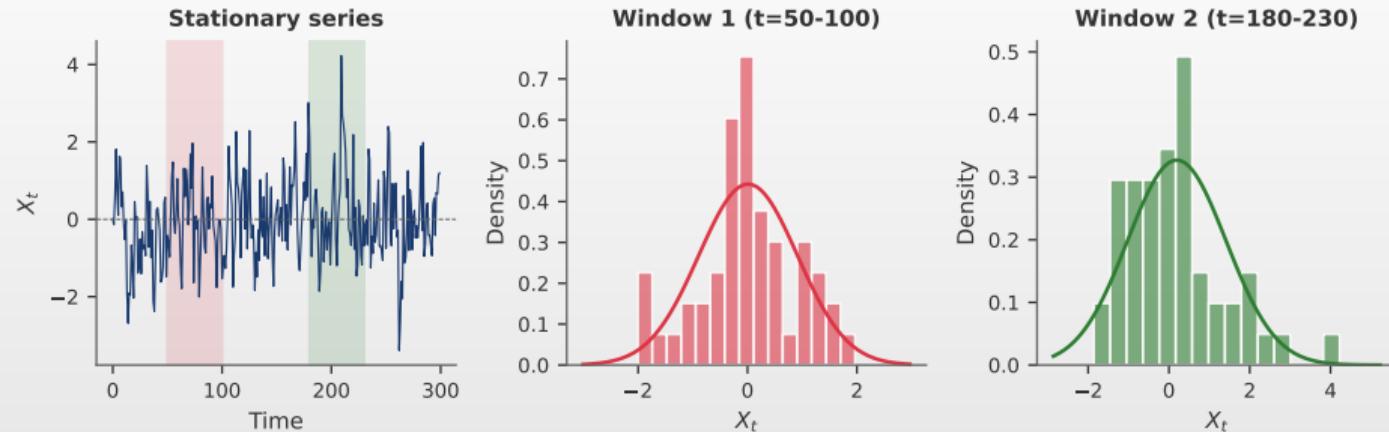
- ◻ Statistical properties constant
 - ▶ Ergodicity justified
- ◻ Can estimate from a single realization
- ◻ Valid inference possible
- ◻ Models are meaningful

Key Principle

- ◻ Most time series models (ARMA, ARIMA, etc.) require stationarity
- ◻ Non-stationary series must be transformed (e.g., differencing) before modeling



Strict stationarity: visual illustration



Interpretation

- Time translation does not change the joint distribution of the variables
- Any two time windows have the same statistical properties
- In practice: we only check the first moments (weak stationarity)



Strict stationarity

Definition 2 (Strict (Strong) Stationarity)

- ◻ A process $\{X_t\}$ is **strictly stationary** if for all k , all t_1, \dots, t_k , and all h :
 - ▶ $(X_{t_1}, \dots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_k+h})$
- ◻ **Notation:** $X \stackrel{d}{=} Y$ means *equality in distribution*
 - ▶ $P(X \leq x) = P(Y \leq x)$

Implications

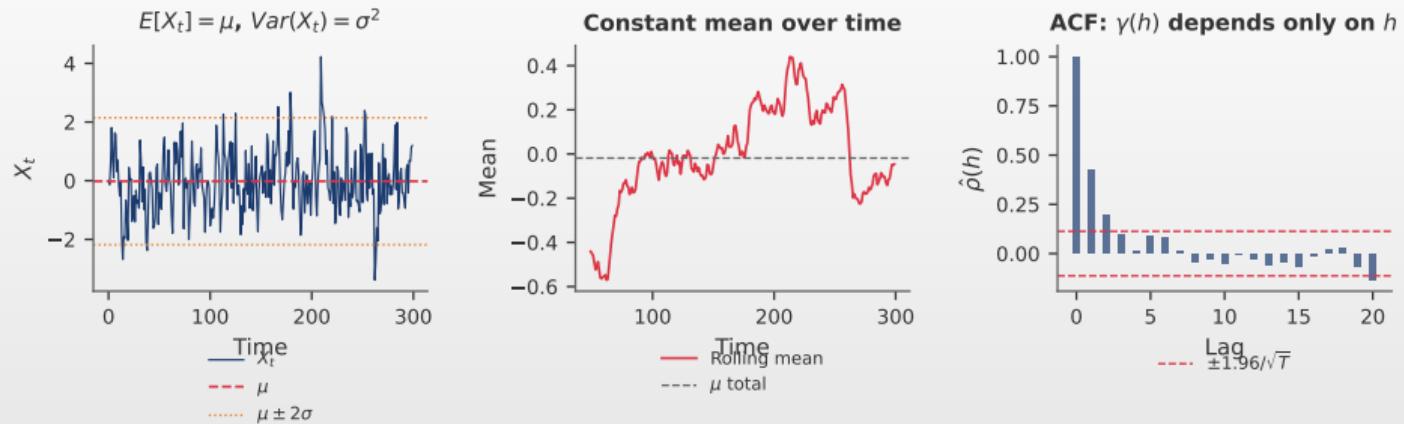
- ◻ **Identical distributions:** $F_{X_t}(x)$ does not depend on t
 - ▶ $\mathbb{E}[X_t] = \mu$ (constant mean, if it exists)
 - ▶ $\text{Var}(X_t) = \sigma^2$ (constant variance, if it exists)
- ◻ **Lag dependence:** Joint distributions depend only on lag

Note

- ◻ Strict stationarity is a strong condition, often impossible to verify in practice



Weak stationarity: visual illustration



The Three Conditions

- $\mathbb{E}[X_t] = \mu$ constant \rightarrow mean does not depend on time
- $\text{Var}(X_t) = \sigma^2$ constant \rightarrow variance does not depend on time
- $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ \rightarrow autocovariance depends only on lag h



Weak (covariance) stationarity

Definition 3 (Weak Stationarity)

- ◻ A process $\{X_t\}$ is **weakly stationary** (or covariance stationary) if:
 - ▶ $\mathbb{E}[X_t^2] < \infty$ for all t — finite second-order moments
 - ▶ $\mathbb{E}[X_t] = \mu$ for all t — constant mean
 - ▶ $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ — covariance depends only on lag h , not on t

Key Properties

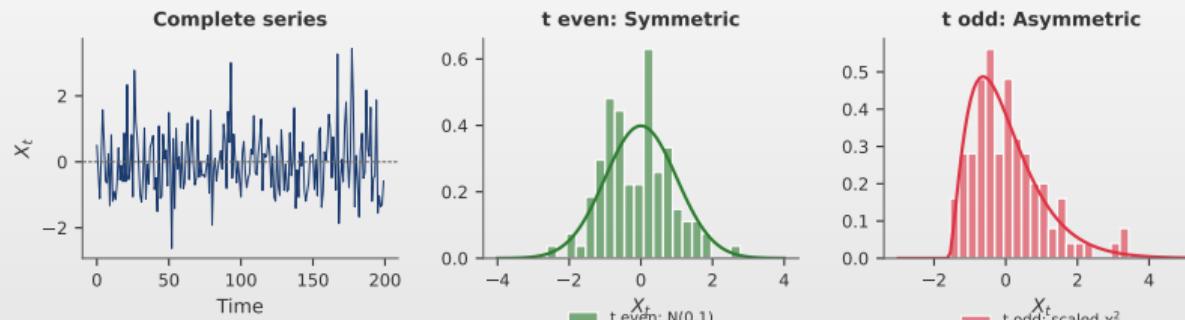
- ◻ **Autocovariance:** $\gamma(h) = \text{Cov}(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)]$
- ◻ **Autocorrelation:** $\rho(h) = \gamma(h)/\gamma(0) = \text{Cov}(X_t, X_{t+h})/\text{Var}(X_t)$
- ◻ **Note:** $\rho(0) = 1$, $|\rho(h)| \leq 1$, $\rho(h) = \rho(-h)$ (symmetry)



Counterexample: weakly stationary but NOT strictly stationary

Construction

- Let $\{X_t\}$ be **independent** random variables with: t even: $X_t \sim N(0, 1)$; t odd: $X_t \sim \frac{\chi^2(5)-5}{\sqrt{10}}$



Weakly stationary ✓

- $\mathbb{E}[X_t] = 0$, $\text{Var}(X_t) = 1$, $\text{Cov}(X_t, X_{t+h}) = 0$

NOT strictly stationary ✗

- Skewness differs (0 vs > 0) $\rightarrow X_1 \neq X_2$

Q TSA_ch1_stationarity



Relationship between strict and weak stationarity

Theorem 1 (Fundamental Implication)

If $\{X_t\}$ is **strictly stationary** and $\mathbb{E}[X_t^2] < \infty$, then $\{X_t\}$ is also **weakly stationary**.

Proof.

- Let t_1, t_2 be arbitrary and h any time shift
- From joint distribution invariance: $(X_{t_1}, X_{t_2}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h})$
- $\mathbb{E}[X_{t_1}] = \mathbb{E}[X_{t_1+h}] = \mu$ (constant mean)
- $\text{Cov}(X_{t_1}, X_{t_2}) = \text{Cov}(X_{t_1+h}, X_{t_2+h})$
- Thus autocovariance depends only on the difference $t_2 - t_1 = h$, not on t_1



Warning: The Converse is NOT True!

- There exist weakly stationary processes that are **not** strictly stationary



Properties of the autocovariance function

Proposition 1

For a weakly stationary process, the ACVF $\gamma(h)$ satisfies:

- **Symmetry:** $\gamma(h) = \gamma(-h)$
- **Maximum at zero:** $|\gamma(h)| \leq \gamma(0) = \text{Var}(X_t)$
- **Non-negative definiteness:** $\sum_{i,j} a_i a_j \gamma(i - j) \geq 0$ for any a_1, \dots, a_n

Proof (property 3)

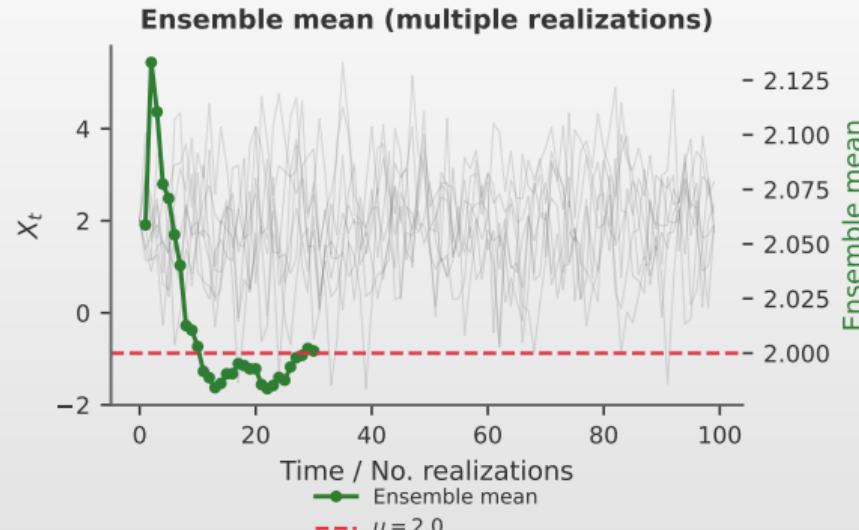
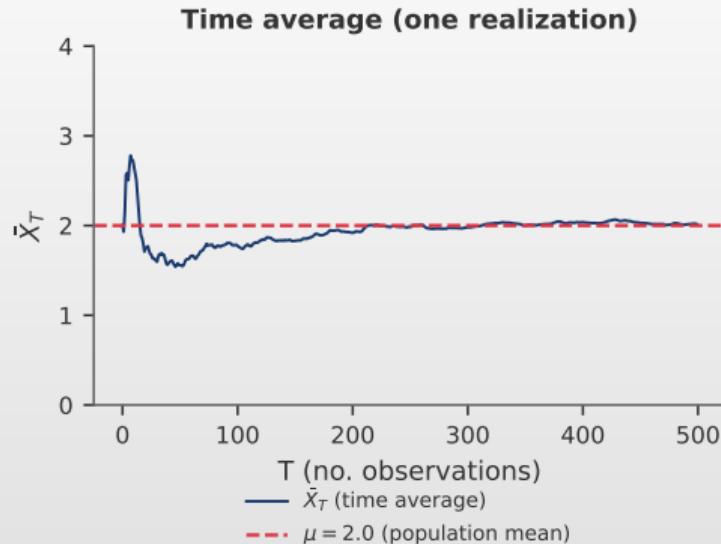
- $\text{Var}(\sum_{i=1}^n a_i X_{t+i}) = \sum_{i,j} a_i a_j \gamma(i - j) \geq 0$ (variance ≥ 0)

Implication

- Not every function can be a valid autocovariance function



Ergodicity: visual illustration



- Time average (single realization) and ensemble average (multiple realizations) both converge to μ
- Ergodicity guarantees that we can estimate μ from a single sufficiently long time series



Ergodicity: the foundation of inference from data

Definition 4 (Ergodicity for Mean)

- A stationary process $\{X_t\}$ is **ergodic for the mean** if:
 - ▶ $\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t \xrightarrow{P} \mathbb{E}[X_t] = \mu$ as $T \rightarrow \infty$

Why does ergodicity matter?

- **Problem:** We have only **one realization** of the stochastic process
- **Solution:** Ergodicity allows estimating μ from \bar{X}_T
 - ▶ The time average converges to the population mean
 - ▶ Without ergodicity, statistical inference is not possible!

Theorem 2 (Sufficient Condition)

If $\sum_{h=0}^{\infty} |\gamma(h)| < \infty$ (absolutely summable autocovariances), the process is ergodic.



The Wold decomposition theorem

Theorem 3 (Wold, 1938)

Any **covariance stationary** process $\{X_t\}$ can be written as: $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \eta_t$

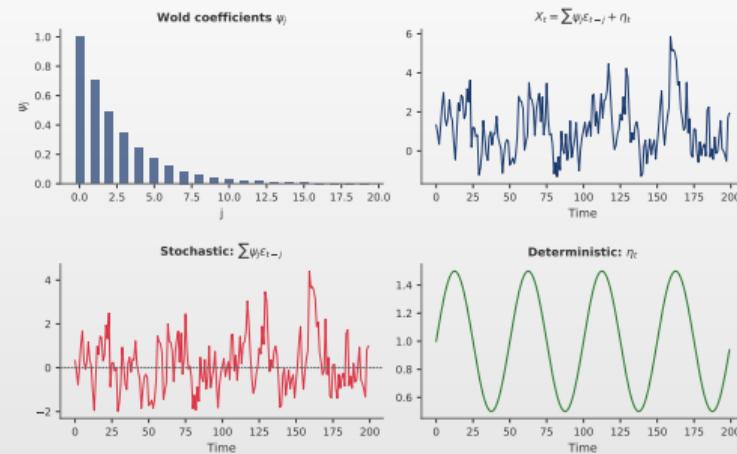
- ◻ $\varepsilon_t \sim WN(0, \sigma^2) \rightarrow$ white noise
 - ▶ $\psi_0 = 1, \sum \psi_j^2 < \infty$
- ◻ $\eta_t \rightarrow$ deterministic component (perfectly predictable)

Significance of the Wold Theorem

- ◻ **Decomposition:** Any stationary process = $MA(\infty) +$ deterministic component
 - ▶ Theoretically justifies $MA(q)$ and $ARMA(p, q)$ models
 - ▶ Coefficients ψ_j measure the impact of past shocks



The Wold theorem: visual illustration

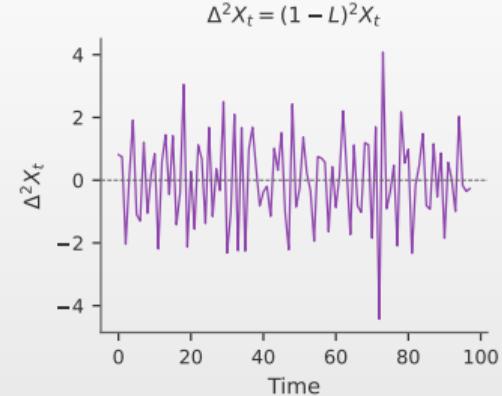
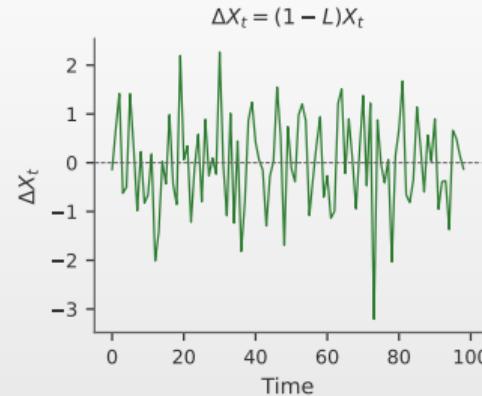
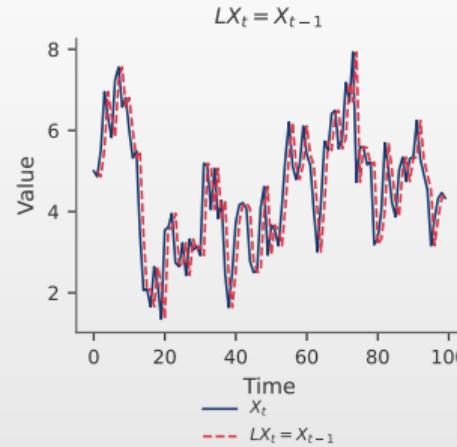


Interpretation

- X_t decomposes into a **stochastic** component ($MA(\infty)$) and a **deterministic** component (η_t)
- Coefficients ψ_j decay → recent shocks have greater impact than distant ones



Lag operator: visual illustration



Properties

- $LX_t = X_{t-1} \rightarrow$ the lag operator shifts the series back by one period
- $L^k X_t = X_{t-k} \rightarrow$ shift by k periods; $L^0 = I$ (identity)
- **Difference operator:** $\Delta = (1 - L)$, so $\Delta X_t = X_t - X_{t-1}$



The lag operator

Definition 5 (Lag Operator)

- The **lag operator** (or backshift operator) L is defined by: $LX_t = X_{t-1}$

Properties

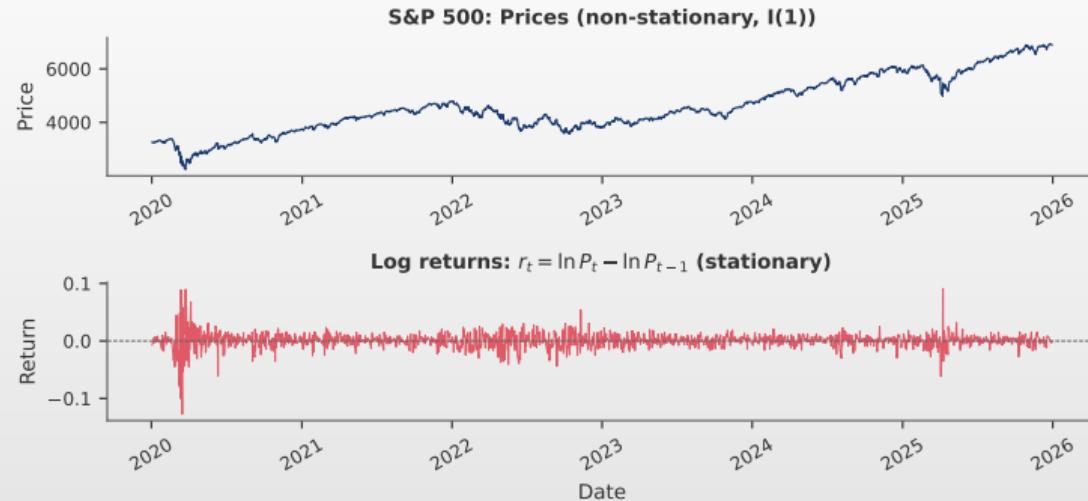
- **Powers:** $L^k X_t = X_{t-k}$ (lag by k periods)
 - ▶ Compact notation for models
- **Identity:** $L^0 = I$
- **Polynomial:** $(1 - \phi L)X_t = X_t - \phi X_{t-1}$

Examples

- **First difference:** $(1 - L)X_t = X_t - X_{t-1}$
- **Second difference:** $(1 - L)^2 X_t = \Delta^2 X_t$
- **Seasonal:** $(1 - L^{12})X_t$



Effect of differencing: S&P 500



Interpretation

- Top:** S&P 500 prices → clear trend, non-stationary ($I(1)$)
- Bottom:** Log returns $r_t = \ln P_t - \ln P_{t-1} \rightarrow$ fluctuates around mean ≈ 0 , stationary



Differencing

Why Do We Difference?

- **First Difference:** $\Delta X_t = X_t - X_{t-1} = (1 - L)X_t$
 - ▶ Removes trend and unit root
 - ▶ Random walk: $\Delta X_t = \varepsilon_t$

Definition 6 (Integrated Process of Order d)

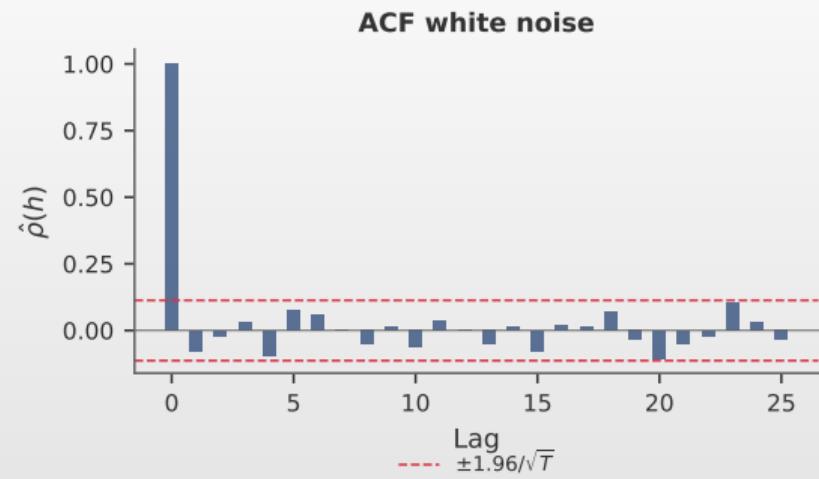
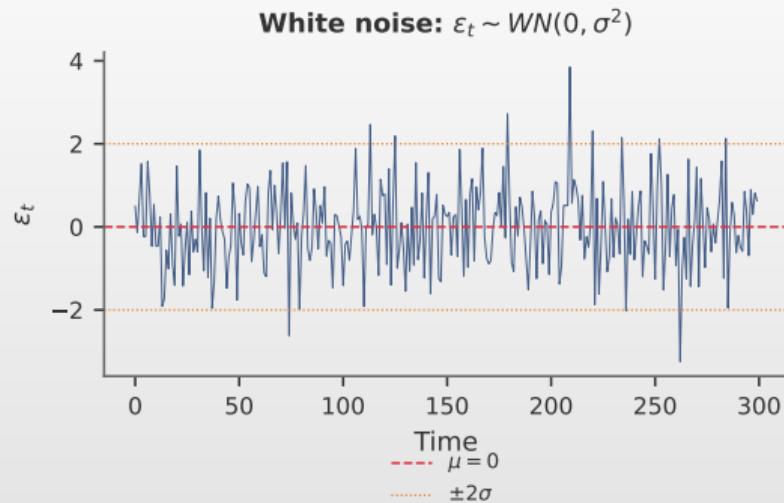
- A process $\{X_t\}$ is **integrated of order d** , denoted $X_t \sim I(d)$, if:
 - ▶ $\Delta^d X_t = (1 - L)^d X_t$ is stationary ($I(0)$ process)
 - ▶ $\Delta^{d-1} X_t$ is **not** stationary

Examples

- $I(0)$: Stationary process (white noise, stationary AR)
- $I(1)$: Random walk $\rightarrow \Delta X_t = \varepsilon_t$ is stationary
- $I(2)$: Requires two differences for stationarity



White noise: visual illustration



Q TSA_ch1_white_noise



White noise process

Definition 7 (White Noise)

- A process $\{\varepsilon_t\}$ is **white noise**, denoted $\varepsilon_t \sim WN(0, \sigma^2)$, if:
 - ▶ $\mathbb{E}[\varepsilon_t] = 0$ for all t (zero mean)
 - ▶ $\text{Var}(\varepsilon_t) = \sigma^2$ for all t (constant variance)
 - ▶ $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$ (uncorrelated)

ACF of White Noise

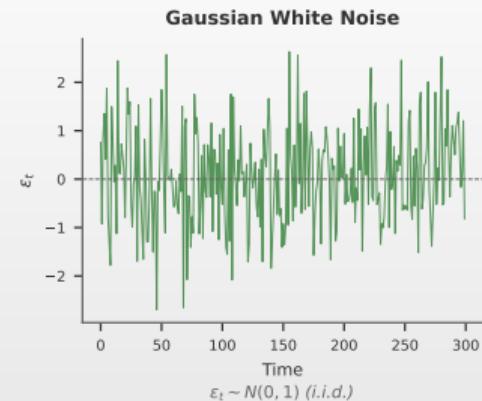
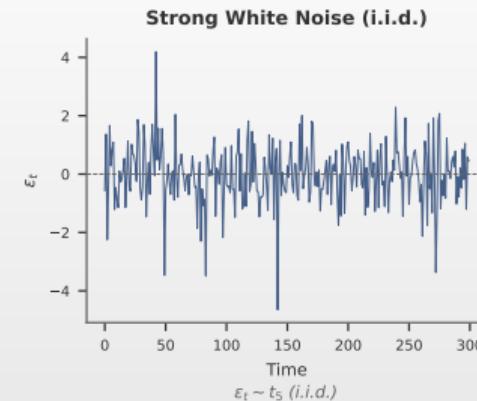
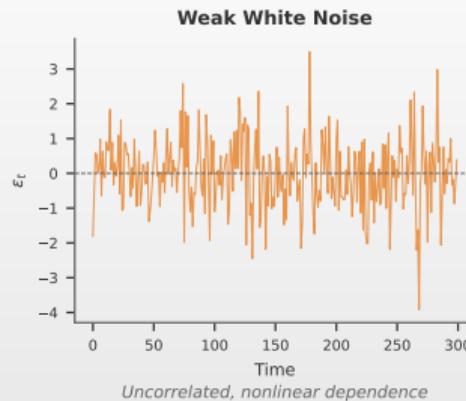
- By definition: $\gamma(0) = \sigma^2$ and $\gamma(h) = 0$ for $h \neq 0$; $\rho(h) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}$

Types of white noise (in order of increasing restrictions)

- **Weak:** uncorrelated, but nonlinear dependencies may exist
- **Strong:** ε_t are *independent* and identically distributed (i.i.d.)
- **Gaussian:** $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$
 - ▶ Uncorrelated \Rightarrow independent



The three types of white noise

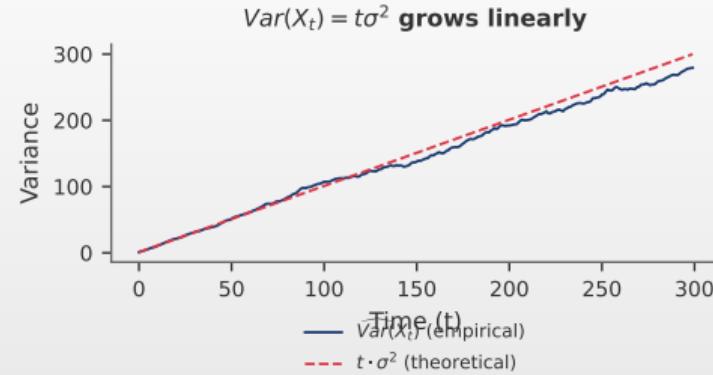
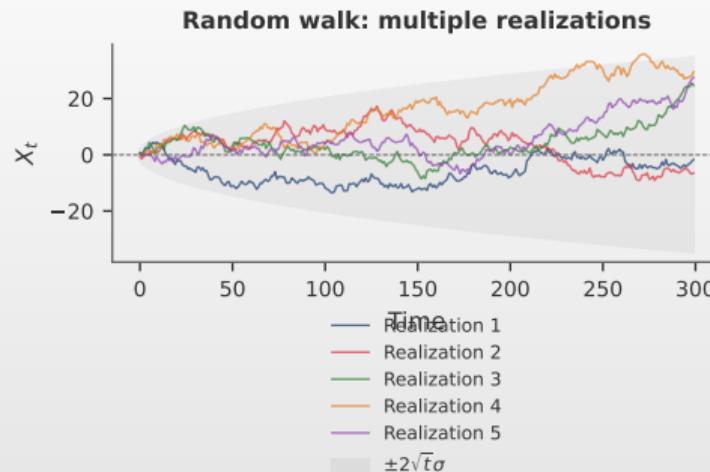


Inclusion relationship: Gaussian \subset Strong (i.i.d.) \subset Weak (uncorrelated)

- **Weak:** $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$, but nonlinear dependencies may exist (e.g. GARCH)
- **Strong:** ε_t are i.i.d. — any distribution (e.g. Student- t)
- **Gaussian:** $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$ — uncorrelated \Leftrightarrow independent



Random walk: visualization



Observations

- Each shock has a **permanent effect**; $\text{Var}(X_t) = t\sigma^2$ grows linearly with time
- Solution** — differencing transforms into white noise, $\Delta X_t = \varepsilon_t$



Random walk process

Definition 8 (Random Walk)

$X_t = X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2), \quad X_0 = 0 \quad \Rightarrow \text{Explicit form: } X_t = \sum_{i=1}^t \varepsilon_i$

Proposition 2 (Properties)

- ◻ $\mathbb{E}[X_t] = 0$
- ◻ $\text{Var}(X_t) = t\sigma^2$ (grows with time!)
- ◻ $\text{Cov}(X_t, X_s) = \min(t, s) \cdot \sigma^2$

Proofs.

- ◻ $\mathbb{E}[X_t] = \mathbb{E}\left[\sum_{i=1}^t \varepsilon_i\right] = 0$
- ◻ $\text{Var}(X_t) = \text{Var}\left(\sum_{i=1}^t \varepsilon_i\right) = \sum_{i=1}^t \text{Var}(\varepsilon_i) = t\sigma^2 \quad (\text{independence})$
- ◻ $\text{Cov}(X_t, X_s) = \min(t, s) \sigma^2 \quad (\text{for } s \leq t)$

□

Non-Stationary!

$\text{Var}(X_t) = t\sigma^2$ depends on $t \rightarrow$ random walk is **not stationary**



Random walk with drift

Definition 9 (Random Walk with Drift)

$X_t = c + X_{t-1} + \varepsilon_t$, $c \neq 0$ is the **drift** \Rightarrow **Explicit form:** $X_t = ct + \sum_{i=1}^t \varepsilon_i$

Proposition 3 (Properties)

- $\mathbb{E}[X_t] = ct$ (linear trend)
- $\text{Var}(X_t) = t\sigma^2$ (grows with time)

Differencing

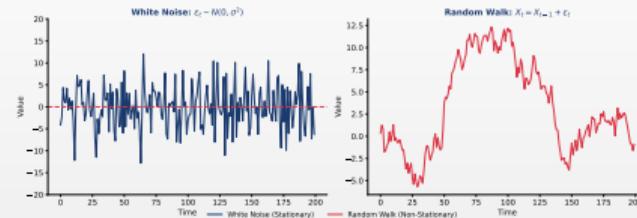
$\Delta X_t = c + \varepsilon_t$ — constant plus white noise \rightarrow the differenced series is stationary

Practical Importance

- Nominal GDP, stock prices \rightarrow often modeled as RW with drift
- The ADF test includes variants: without constant, with constant, with constant and trend



White noise vs random walk: comparison



Q TSA_ch1_random_walk

White Noise

- ☐ Stationary; $\text{Var} = \sigma^2$ (const.); $\text{ACF} = 0, h \neq 0$; no memory

Random Walk

- ☐ Non-stationary; $\text{Var} = t\sigma^2$ (grows); $\text{ACF} \approx 1$ (slow); permanent shocks

Link

- ☐ $\Delta X_t = \varepsilon_t$



Trend-stationary vs. difference-stationary

Trend-Stationary (TS)

- ◻ **Model:** $Y_t = \alpha + \beta t + \varepsilon_t$
 - ▶ **Deterministic trend**
 - ▶ Deviations from the trend are temporary
- ◻ **Solution:** regression on t , extract residuals
- ◻ **Effect:** Shocks do NOT have a permanent effect

Difference-Stationary (DS)

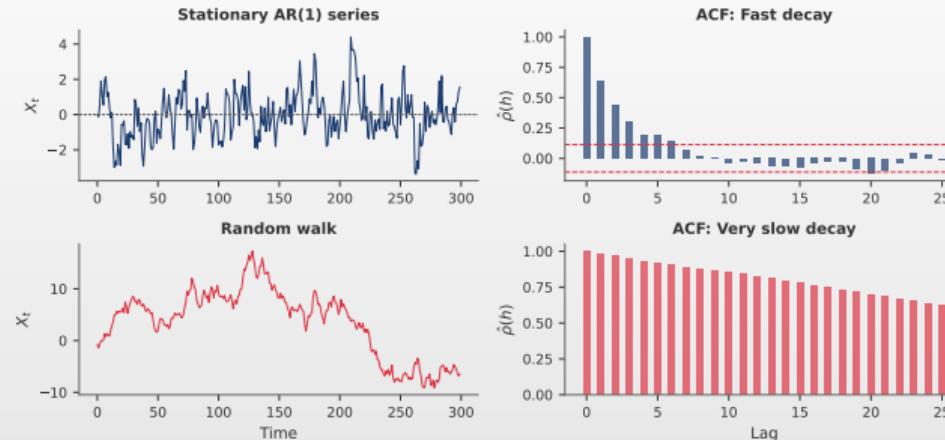
- ◻ **Model:** $Y_t = c + Y_{t-1} + \varepsilon_t$
 - ▶ **Stochastic trend**
 - ▶ Deviations from the trend are permanent
- ◻ **Solution:** differencing (ΔY_t)
- ◻ **Effect:** Shocks HAVE a permanent effect

Why does the distinction matter?

- ◻ **Differencing a TS process:** introduces an artificial unit root in the MA part
- ◻ **Regression on a DS process:** produces residuals that are still non-stationary
- ◻ **Solution:** ADF and KPSS tests help distinguish between the two



ACF comparison: stationary vs random walk

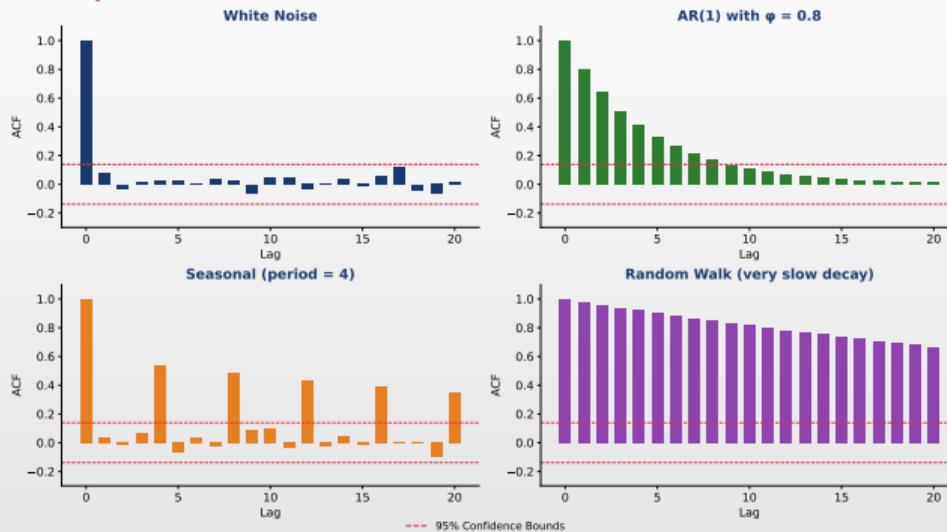


Interpretation

- **Stationary:** ACF decays rapidly (exponentially or oscillating) toward zero
- **Random walk:** ACF decays very slowly, stays close to 1
- **Rule of thumb:** Slow ACF decay → suspect unit root → ADF test



ACF patterns for different processes



Interpretation

- White noise: $ACF = 0$; Stationary: decays fast; Non-stationary: decays slowly
- Seasonal: Spikes at seasonal lags (12, 24 for monthly data)



Sample autocorrelation function

Sample ACF at Lag h

$$\square \hat{\rho}(h) = \frac{\sum_{t=1}^{T-h} (x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

- ▶ Properties: $\hat{\rho}(0) = 1$, $|\hat{\rho}(h)| \leq 1$

Theorem 4 (Bartlett, 1946)

Under H_0 : white noise, for large T : $\hat{\rho}(h) \approx N(0, 1/T)$

95% Confidence Interval

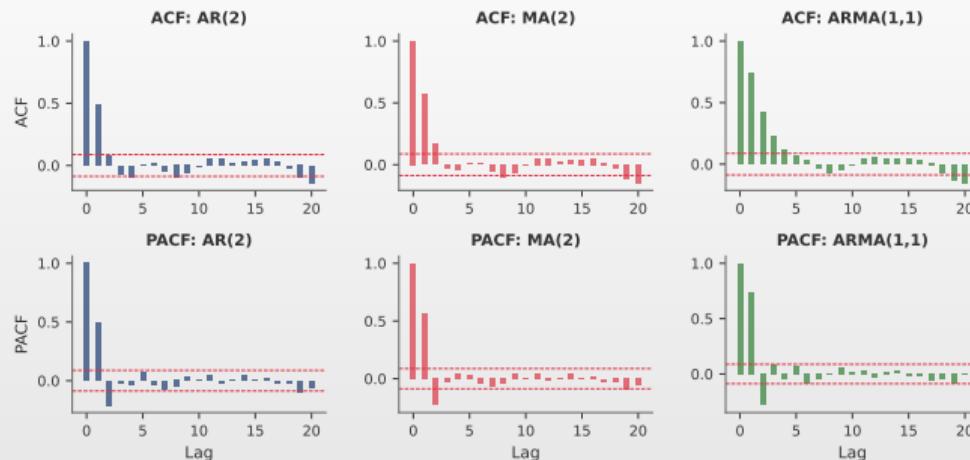
- $\pm 1.96/\sqrt{T}$ (the bands in ACF plots)

Caution

- Bartlett's formula is valid **only under H_0 : white noise**
- For AR/MA, the asymptotic variance differs



ACF and PACF patterns



Identification Rules

- **AR(p):** ACF decays exponentially, PACF cuts off after lag p
- **MA(q):** ACF cuts off after lag q , PACF decays exponentially
- **ARMA(p, q):** Both decay exponentially → identification requires information criteria



Partial autocorrelation function (PACF)

Definition 10 (Partial Autocorrelation)

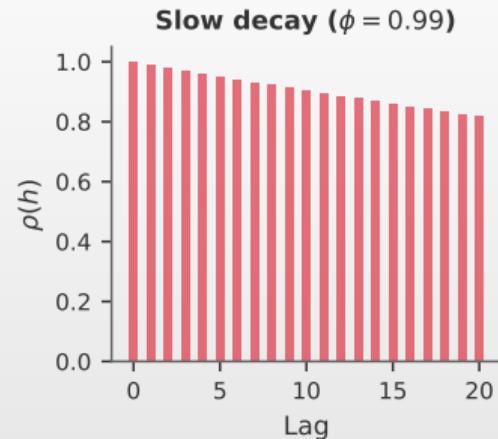
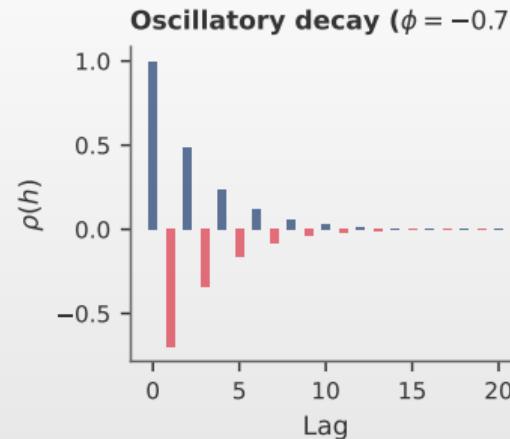
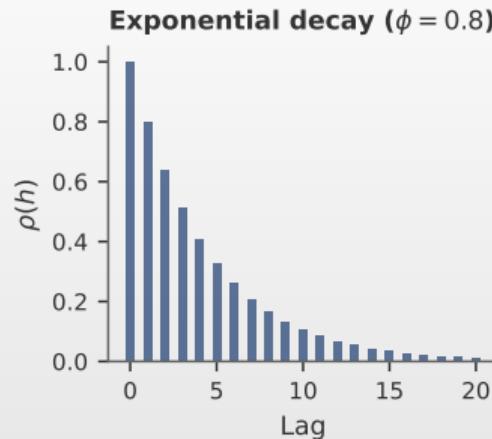
- PACF at lag h , denoted ϕ_{hh} : the last coefficient in the regression:
 - ▶ $X_t = \phi_{h1}X_{t-1} + \phi_{h2}X_{t-2} + \cdots + \phi_{hh}X_{t-h} + e_t$
- Alternatively:
 - ▶ $\phi_{hh} = \text{Corr}(X_t - \hat{X}_t^{(h-1)}, X_{t-h} - \hat{X}_{t-h}^{(h-1)})$
- Interpretation: Direct dependence at lag h
 - ▶ Removes the effect of intermediate lags

Key Application: Model Order Identification

- AR(p): PACF cuts off after lag p
 - ▶ ACF decays exponentially or oscillates
- MA(q): ACF cuts off after lag q
 - ▶ PACF decays exponentially or oscillates



ACF decay patterns



Interpretation

- ☐ **Exponential decay:** Persistent positive dependence (AR with $\phi > 0$)
- ☐ **Oscillating decay:** Alternating dependence (AR with $\phi < 0$)
- ☐ The decay rate indicates the strength of the process memory



Augmented Dickey-Fuller (ADF) test

ADF Model

$$\Delta X_t = \alpha + \gamma X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t, \quad \gamma = \rho - 1, \quad H_0 : \gamma = 0 \Leftrightarrow \rho = 1$$

Hypotheses

- $H_0: \gamma = 0$ (unit root)
- $H_1: \gamma < 0$ (stationary)

Test Statistic

- $\tau_{ADF} = \hat{\gamma} / SE(\hat{\gamma})$
- $\hat{\gamma}$ = OLS coefficient of X_{t-1}
- $SE(\hat{\gamma})$ from the OLS regression

Decision Rule

- $\tau_{ADF} <$ critical value \rightarrow Reject $H_0 \rightarrow$ Stationary
- $\tau_{ADF} \geq$ critical value \rightarrow Non-stationary (unit root)
- Critical values follow the Dickey-Fuller distribution (**not t-Student!**)



KPSS test

Model

- $X_t = \xi t + r_t + \varepsilon_t$ where $r_t = r_{t-1} + u_t$

Hypotheses (opposite of ADF)

- $H_0: \sigma_u^2 = 0$ (stationary)
- $H_1: \sigma_u^2 > 0$ (unit root)

Test Statistic

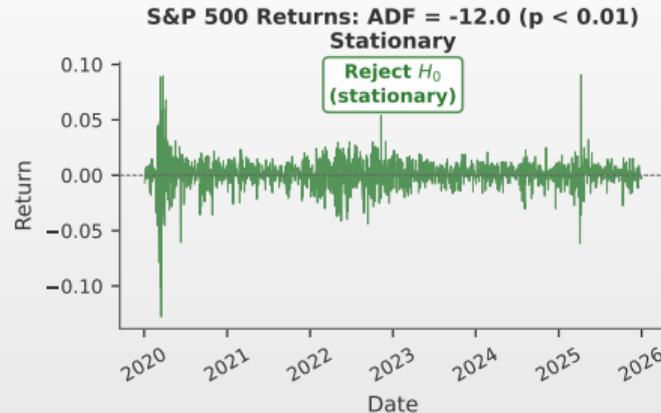
- $LM = \frac{\sum_{t=1}^T S_t^2}{T^2 \hat{\sigma}_{LR}^2}$
- $S_t = \sum_{i=1}^t \hat{e}_i, \quad \hat{\sigma}_{LR}^2 = \text{long-run variance}$

Decision Rule

- $LM > \text{critical value} \rightarrow \text{Reject } H_0 \rightarrow \text{Non-stationary}$
- $LM \leq \text{critical value} \rightarrow \text{Stationary}$



ADF test: visualization with S&P 500



Q TSA_ch1_unit_root_tests

Interpreting the ADF Test

- Hypothesis:** H_0 : Unit root
 - ▶ Critical values: -3.43 (1%), -2.86 (5%), -2.57 (10%)
 - ▶ $\tau <$ critical value \rightarrow reject $H_0 \rightarrow$ stationary series
- S&P 500:** Prices non-stationary; Returns stationary



Using ADF and KPSS together

Confirmatory Testing

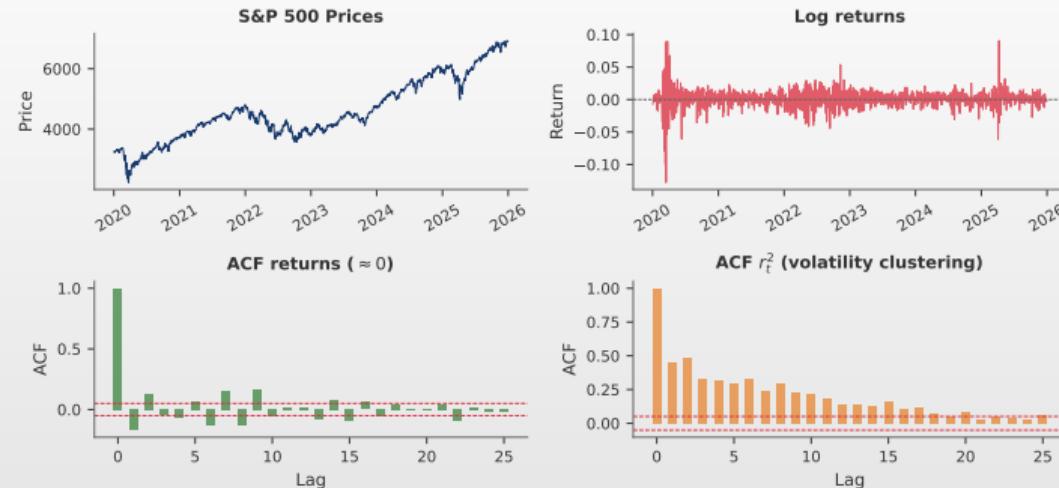
- ADF rejects H_0 + KPSS fails to reject: **Stationary**
- ADF fails to reject + KPSS rejects H_0 : **Unit Root**
- Both reject or both fail to reject: **Inconclusive**
 - ▶ Additional tests required (PP, DF-GLS)

Workflow

- Step 1:** ADF test (H_0 : unit root)
- Step 2:** KPSS test (H_0 : stationary)
- Step 3:** Concordant results → OK
 - ▶ Otherwise: PP, DF-GLS tests



S&P 500 analysis: overview

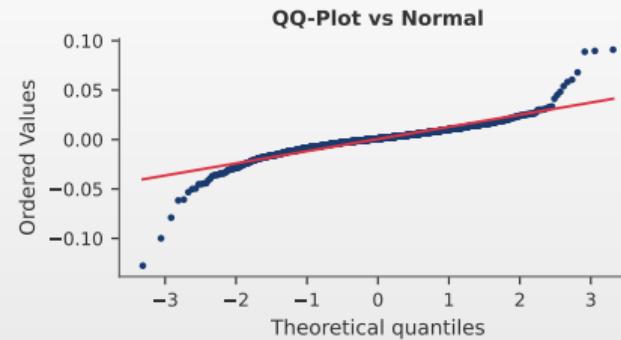
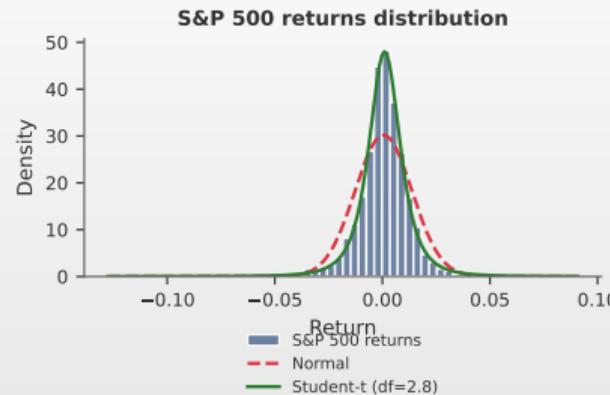


Observations

- **Prices:** Upward trend, non-stationary; **Returns:** Mean ≈ 0 , stationary
- **ACF returns:** ≈ 0 (efficient); **ACF r_t^2 :** Significant (volatility clustering)



Stylized facts of financial returns



Observed Properties

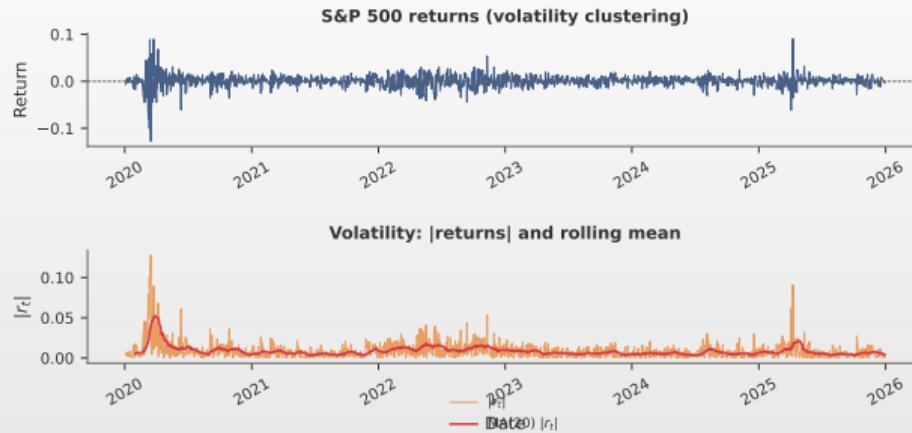
- ☐ Negative skewness (left tail)
- ☐ Excess kurtosis ($\gg 3$)
- ☐ Heavy tails (fat tails)

Implications

- ☐ Normal distribution inadequate
- ☐ Extreme events more likely
- ☐ Student-t or GED required



Volatility clustering

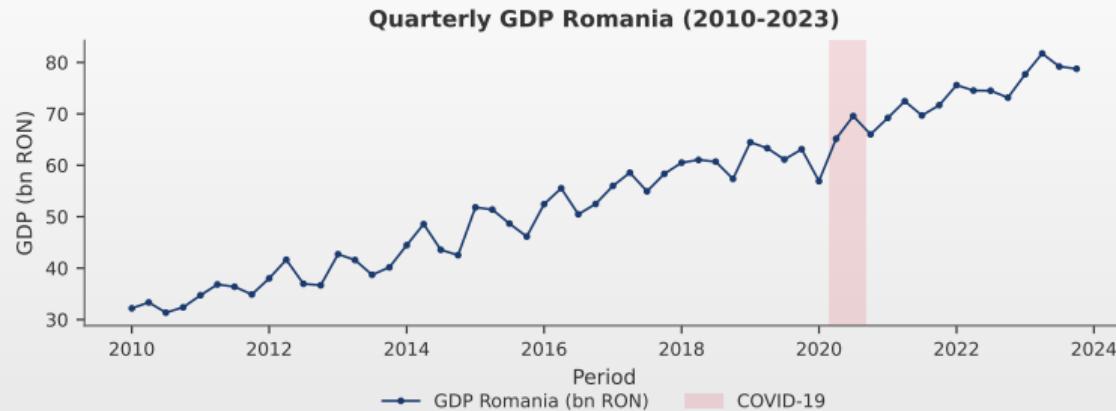


Observations

- Large returns (in absolute value) followed by large returns
- Calm periods followed by calm periods
- Time-varying volatility** → ARCH/GARCH models (Ch. 5)



Case study: Romanian quarterly GDP



 TSA_ch1_case_gdp

Initial Analysis

- Data:** Romanian quarterly GDP 2010–2023 (56 obs., INS/Eurostat)
- Observations:** Upward trend, possibly seasonal
 - ▶ COVID-19 structural shock visible
- Hypothesis:** Non-stationary series → test with ADF and KPSS



Stationarity testing: ADF and KPSS

ADF Test

- ◻ Hypothesis: H_0 : Unit root
- ◻ Result: ADF stat.: -1.23
 - ▶ Critical value: -2.89
 - ▶ Fail to reject H_0

KPSS Test

- ◻ Hypothesis: H_0 : Stationary
- ◻ Result: KPSS stat.: 0.89
 - ▶ Critical value: 0.46
 - ▶ Reject H_0

Conclusion: Both Tests Agree

- ◻ The GDP series is **non-stationary** → requires differencing



Differencing: transformation to stationarity

After Differencing

- **Tests:** Both confirm stationarity
 - ▶ ADF: -4.56 ($p < 0.01$)
 - ▶ KPSS: 0.21 ($p > 0.10$)

Conclusion

- **GDP level:** non-stationary
- ΔGDP : stationary
 - ▶ Use ΔGDP_t for modeling

Final Result

- GDP requires one differencing to become stationary



AI Exercise: Critical Thinking

Prompt to test in ChatGPT / Claude / Copilot

"Using yfinance, download daily EUR/RON exchange rate data (EURRON=X) from 2020-01-01 to 2024-12-31 (approx. 1,250 observations). Test whether the series is stationary using ADF and KPSS tests. Fit an appropriate model and forecast the exchange rate for the next 5 trading days. Tell me if the forecast is reliable."

Exercise:

1. Run the prompt in an LLM of your choice and critically analyze the response.
2. Download real EUR/RON data and reproduce the analysis. Do the results match?
3. Is the ADF test correctly specified (trend, lags)? What changes if you modify the options?
4. Compare the AI model's forecast against a naïve benchmark ($\hat{X}_{t+1} = X_t$).
5. If the series is a random walk, does fitting an ARMA model make sense?

Warning: Low RMSE and significant coefficients *do not guarantee* a useful forecast.



Key takeaways

Summary

- **Stochastic process:** collection of random variables indexed by time
- **Weak stationarity:** constant mean, variance, autocovariance
- **White noise:** $\varepsilon_t \sim WN(0, \sigma^2)$
 - ▶ Stationary, ACF = 0 for $h \neq 0$
- **Random walk:** $X_t = X_{t-1} + \varepsilon_t$
 - ▶ Non-stationary, $\text{Var}(X_t) = t\sigma^2$
- **ACF/PACF:** key tools for identifying structure
- **Differencing:** transforms non-stationary series into stationary ones
- **Unit root tests:**
 - ▶ ADF (H_0 : unit root) vs KPSS (H_0 : stationary)



Important formulas

Weak Stationarity

- **Constant moments:**
 - ▶ $\mathbb{E}[X_t] = \mu$ (constant mean)
 - ▶ $\text{Var}(X_t) = \sigma^2$ (constant variance)
- **Autocovariance:** $\gamma(h) = \text{Cov}(X_t, X_{t+h})$
- **Autocorrelation:** $\rho(h) = \gamma(h)/\gamma(0)$

Lag Operator

- **Lag:** $LX_t = X_{t-1}$
- **Difference:** $\Delta X_t = (1 - L)X_t$

White Noise (WN)

- **Model:** $\varepsilon_t \sim WN(0, \sigma^2)$
- **ACF:** $\rho(h) = 0$ for $h \neq 0$

Random Walk (RW)

- **Model:** $X_t = X_{t-1} + \varepsilon_t$
- **Variance:** $\text{Var}(X_t) = t\sigma^2$ (grows!)

Next chapter preview

Chapter 2: ARMA Models

- **AR(p):** Autoregressive Models
- **MA(q):** Moving Average Models
- **ARMA(p, q):** Combined Models
- **Identification:** Using ACF/PACF

What We Will Learn

- **Estimation:** Model parameters
- **Diagnostics:** Model validation
- **Forecasting:** Confidence intervals
- **Selection:** AIC, BIC



Question 1

Question

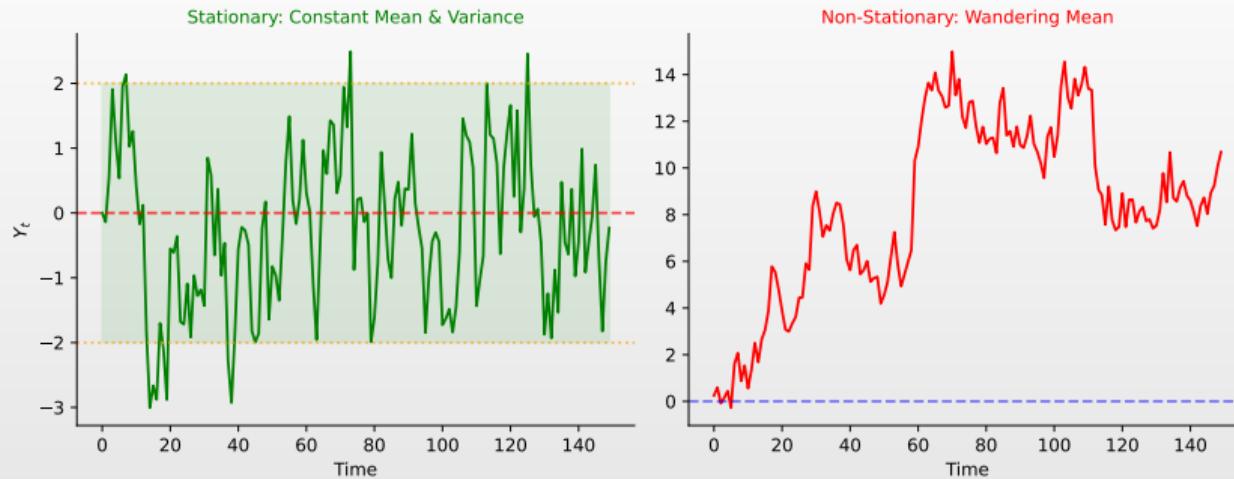
- What are the three conditions for weak (covariance) stationarity?

Answer Choices

- (A)** Zero mean, infinite variance, time-dependent covariance
- (B)** Constant mean, constant variance, autocovariance depends only on lag
- (C)** Normal distribution, independence, unit variance
- (D)** Linear trend, constant seasonality, white residuals



Question 1: Answer



Answer: (B)

- $\mathbb{E}[X_t] = \mu, \text{Var}(X_t) = \sigma^2, \gamma(t, s) = \gamma(|t - s|)$



Question 2

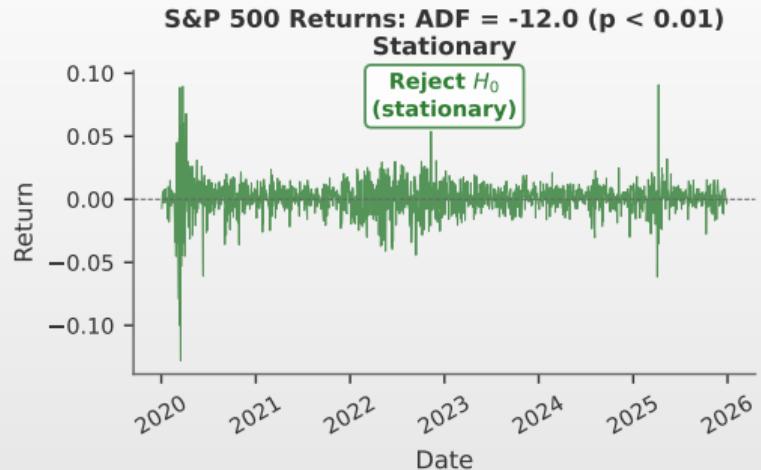
Question

- What is the null hypothesis (H_0) of the ADF (Augmented Dickey-Fuller) test?

Answer Choices

- (A)** The series is stationary
- (B)** The series has a unit root (is non-stationary)
- (C)** The series has no autocorrelation
- (D)** The series has a normal distribution

Question 2: Answer



Answer: (B)

- H_0 : unit root; $\tau <$ critical value \rightarrow stationary



Question 3

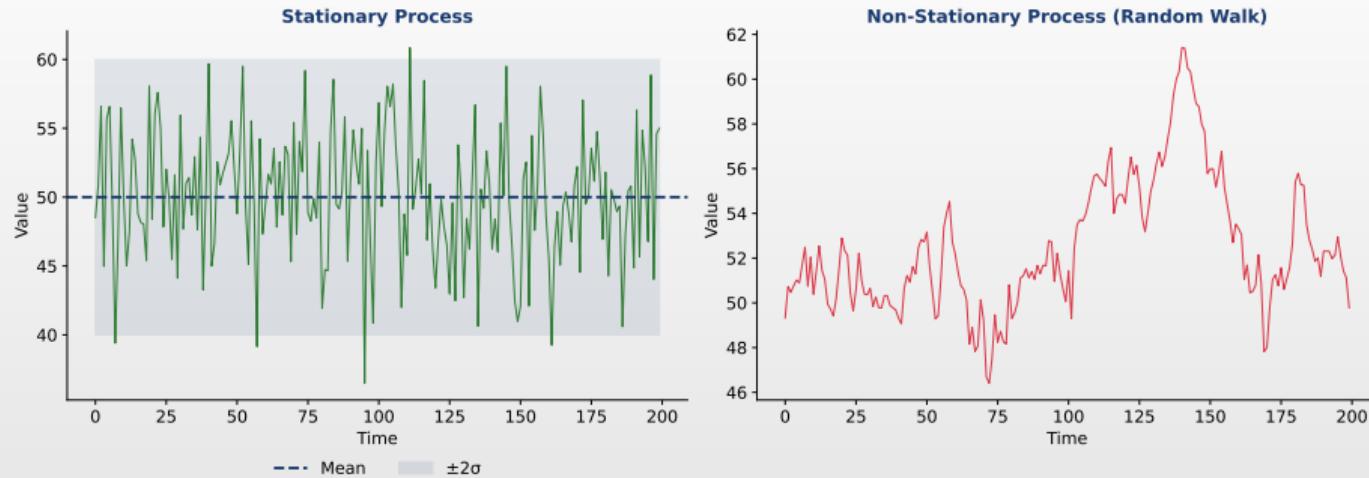
Question

- What is the null hypothesis (H_0) of the KPSS test?

Answer Choices

- (A)** The series has a unit root (non-stationary)
- (B)** The series is stationary
- (C)** The series is a random walk
- (D)** The series has a deterministic trend

Question 3: Answer



Answer: (B)

- KPSS: H_0 stationary (opposite of ADF). Use both tests!



Question 4

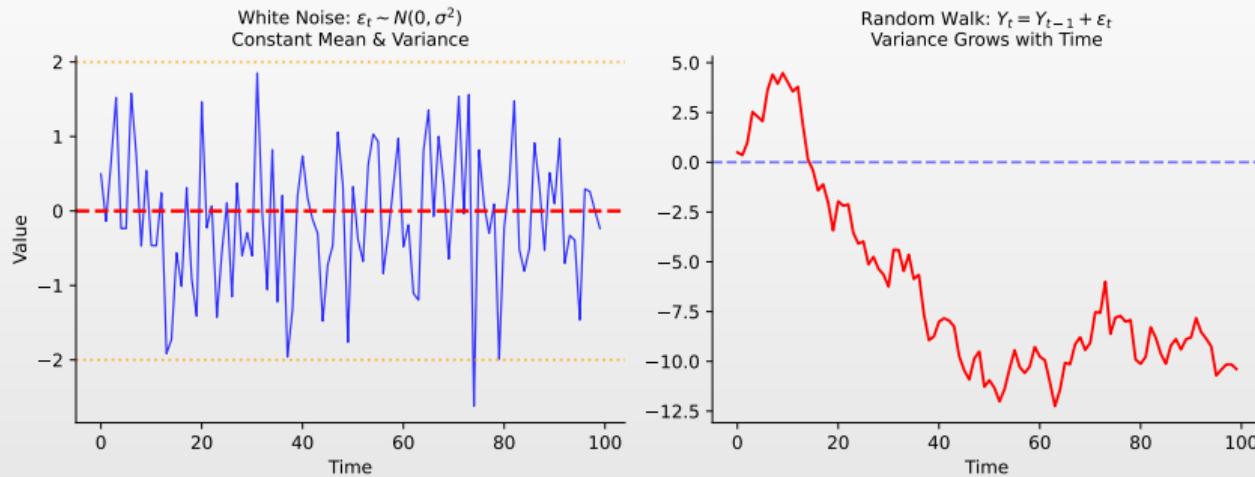
Question

- What is the key property of the variance of a random walk $X_t = X_{t-1} + \varepsilon_t$?

Answer Choices

- (A)** Variance is constant: $\text{Var}(X_t) = \sigma^2$
- (B)** Variance grows linearly with time: $\text{Var}(X_t) = t\sigma^2$
- (C)** Variance decreases with time
- (D)** Variance is zero

Question 4: Answer



Answer: (B)

- $\text{Var}(X_t) = t\sigma^2$ grows linearly \rightarrow non-stationary



Question 5

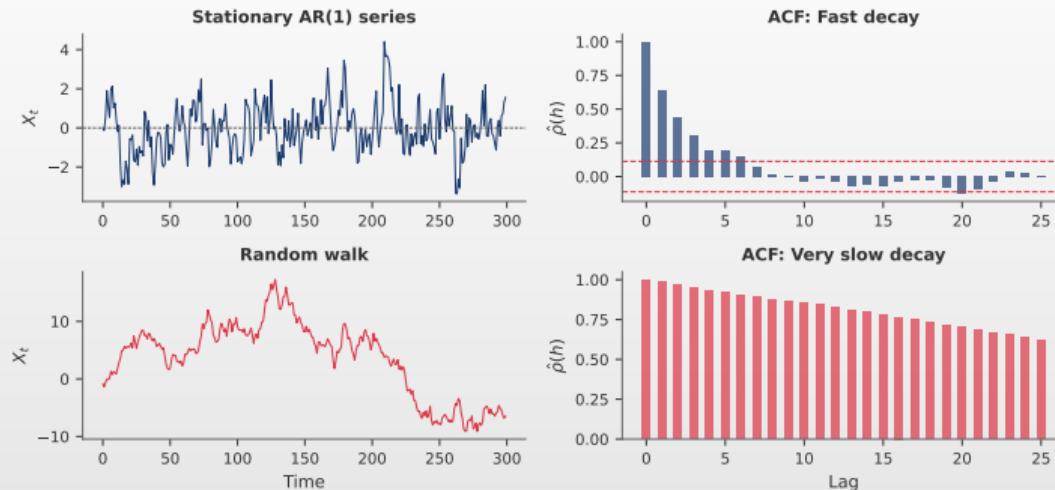
Question

- What does the ACF of a random walk (non-stationary series with unit root) look like?

Answer Choices

- (A)** All values are zero after lag 0
- (B)** Decays exponentially fast
- (C)** Decays very slowly (high persistence)
- (D)** Oscillates between positive and negative

Question 5: Answer



Answer: (C)

- ACF ≈ 1 for many lags, slow decay \rightarrow ADF test

TSA_ch1_random_walk



Question 6

Question

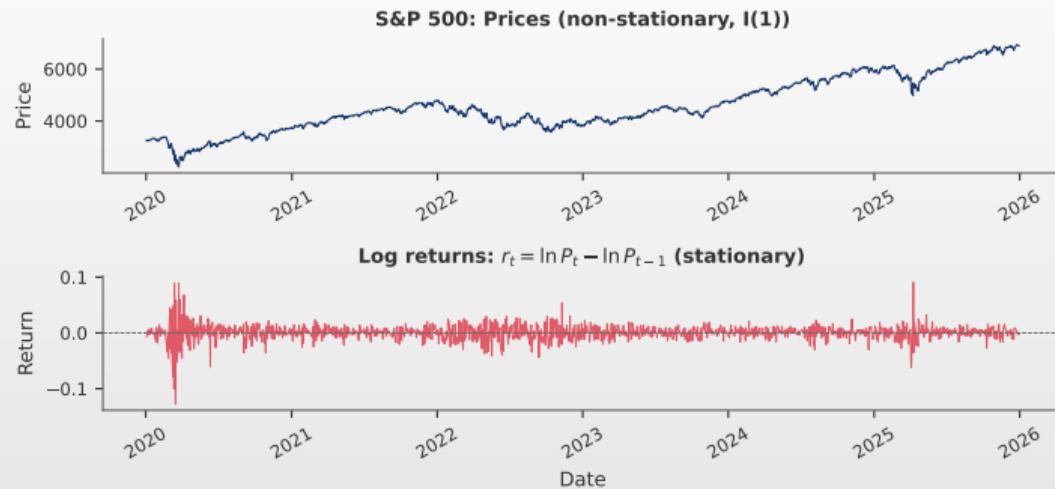
- How do we obtain stationary returns from a financial price series P_t ?

Answer Choices

- (A)** Simple differencing: $\Delta P_t = P_t - P_{t-1}$
- (B)** Log then differencing: $r_t = \ln P_t - \ln P_{t-1}$
- (C)** Log only: $\ln P_t$
- (D)** Standardization: $(P_t - \bar{P})/s_P$



Question 6: Answer



Answer: (B)

- Log returns: $r_t = \ln P_t - \ln P_{t-1}$
- First \ln (stabilizes variance), then Δ (removes trend) \rightarrow stationary series



References

Core Textbooks

- Hyndman & Athanasopoulos (2021). *Forecasting*, OTexts
- Shumway & Stoffer (2017). *Time Series Analysis*, Springer
- Hamilton (1994). *Time Series Analysis*, Princeton

Classic References

- Wold (1938). *Analysis of Stationary Time Series*
- Bartlett (1946). "Sampling Properties", *JRSS*



Thank You!

Questions?

Course materials available at: <https://danpele.github.io/Time-Series-Analysis/>

