Stationarity and ARIMA

BSAD 8310: Business Forecasting — Lecture 4

Department of Economics

University of Nebraska at Omaha

Spring 2026

Stationarity

All classical forecasting models implicitly assume a stable data-generating process.

A time series $\{y_t\}$ is **weakly stationary** if:

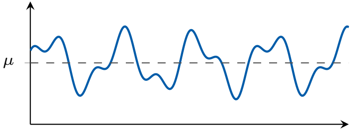
1. $\mathbb{E}[y_t] = \mu \quad$ (constant mean)
2. $\mathrm{Var}(y_t) = \sigma^2 < \infty \quad$ (constant, finite variance)
3. $\mathrm{Cov}(y_t, y_{t-k}) = \gamma_k$ depends only on lag k, not on t

Why it matters for forecasting:

- If $\mathbb{E}[y_t]$ changes over time, no single mean is forecastable
- If $\mathrm{Var}(y_t) \to \infty$, prediction intervals become unbounded
- Stationarity is what makes past patterns informative about the future

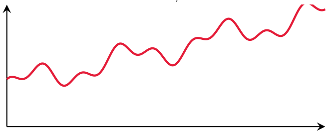
Socratic: if $y_t = t + \varepsilon_t$ (linear trend plus white noise), which stationarity condition does it violate?

AR(1)-like series, $\phi_1 = 0.7$ (simulated):



Fluctuates around μ; bounded variance.

Random-walk-like series, $\phi_1 = 1$ (simulated):



Drifts without bound; variance grows with t.

ETS and AR models require stationarity (or achieve it via trend/seasonal components). ARIMA handles non-stationarity through **differencing**.

Unit Roots and Differencing

A random walk has a unit root — shocks accumulate permanently.

Consider the AR(1) model: $y_t = \phi_1 \, y_{t-1} + \varepsilon_t$

Condition	Behavior	Process
$\lvert\phi_1\rvert < 1$	Shock decays geometrically	Stationary AR(1)
$\phi_1 = 1$	Shock persists permanently	Random walk (unit root)
$\lvert\phi_1\rvert > 1$	Explosion	Explosive (not forecastable)

Random walk expanded: $y_T = y_0 + \sum_{t=1}^{T} \varepsilon_t$

With a unit root, the **effect of every past shock is permanent**. The naïve forecast $\hat{y}_{T+h|T} = y_T$ is optimal for a pure random walk — and the forecast variance grows as $h\sigma^2$.

Rewrite the AR(1) as a regression:

$$\Delta y_t = \delta\, y_{t-1} + \varepsilon_t, \qquad \delta = \phi_1 - 1$$

$H_0 : \delta = 0$ (unit root — non-stationary)

$H_1 : \delta < 0$ (stationary)

Test statistic: $\tau = \hat{\delta}/\mathrm{SE}(\hat{\delta})$ follows a non-standard distribution; critical values from Dickey and Fuller (1979).

Augmentation: include $\Delta y_{t-1}, \ldots, \Delta y_{t-p}$ lags to remove residual autocorrelation. Also include a constant and/or linear trend as appropriate. *Socratic: the ADF statistic does not follow a standard t-distribution even in large samples. What does this mean for using standard regression p-values to test for a unit root?*

First difference removes a stochastic trend (unit root):

$$\Delta y_t = y_t - y_{t-1} = (1 - B)\, y_t$$

where B is the **backshift operator** $(By_t = y_{t-1})$.

Seasonal difference removes seasonal non-stationarity:

$$\Delta_m y_t = y_t - y_{t-m} = (1 - B^m)\, y_t$$

For monthly data: $\Delta_{12} y_t = y_t - y_{t-12}$.

$d = 0$: already stationary
$d = 1$: one first difference
$d = 2$: rarely needed
$D = 1$: one seasonal difference

Over-differencing induces negative autocorrelation. Re-apply ADF/KPSS after differencing: if Δy_t is stationary, stop at $d = 1$. Over-differencing introduces MA unit roots.

The ADF and KPSS (Kwiatkowski–Phillips–Schmidt–Shin) tests address *opposite* null hypotheses:

Test	H_0	H_1
ADF (Dickey and Fuller 1979)	Unit root (non-stationary)	Stationary
KPSS (Kwiatkowski et al. 1992)	Stationary	Unit root

Practical workflow:

1. ADF: fail to reject $\Rightarrow$ evidence of unit root
2. KPSS: reject $\Rightarrow$ evidence against stationarity
3. If both agree $\Rightarrow$ high confidence; if conflicting $\Rightarrow$ inspect ACF and apply domain knowledge

Automatic order selection (`pmdarima.auto_arima`) runs ADF internally to determine d before fitting ARIMA.

ACF and PACF

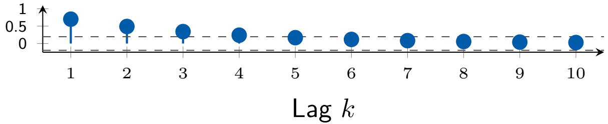
The autocorrelation function is the fingerprint of a time series.

The **lag-k autocorrelation** of a stationary series:

$$\rho_k = \frac{\mathrm{Cov}(y_t,\, y_{t-k})}{\mathrm{Var}(y_t)} = \frac{\gamma_k}{\gamma_0}, \qquad k = 1, 2, \ldots$$

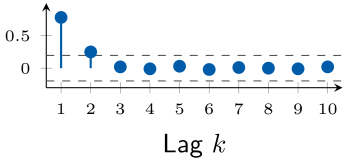
Estimated by $\hat{\rho}_k$; 95% confidence bounds $\approx \pm 1.96/\sqrt{T}$.

Theoretical ACF for AR(1): $\rho_k = \phi_1^k$ (geometric decay):



Dashed lines: $\pm 1.96/\sqrt{T}$ bounds ($T = 100$). Bars outside bounds are statistically significant.

Theoretical PACF for AR(2):



Cuts off after lag 2: AR(2) signature.

The **lag-k PACF** ϕ_{kk} is the correlation between y_t and y_{t-k} *after removing* the linear effects of $y_{t-1}, \ldots, y_{t-k+1}$.

For model identification:

- AR(p): $\phi_{kk} = 0$ for $k > p \Rightarrow$ PACF **cuts off** after lag p
- MA(q): $\phi_{kk} \to 0$ geometrically as $k \to \infty \Rightarrow$ PACF **tails off**

ACF identifies MA order (q); PACF identifies AR order (p). Together they form the Box-Jenkins identification toolkit.

Model	ACF	PACF	d
White noise	No significant spikes	No significant spikes	0
AR(p)	Tails off (decays)	Cuts off after p	0
MA(q)	Cuts off after q	Tails off (decays)	0
ARMA(p,q)	Tails off	Tails off	0
Random walk	Decays very slowly	Large spike at lag 1	1

Always unit-root test and difference *before* reading ACF/PACF. The ACF/PACF of a non-stationary series are not interpretable.

The ARIMA Model

ARIMA = differencing to achieve stationarity + ARMA on the result.

AR(p): $y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t$
MA(q): $y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$ *MA terms arise naturally from aggregation and averaging — notably, ARIMA(0,1,1) is algebraically equivalent to SES.*

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \qquad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

MA terms capture autocorrelation in *shocks*; AR terms capture autocorrelation in *levels*.

Special cases:
ARMA(p,0) = AR(p)
ARMA(0,q) = MA(q)

Requirements:
Stationarity: roots of AR polynomial outside unit circle
Invertibility: roots outside unit circle (unique MA representation; allows rewriting as AR(∞))

Apply ARMA(p,q) to the d-times differenced series $(1-B)^d y_t$:

$$\underbrace{\left(1-\phi_1 B-\cdots-\phi_p B^p\right)}_{\text{AR polynomial}}\underbrace{(1-B)^d}_{\text{differencing}} y_t = c+\underbrace{\left(1+\theta_1 B+\cdots+\theta_q B^q\right)}_{\text{MA polynomial}} \varepsilon_t$$

Parameter meaning:

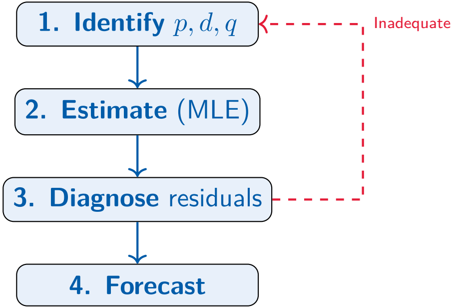
- p: AR order (lags of y)
- d: degree of differencing
- q: MA order (lags of ε)

Common special cases:

- ARIMA(0,1,0): random walk
- ARIMA(1,1,0): differenced AR(1)
- ARIMA(0,1,1): equivalent to SES (Hyndman and Athanasopoulos 2021, Ch. 9)

Example (ARIMA(1,1,0), $\hat{\phi}_1 = 0.5$, $y_{100} = 120$, $y_{99} = 118$): $\Delta y_{100} = 2 \Rightarrow \hat{y}_{101|100} = 120 + 0.5 \times 2 = \mathbf{121}$

2015



Step 1 (Identify): unit root tests $\to d$; ACF/PACF $\to p, q$
Step 3 (Diagnose): residual ACF should show no structure; Ljung-Box H_0: white noise (test at lag $L = \min(10, T/5)$; p-value < 0.05 signals residual autocorrelation — re-identify)

Multi-step forecasting uses the recursive substitution principle:

$$\hat{y}_{T+h|T} = \hat{c} + \sum_{i=1}^{p} \hat{\phi}_i \, \hat{y}_{T+h-i|T} + \sum_{j=1}^{q} \hat{\theta}_j \, \hat{\varepsilon}_{T+h-j|T}$$

where $\hat{y}_{T+k|T} = y_{T+k}$ for $k \leq 0$ and $\hat{\varepsilon}_{T+k|T} = 0$ for $k > 0$.

ARIMA forecast uncertainty **grows** with horizon h: prediction interval width scales as $\sigma\sqrt{h}$ for ARIMA(0,1,0). ARIMA(0,1,1) $\equiv$ SES shares this same growing uncertainty.

Socratic: ARIMA(0,1,1) produces the same point forecast as SES with $\hat{\alpha} = 1 - \hat{\theta}_1$. Why might their prediction intervals still differ?

Seasonal ARIMA

Retail and economic data have seasonal autocorrelation at lags m, $2m$, $3m$,...

Monthly retail sales have autocorrelation at lag 12, 24, ... — beyond standard ARIMA. Denote the AR and MA lag polynomials $\phi_p(B)=1-\phi_1 B-\cdots-\phi_p B^p$ and $\theta_q(B)=1+\theta_1 B+\cdots+\theta_q B^q$; SARIMA adds seasonal counterparts $\Phi_P(B^m)$, $\Theta_Q(B^m)$:

$$\underbrace{\Phi_P(B^m)}_{\text{seasonal AR}}\,\underbrace{\phi_p(B)}_{\text{AR}}\,\underbrace{(1-B^m)^D}_{\text{seas. diff.}}\,\underbrace{(1-B)^d}_{\text{diff.}}\,y_t \;=\; c+\underbrace{\Theta_Q(B^m)}_{\text{seasonal MA}}\,\underbrace{\theta_q(B)}_{\text{MA}}\,\varepsilon_t$$

Parameters:

- (p, d, q): non-seasonal orders
- (P, D, Q): seasonal orders at period m
- $m = 12$ for monthly; $m = 4$ for quarterly

SARIMA(1,1,1)(0,1,1)[12] is a common starting point for monthly retail series.

Manual identification via ACF/PACF is time-consuming and subjective. **Auto-ARIMA** automates the process:

1. Apply unit-root tests to select d (and D)
2. Search over a grid of (p, q, P, Q) values
3. Select the model minimizing AIC (or BIC)
4. Return selected model + fitted parameters

Python: `pmdarima.auto_arima()` with `seasonal=True`, `m=12` searches SARIMA models automatically. `statsmodels.tsa.statespace.sarimax.SARIMAX` fits any manually specified SARIMA.

AIC-selected models can differ from ACF/PACF-identified models. Both approaches should produce white-noise residuals — if they disagree strongly, investigate for outliers or structural breaks.

- **Stationarity** (constant mean, variance, and autocorrelation) is required by all classical forecasting models.
- **Unit root tests** (ADF, KPSS) determine the differencing order d before model fitting.
- **ACF and PACF** fingerprint the autocorrelation structure: AR cuts off in PACF; MA cuts off in ACF.
- **ARIMA(p,d,q)** combines differencing with an ARMA model — subsumes random walk, AR, MA, and SES as special cases.
- **SARIMA** adds seasonal polynomials for periodic autocorrelation at multiples of lag m.

ARIMA(0,1,1) $\equiv$ *SES and ARIMA(0,2,2)* $\equiv$ *Holt linear — the ARIMA family unifies exponential smoothing and classical ARMA models.*

We now have four model families in our toolkit:

- Benchmarks (naïve, seasonal naïve, mean, drift)
- Regression and AR models (Lecture 2)
- Exponential smoothing / ETS (Lecture 3)
- ARIMA / SARIMA (Lecture 4)

Key question: what if two series are *related* to each other?

Lecture 5: Multivariate forecasting — VAR models, ARIMAX, and Granger causality tests.

Lab 4: ADF/KPSS tests, ACF/PACF plots, manual ARIMA identification, and auto-ARIMA on RSXFS.

- Box, George E. P. et al. (2015). *Time Series Analysis: Forecasting and Control*. 5th. Hoboken, NJ: Wiley.
- Dickey, David A. and Wayne A. Fuller (1979). “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”. In: *Journal of the American Statistical Association* 74.366, pp. 427–431.
- Hyndman, Rob J. and George Athanasopoulos (2021). *Forecasting: Principles and Practice*. 3rd. Open access. Melbourne, Australia: OTexts. URL: `https://otexts.com/fpp3/`.
- Kwiatkowski, Denis et al. (1992). “Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root”. In: *Journal of Econometrics* 54.1–3, pp. 159–178.