

# VU LEARNING HUB

[vulearninghub.xyz](https://vulearninghub.xyz)

**HANDOUT**

## **STA621**

Time Series Analysis

**STA621 by VU Learning HUB**

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

VIRTUAL UNIVERSITY OF PAKISTAN

# STA621

## Time Series Analysis

---

### STUDY HANDOUT

Time Series Analysis studies data collected over time - stock prices, rainfall, sales figures. This handout covers the classical components of a time series, smoothing methods, stationarity, AR/MA/ARMA/ARIMA models, the Box-Jenkins methodology and forecasting, with formulas, key terms and exam pointers.

Compiled as a structured study resource for VU students.  
Covers the key topics, definitions, concepts and exam pointers of the course.

# Table of Contents

1. Introduction to Time Series
2. Smoothing and Moving Averages
3. Stationarity and Differencing
4. Autocorrelation: ACF and PACF
5. Autoregressive (AR) Models
6. Moving Average (MA) Models
7. ARMA and ARIMA Models
8. Box-Jenkins Methodology
9. Forecasting with ARIMA
10. Seasonal Models (SARIMA) and Applications

# Unit 1: Introduction to Time Series

## Learning Objectives

- Define a time series and give examples from economics, business and science
- Identify the four classical components: trend, seasonal, cyclical, irregular
- Distinguish additive from multiplicative decomposition models
- Explain the goals of time series analysis: description, explanation, forecasting

A time series is a sequence of observations recorded over time, usually at equally spaced intervals:  $X_1, X_2, \dots, X_t, \dots, X_n$ . Examples: monthly inflation rates, daily closing prices of a stock, annual wheat production, hourly temperature readings.

Time series analysis has three main goals: description (summarize the patterns in the data), explanation (understand what drives the movements), and forecasting (predict future values). In business, forecasting sales or demand is often the most valuable application.

The classical decomposition splits a series into four components: Trend (T) - the long-term direction (rising, falling, flat); Seasonal (S) - regular repeating patterns within a year (e.g. ice-cream sales peak every summer); Cyclical (C) - longer wave-like movements around the trend, often tied to business cycles of several years; Irregular (I) - the leftover random noise after the other components are removed.

In the additive model,  $X_t = T_t + S_t + C_t + I_t$  - appropriate when seasonal swings are roughly constant in size. In the multiplicative model,  $X_t = T_t \times S_t \times C_t \times I_t$  - appropriate when seasonal swings grow with the level of the series (common in economic data). Taking logs converts a multiplicative model into an additive one.

Plotting the series against time is always the first step. The time plot reveals trend, seasonality, outliers and structural breaks - features no formula should be trusted to find on its own.

## Key Terms

|                             |                                                                                      |
|-----------------------------|--------------------------------------------------------------------------------------|
| <b>Time series</b>          | Sequence of observations $X_1 \dots X_n$ recorded at (usually) equally spaced times. |
| <b>Trend (T)</b>            | Long-term direction of the series.                                                   |
| <b>Seasonal (S)</b>         | Regular repeating pattern within a fixed period (e.g. yearly).                       |
| <b>Cyclical (C)</b>         | Longer wave-like movement around trend, often business-cycle related.                |
| <b>Irregular (I)</b>        | Random leftover component after T, S, C are removed.                                 |
| <b>Additive model</b>       | $X_t = T_t + S_t + C_t + I_t$ ; constant-size seasonal swings.                       |
| <b>Multiplicative model</b> | $X_t = T_t \times S_t \times C_t \times I_t$ ; swings proportional to level.         |

Example - decomposition: Monthly electricity demand shows an upward trend (population growth), a seasonal peak every summer (air conditioning), a cyclical dip during a recession year, and irregular spikes during a heatwave week.

Example - multiplicative: Airline passenger numbers grow from 1M to 10M over 20 years; the summer peak grows from +200k to +2M - the seasonal swing is proportional to the level, so the multiplicative model fits.

## Summary

- ✓ A time series is data ordered in time:  $X_1, X_2, \dots, X_n$ .
- ✓ Four components: trend, seasonal, cyclical, irregular.
- ✓ Additive model for constant swings; multiplicative for proportional swings (logs make it additive).
- ✓ Always plot the series first; goals are description, explanation, forecasting.

## Exam Pointers

- ★ Define a time series and name its four classical components with one example each.
- ★ Distinguish the additive and multiplicative decomposition models.
- ★ A company's quarterly sales show peaks every Q4 that grow larger each year. Which model fits and why?

# Unit 2: Smoothing and Moving Averages

## Learning Objectives

- Explain the purpose of smoothing a time series
- Compute simple and centered moving averages
- Describe single exponential smoothing and the role of alpha
- Compare moving averages with exponential smoothing

Smoothing removes short-term noise to reveal the underlying trend and seasonal pattern. A smoothed series is easier to read, and smoothing is the basis of several forecasting methods.

A simple moving average of order  $m$  replaces each value by the average of itself and its neighbors. For odd  $m$ , the average is centered naturally:  $MA(3): S_t = (X_{t-1} + X_t + X_{t+1})/3$ . For even  $m$  (e.g. 4 for quarterly data, 12 for monthly), a centered moving average averages two consecutive moving averages so the result lines up with the time points.

Longer windows give smoother results but react more slowly to genuine changes, and values are lost at the start and end of the series. A 12-month centered moving average of monthly data removes seasonality and reveals the trend-cycle.

Single exponential smoothing produces the smoothed value  $S_t = \alpha * X_t + (1 - \alpha) * S_{t-1}$ , where  $0 < \alpha < 1$  is the smoothing constant. It is a weighted average giving the most weight to the most recent observation, with weights declining exponentially into

the past.

A large alpha (near 1) reacts quickly to new data but stays noisy; a small alpha gives a smooth series that lags behind turning points. Alpha is usually chosen (e.g. between 0.1 and 0.3) to minimize the forecast error on past data.

Exponential smoothing also gives a simple forecast: the forecast for the next period is just the current smoothed value,  $F_{t+1} = S_t$ . It works well for series with no trend and no seasonality; trend and seasonal versions (Holt's and Holt-Winters methods) extend it.

## Key Terms

|                                     |                                                                              |
|-------------------------------------|------------------------------------------------------------------------------|
| <b>Smoothing</b>                    | Removing short-term noise to reveal trend/pattern.                           |
| <b>Moving average<br/>MA(m)</b>     | Average of m neighboring values; centered version aligns with time points.   |
| <b>Exponential<br/>smoothing</b>    | $S_t = \alpha X_t + (1-\alpha)S_{t-1}$ ; exponentially declining weights.    |
| <b>Smoothing<br/>constant alpha</b> | $0 < \alpha < 1$ ; large = responsive but noisy, small = smooth but lagging. |
| <b>Holt's method</b>                | Exponential smoothing extended with a trend component.                       |
| <b>Holt-Winters</b>                 | Exponential smoothing with trend and seasonal components.                    |

Example - MA(3): Data: 10, 12, 14, 11, 13. MA(3) at  $t=2$ :  $(10+12+14)/3 = 12$ ; at  $t=3$ :  $(12+14+11)/3 = 12.33$ ; at  $t=4$ :  $(14+11+13)/3 = 12.67$ . The smoothed series (12, 12.33, 12.67) is calmer than the raw data.

Example - exponential smoothing:  $\alpha = 0.2$ ,  $S_1 = X_1 = 100$ .  $X_2 = 110 \rightarrow S_2 = 0.2*110 + 0.8*100 = 102$ .  $X_3 = 105 \rightarrow S_3 = 0.2*105 + 0.8*102 = 102.6$ . Forecast  $F_4 = 102.6$ .

## Summary

- ✓ Smoothing reveals trend by removing noise; moving averages and exponential smoothing are the basic tools.
- ✓ Center moving averages for even window sizes (4, 12).
- ✓ Exponential smoothing:  $S_t = \alpha X_t + (1-\alpha)S_{t-1}$ ; forecast = last smoothed value.
- ✓ Alpha trades responsiveness against smoothness.

## Exam Pointers

- ★ Compute a centered MA(4) for quarterly data and explain why centering is needed.
- ★ Write the exponential smoothing formula and explain the effect of alpha.
- ★ When would you prefer exponential smoothing over a moving average?

## Unit 3: Stationarity and Differencing

### Learning Objectives

- Define weak (covariance) stationarity and its three conditions
- Explain why stationarity matters for ARMA modeling
- Apply first and seasonal differencing to remove trend and seasonality
- Use the backward shift operator B and difference operator

A series is (weakly / covariance) stationary if: (1) the mean  $E(X_t) = \mu$  is constant over time; (2) the variance  $\text{Var}(X_t) = \sigma^2$  is constant; (3) the autocovariance  $\text{Cov}(X_t, X_{t+k}) = \gamma_k$  depends only on the lag  $k$ , not on  $t$ . In words: the statistical behavior does not change over time.

Stationarity matters because ARMA models assume it - their parameters describe a stable probabilistic mechanism. A trending or seasonal series violates the constant-mean condition, so it must be transformed to stationarity before modeling.

Differencing removes trend: the first difference is  $\nabla X_t = X_t - X_{t-1}$ . A random walk with drift becomes (approximately) stationary after one difference. If trend remains, difference again (second differencing), but over-differencing adds artificial noise, so difference only as much as needed.

Seasonal differencing removes seasonality: for monthly data with yearly seasonality,  $\nabla_{12} X_t = X_t - X_{t-12}$ . A series with both trend and seasonality may need both differences.

The backward shift operator B is defined by  $B X_t = X_{t-1}$ , so  $\nabla = 1 - B$  and the first difference is  $(1-B)X_t$ . Operator notation keeps ARIMA formulas compact: e.g. an ARIMA(1,1,0) is  $(1 - \phi B)(1 - B)X_t = e_t$ .

A quick stationarity check: the time plot should show constant mean and spread, and the sample ACF should die out quickly rather than decaying very slowly (slow decay signals non-stationarity).

### Key Terms

|                                  |                                                                         |
|----------------------------------|-------------------------------------------------------------------------|
| <b>Stationarity</b>              | Constant mean, constant variance, autocovariance depending only on lag. |
| <b>First difference</b>          | $\nabla X_t = X_t - X_{t-1}$ ; removes trend.                           |
| <b>Seasonal difference</b>       | $\nabla_s X_t = X_t - X_{t-s}$ ; removes period-s seasonality.          |
| <b>Backward shift operator B</b> | $B X_t = X_{t-1}$ ; $\nabla = 1 - B$ .                                  |
| <b>Over-differencing</b>         | Differencing more than needed; injects artificial noise.                |

Example:  $X_t = 5, 7, 10, 12, 15$  (trending). First differences: 2, 3, 2, 3 - now fluctuating around a constant mean of 2.5; the trend is gone.

Example - operator form: AR(1) on differenced data:  $(1 - 0.6B)(1 - B)X_t = e_t$  expands to  $X_t = 1.6 X_{t-1} - 0.6 X_{t-2} + e_t$ .

## Summary

- ✓ Stationary: constant mean, variance; autocovariance depends only on lag.
- ✓ ARMA needs stationarity; difference trending/seasonal data first.
- ✓ First difference kills trend; seasonal difference kills seasonality.
- ✓  $B X_t = X_{t-1}$ ; use operator notation for compact models.

## Exam Pointers

- ★ State the three conditions of weak stationarity.
- ★ Why must a series be stationary before fitting an ARMA model?
- ★ How do first and seasonal differencing work? Give the operator forms.

# Unit 4: Autocorrelation: ACF and PACF

## Learning Objectives

- Define the autocorrelation function (ACF) and partial autocorrelation (PACF)
- Read a correlogram to judge stationarity and model order
- State the ACF/PACF signatures of AR(p) and MA(q) processes
- Explain the standard error bands on a correlogram

The autocorrelation at lag  $k$ ,  $\rho_k$ , is the correlation between  $X_t$  and  $X_{t-k}$ :  $\rho_k = \gamma_k / \gamma_0$ . The plot of  $\rho_k$  against  $k$  is the correlogram (ACF plot) - the single most useful diagnostic in time series.

The partial autocorrelation (PACF) at lag  $k$  is the correlation between  $X_t$  and  $X_{t-k}$  after removing the linear effect of the in-between values  $X_{t-1} \dots X_{t-k+1}$ . It measures the 'direct' dependence at exactly lag  $k$ .

Under the null of white noise, sample autocorrelations are approximately normal with standard error  $1/\sqrt{n}$ ; values outside about  $\pm 2/\sqrt{n}$  are drawn as significant (the dashed bands on the correlogram).

Signatures: for an AR(p) process, the PACF cuts off (becomes insignificant) after lag  $p$ , while the ACF tails off gradually (exponential decay or damped sine). For an MA(q) process, the reverse: the ACF cuts off after lag  $q$ , while the PACF tails off. For ARMA, both tail off.

A very slowly decaying ACF (staying significant for many lags) is the classic symptom of non-stationarity - difference the series and look again.

## Key Terms

|                                |                                                            |
|--------------------------------|------------------------------------------------------------|
| <b>ACF <math>\rho_k</math></b> | Correlation between $X_t$ and $X_{t-k}$ ; the correlogram. |
|--------------------------------|------------------------------------------------------------|

|                    |                                                                              |
|--------------------|------------------------------------------------------------------------------|
| <b>PACF</b>        | Correlation between $X_t$ and $X_{t-k}$ with in-between lags partialled out. |
| <b>Cut-off</b>     | ACF/PACF becomes insignificant after some lag (signature of MA/AR order).    |
| <b>Tail-off</b>    | Gradual decay of ACF/PACF (signature of the 'other' component).              |
| <b>White noise</b> | Uncorrelated zero-mean series; all $\rho_k = 0$ for $k > 0$ .                |

Example - AR(1) signature: PACF significant only at lag 1 (value  $\sim 0.7$ ), ACF decaying 0.7, 0.49, 0.34, ... - the classic AR(1) fingerprint.

Example - MA(1) signature: ACF significant only at lag 1 (value  $\sim -0.4$ ), PACF alternating and decaying - the MA(1) fingerprint.

## Summary

- ✓ ACF: correlation with past values; PACF: direct effect at exactly lag  $k$ .
- ✓ AR( $p$ ): PACF cuts off after  $p$ , ACF tails off. MA( $q$ ): ACF cuts off after  $q$ , PACF tails off.
- ✓ Significance bands approx  $\pm 2/\sqrt{n}$ .
- ✓ Slow ACF decay = non-stationarity; difference first.

## Exam Pointers

- ★ Define ACF and PACF and explain how they differ.
- ★ How do you identify AR(1) vs MA(1) from a correlogram?
- ★ What does a very slowly decaying sample ACF tell you?

# Unit 5: Autoregressive (AR) Models

## Learning Objectives

- Write the AR( $p$ ) model and explain each term
- State the stationarity condition for AR(1) and AR(2)
- Derive the Yule-Walker equations for estimating AR parameters
- Compute simple AR(1) forecasts

In an autoregressive model of order  $p$ , AR( $p$ ), the current value is a linear regression on its own past:  $X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t$ , where  $e_t$  is white noise with variance  $\sigma^2$ . With zero mean ( $c = 0$ ):  $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + e_t$ .

The AR(1):  $X_t = \phi X_{t-1} + e_t$ . It is stationary when  $|\phi| < 1$ . Its ACF is  $\rho_k = \phi^k$  - exponential decay, and its variance is  $\sigma^2/(1-\phi^2)$ . Positive  $\phi$  gives persistent, smooth wandering; negative  $\phi$  gives rapid oscillation.

The AR(2):  $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t$  is stationary when the roots of  $1 - \phi_1 z - \phi_2 z^2 = 0$  lie outside the unit circle (equivalently:  $\phi_2 - \phi_1 < 1$ ,  $\phi_2 + \phi_1 < 1$ ,  $|\phi_2| < 1$ ). It

can produce pseudo-periodic (cyclical) behavior.

Parameters are estimated by the Yule-Walker equations: multiply the AR(p) equation by  $X_{t-k}$  and take expectations to get  $\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}$  for  $k = 1..p$  - p linear equations in the p unknowns  $\phi_i$ . In practice, least squares / maximum likelihood give similar, slightly better estimates.

Forecasting is easy: the 1-step forecast  $\hat{X}_{t+1} = \phi_1 X_t + \dots + \phi_p X_{t-p+1}$  (replace future e's by zero, future X's by their forecasts). Multi-step forecasts revert toward the mean - the 'forecast function' of a stationary AR dies out.

## Key Terms

|                                     |                                                                                                      |
|-------------------------------------|------------------------------------------------------------------------------------------------------|
| <b>AR(p)</b>                        | $X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + e_t$ ; regression on own past.                  |
| <b>Stationarity condition</b>       | Roots of $1 - \phi_1 z - \dots - \phi_p z^p = 0$ outside the unit circle;<br>AR(1): $ \phi_1  < 1$ . |
| <b>Yule-Walker equations</b>        | Moment equations $\gamma_k = \sum \phi_j \gamma_{k-j}$ used to estimate $\phi_i$ .                   |
| <b>White noise <math>e_t</math></b> | Zero-mean, constant variance $\sigma^2$ , uncorrelated shocks driving the model.                     |

Example - AR(1) forecast:  $X_t = 0.7 X_{t-1} + e_t$ ,  $X_{100} = 10$ . Forecast  $X_{101} = 0.7 \cdot 10 = 7$ ;  $X_{102} = 0.7 \cdot 7 = 4.9$  - reverting toward the zero mean.

Example - variance: AR(1) with  $\phi = 0.6$ ,  $\sigma^2 = 1$ :  $\text{Var}(X) = 1/(1-0.36) = 1.5625$ .

## Summary

- ✓ AR(p): current value regressed on p past values plus white noise.
- ✓ AR(1) stationary iff  $|\phi| < 1$ ;  $\text{ACF} = \phi^k$ .
- ✓ Yule-Walker equations estimate  $\phi$  from sample autocovariances.
- ✓ Forecasts revert to the mean; replace future shocks by zero.

## Exam Pointers

- ★ Write the AR(2) model and state its stationarity conditions.
- ★ Derive the Yule-Walker equations for AR(1).
- ★ For  $X_t = 0.8 X_{t-1} + e_t$  with  $X_{50} = 5$ , compute the 1-step and 2-step forecasts.

# Unit 6: Moving Average (MA) Models

## Learning Objectives

- Write the MA(q) model and explain each term
- State the invertibility condition for MA(1)
- Explain the duality between AR and MA models

- Compute the ACF of MA(1) and MA(2)

In a moving average model of order  $q$ ,  $MA(q)$ , the current value is a linear combination of current and past shocks:  $X_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$ . (Sign conventions vary; some texts use plus signs.)

The MA(1):  $X_t = e_t - \theta e_{t-1}$ . It is always stationary (finite sum of shocks), with variance  $\sigma^2(1+\theta^2)$  and ACF:  $\rho_1 = -\theta/(1+\theta^2)$ ,  $\rho_k = 0$  for  $k > 1$  - the cut-off signature.

An MA model is invertible if it can be rewritten as an infinite AR - i.e. the shocks can be recovered from the observed  $X$ 's. MA(1) is invertible when  $|\theta| < 1$ . Invertibility is the MA counterpart of stationarity, with the same root condition on the  $\theta$  polynomial.

There is a duality: a stationary AR( $p$ ) can be written as an infinite MA, and an invertible MA( $q$ ) as an infinite AR. In practice this means a high-order AR can approximate an MA and vice versa - which is why correlogram signatures (cut-off vs tail-off) guide the choice.

MA terms capture short-lived shock effects: a one-off event affects the series for exactly  $q$  periods then vanishes. AR terms capture persistent, decaying memory. Real data often needs both - the ARMA model.

## Key Terms

|                           |                                                                                                 |
|---------------------------|-------------------------------------------------------------------------------------------------|
| <b>MA(<math>q</math>)</b> | $X_t = \mu + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$ ; combination of recent shocks. |
| <b>Invertibility</b>      | MA can be expressed as infinite AR; MA(1) needs $ \theta  < 1$ .                                |
| <b>AR-MA duality</b>      | Stationary AR = infinite MA; invertible MA = infinite AR.                                       |
| <b>Shock persistence</b>  | MA: effect lasts exactly $q$ periods; AR: effect decays gradually.                              |

Example - MA(1) ACF:  $\theta = 0.5$ :  $\rho_1 = -0.5/1.25 = -0.4$ ,  $\rho_2 = \rho_3 = \dots = 0$ .

Example - interpretation: Daily sales = baseline + today's shock - 0.6 x yesterday's shock: a promotion's effect washes out after one day - classic MA(1) behavior.

## Summary

- ✓ MA( $q$ ): linear combination of current and  $q$  past white-noise shocks.
- ✓ Always stationary; invertible iff  $\theta$ -roots outside unit circle ( $|\theta| < 1$  for MA(1)).
- ✓ ACF cuts off after lag  $q$  - the MA fingerprint.
- ✓ MA = short memory of shocks; AR = long decaying memory.

## Exam Pointers

- ★ Write the MA(2) model and derive its autocovariances.
- ★ State and explain the invertibility condition for MA(1).
- ★ How does the correlogram distinguish MA(1) from AR(1)?

## Unit 7: ARMA and ARIMA Models

### Learning Objectives

- Write the ARMA(p,q) and ARIMA(p,d,q) models
- Explain the role of the differencing order d
- Use operator notation for compact model writing
- Explain when ARIMA is preferred over plain ARMA

The ARMA(p,q) combines both:  $X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$ . In operator form:  $\phi(B)X_t = c + \theta(B)e_t$ , where  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  and  $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ .

ARMA needs stationarity: the phi-roots outside the unit circle, and (for unique representation) the theta-roots outside it too, with no common roots. Both its ACF and PACF tail off - which is why ARMA orders are harder to identify than pure AR or MA.

The ARIMA(p,d,q) extends ARMA to non-stationary series: difference the data d times, then fit ARMA(p,q) to the differences. Operator form:  $\phi(B)(1-B)^d X_t = c + \theta(B)e_t$ . The 'I' stands for integrated - the original series is the 'integral' (cumulative sum) of a stationary ARMA.

Common cases: ARIMA(0,1,0) is the random walk  $X_t = X_{t-1} + e_t$  (with drift if  $c \neq 0$ ) - the classic stock-price model. ARIMA(0,1,1) is equivalent to simple exponential smoothing. ARIMA(1,1,0) is AR(1) on first differences.

Choose d as the smallest differencing that achieves stationarity (usually 0, 1, or 2 - check the ACF decay). Then choose p, q from the ACF/PACF of the differenced series, keeping the model parsimonious: fewer parameters forecast better out of sample.

### Key Terms

|                     |                                                                              |
|---------------------|------------------------------------------------------------------------------|
| <b>ARMA(p,q)</b>    | $\phi(B)X_t = c + \theta(B)e_t$ ; AR and MA combined on stationary data.     |
| <b>ARIMA(p,d,q)</b> | ARMA(p,q) on the d-times differenced series; I = integrated.                 |
| <b>Random walk</b>  | ARIMA(0,1,0): $X_t = X_{t-1} + e_t$ ; non-stationary, variance grows with t. |
| <b>Parsimony</b>    | Preferring the simplest adequate model for better forecasts.                 |

Example - operator expansion: ARIMA(1,1,1):  $(1-0.5B)(1-B)X_t = (1-0.3B)e_t$  expands to  $X_t = 1.5 X_{t-1} - 0.5 X_{t-2} + e_t - 0.3 e_{t-1}$ .

Example - random walk forecast:  $X_t = X_{t-1} + e_t$ ,  $X_{100} = 250$ . All future forecasts = 250 (flat line), but prediction intervals widen with the square root of the horizon.

### Summary

- ✓ ARMA(p,q):  $\phi(B)X_t = c + \theta(B)e_t$  on stationary data; both ACF and PACF tail off.
- ✓ ARIMA(p,d,q): difference d times first, then ARMA. I = integrated.

- ✓  $ARIMA(0,1,0)$  = random walk;  $ARIMA(0,1,1)$  = exponential smoothing.
- ✓ Keep  $d$  minimal and the model parsimonious.

## Exam Pointers

- ★ Write  $ARIMA(1,1,1)$  in full and in operator form.
- ★ Explain what the 'I' in ARIMA means.
- ★ How do you choose the differencing order  $d$  in practice?

## Unit 8: Box-Jenkins Methodology

### Learning Objectives

- List the three stages: identification, estimation, diagnostic checking
- Identify candidate models from ACF/PACF patterns
- Explain parameter estimation by least squares / maximum likelihood
- Check residuals and use information criteria (AIC/BIC) for model selection

The Box-Jenkins methodology is the standard iterative workflow for ARIMA modeling, with three stages: (1) Identification - choose tentative  $p$ ,  $d$ ,  $q$ ; (2) Estimation - fit the parameters; (3) Diagnostic checking - verify the model; repeat until adequate.

Identification: difference to stationarity (choose  $d$  from ACF decay), then match ACF/PACF signatures - PACF cut-off after  $p$  suggests  $AR(p)$ ; ACF cut-off after  $q$  suggests  $MA(q)$ ; both tailing off suggests ARMA. Keep candidate models simple ( $p+q$  small).

Estimation: parameters ( $\phi$ ,  $\theta$ ,  $\sigma^2$ ) are fitted by conditional least squares or maximum likelihood (ML). ML is the modern default; software reports estimates with standard errors, and parameters with  $|t| < 2$  are candidates for dropping.

Diagnostic checking: the residuals (one-step forecast errors) should look like white noise - plot their ACF (no significant spikes) and run the Ljung-Box test for overall whiteness. Also check residual normality roughly, and that no parameter is near the non-stationarity/non-invertibility boundary (redundant parameters).

Model selection: compare adequate models with AIC (Akaike) or BIC/SBC (Bayesian/Schwarz): lower is better; BIC penalizes parameters more heavily. Never choose by in-sample fit alone - overfitted models forecast poorly.

### Key Terms

|                            |                                                                   |
|----------------------------|-------------------------------------------------------------------|
| <b>Identification</b>      | Choosing tentative $p$ , $d$ , $q$ from plots and signatures.     |
| <b>Estimation</b>          | Fitting $\phi$ , $\theta$ by least squares or maximum likelihood. |
| <b>Diagnostic checking</b> | Verifying residuals are white noise; Ljung-Box test.              |

|                       |                                                                       |
|-----------------------|-----------------------------------------------------------------------|
| <b>AIC / BIC</b>      | Information criteria trading fit against complexity; lower is better. |
| <b>Ljung-Box test</b> | Tests whether residual autocorrelations are jointly zero.             |

Example - workflow: Monthly sales: ACF decays slowly -> first difference ( $d=1$ ). Differenced ACF cuts off after lag 1, PACF tails off -> try ARIMA(0,1,1). Residuals white, AIC lowest among candidates -> adopt.

Example - failure: Residual ACF shows a spike at lag 12 -> yearly seasonality missed -> add seasonal difference or seasonal terms (SARIMA).

## Summary

- ✓ Three stages: identify ( $p, d, q$  from ACF/PACF), estimate (ML), diagnose (white-noise residuals).
- ✓ Ljung-Box tests residual whiteness; AIC/BIC compare models.
- ✓ Iterate until residuals are white and the model is parsimonious.

## Exam Pointers

- ★ Describe the three stages of the Box-Jenkins methodology.
- ★ How do ACF/PACF signatures guide identification?
- ★ What should the residuals of an adequate model look like, and how do you test it?
- ★ When would you prefer BIC over AIC?

# Unit 9: Forecasting with ARIMA

## Learning Objectives

- Compute point forecasts from ARIMA models recursively
- Explain how prediction intervals are constructed and why they widen
- Distinguish in-sample fit from out-of-sample forecast evaluation
- Describe forecast updating as new data arrives

Point forecasts are computed recursively: replace future shocks  $e_{t+h}$  by zero and future  $X$ 's by their own forecasts. For AR(1):  $X_{t+1}^{\wedge} = \phi X_t$ ,  $X_{t+2}^{\wedge} = \phi X_{t+1}^{\wedge} = \phi^2 X_t$ , and so on. For ARIMA with  $d > 0$ , forecast the differenced series then 'integrate' (cumulate) back.

Forecasts from stationary models revert to the mean; forecasts from random-walk-type models stay flat at the last value. Seasonal models repeat the seasonal pattern. The forecast function eventually follows the deterministic part of the model.

Prediction intervals: the  $h$ -step forecast error variance grows with  $h$  (for AR(1):  $\sigma^2(1 + \phi^2 + \phi^4 + \dots + \phi^{2(h-1)})$ ). A 95% interval is roughly forecast  $\pm 2 \times$  standard error. Intervals widen with horizon - long-range forecasts are very uncertain, especially for non-stationary models where the width grows without bound.

Evaluation: judge models by out-of-sample forecast accuracy (hold out the last  $m$  observations, forecast them, compare), not in-sample fit. Common measures: MAE (mean absolute error), RMSE (root mean squared error), MAPE (mean absolute percentage error). Simpler models often win out-of-sample.

In practice, forecasts are updated as each new observation arrives: re-estimate (or just roll forward) and reforecast. Forecast errors should themselves be white noise - a pattern in the errors means the model missed something.

## Key Terms

|                                 |                                                                           |
|---------------------------------|---------------------------------------------------------------------------|
| <b>Point forecast</b>           | Single best-guess future value; future shocks set to zero.                |
| <b>Prediction interval</b>      | Forecast $\pm z * \text{s.e.}$ ; widens with horizon.                     |
| <b>Out-of-sample evaluation</b> | Testing forecasts on data withheld from fitting.                          |
| <b>MAE / RMSE / MAPE</b>        | Mean absolute, root-mean-squared, and percentage forecast error measures. |

Example - intervals: AR(1),  $\phi=0.7$ ,  $\sigma=2$ ,  $X_{100}=10$ . 1-step: forecast 7, s.e. 2  $\rightarrow$  95% interval (3, 11). 2-step: forecast 4.9, s.e.  $2 * \sqrt{1.49}=2.44 \rightarrow$  interval (0, 9.8).

Example - MAPE: Forecasts 102, 98 vs actuals 100, 100: errors 2%, 2%  $\rightarrow$  MAPE = 2%.

## Summary

- ✓ Recursive forecasts: future shocks = 0, future  $X$ 's = their forecasts.
- ✓ Stationary forecasts revert to mean; intervals widen with horizon.
- ✓ Evaluate out-of-sample with MAE/RMSE/MAPE.
- ✓ Update forecasts as new data arrives; errors should be white noise.

## Exam Pointers

- ★ Compute 1-step and 2-step forecasts for an AR(1) with  $\phi = 0.6$ ,  $X_{50} = 8$ .
- ★ Why do prediction intervals widen as the forecast horizon grows?
- ★ Distinguish in-sample fit from out-of-sample evaluation; name three error measures.

# Unit 10: Seasonal Models (SARIMA) and Applications

## Learning Objectives

- Write the seasonal ARIMA (SARIMA) model in operator form
- Identify seasonal signatures on the ACF/PACF
- Describe applications in economics, business and environmental science
- List common pitfalls: over-differencing, outliers, structural breaks

Many series have seasonality: monthly retail sales, quarterly GDP, daily website traffic. The  $SARIMA(p,d,q)(P,D,Q)_s$  model adds seasonal AR/MA/differencing at lag  $s$ :  $\Phi(B^s)(1-B^s)^D \phi(B)(1-B)^d X_t = \Theta(B^s)\theta(B) e_t$ . The classic 'airline model' is  $SARIMA(0,1,1)(0,1,1)_{12}$  - one non-seasonal and one seasonal difference plus two MA terms.

Seasonal signatures: the ACF/PACF show significant spikes at lags  $s, 2s, 3s, \dots$ . A seasonal AR(1) gives PACF cut-off after lag  $s$ ; a seasonal MA(1) gives ACF cut-off after lag  $s$ . Always check the seasonal lags - many beginners model only the short lags and leave seasonality in the residuals.

Applications: economics (inflation, GDP, unemployment forecasting for policy); business (demand planning, inventory, staffing - Holt-Winters and SARIMA are workhorses); finance (volatility modeling extends to GARCH); environment (rainfall, river flow, air quality); demography (population projections).

Pitfalls: over-differencing (differencing a stationary series injects MA noise); outliers (a single extreme value distorts ACF and estimates - detect and adjust); structural breaks (policy changes shift the level/trend - split the sample or add intervention terms); ignoring seasonality (the most common beginner error).

Modern practice complements Box-Jenkins with automatic selection (e.g. auto.arima using information criteria over a grid), but the diagnostic logic - stationarity, signatures, white-noise residuals - remains the same.

## Key Terms

|                                         |                                                                            |
|-----------------------------------------|----------------------------------------------------------------------------|
| <b>SARIMA(p,d,q)(P,D,Q)<sub>s</sub></b> | ARIMA with seasonal AR/MA/differencing at period $s$ .                     |
| <b>Airline model</b>                    | $SARIMA(0,1,1)(0,1,1)_{12}$ ; famous all-purpose seasonal model.           |
| <b>Seasonal signature</b>               | ACF/PACF spikes at lags $s, 2s, 3s, \dots$                                 |
| <b>Structural break</b>                 | Abrupt change in level/trend (e.g. policy change); breaks model stability. |

Example - airline model: Monthly international air passengers: first difference + seasonal (12) difference, MA(1) + seasonal MA(1). Fits the growing, seasonal series with just 3 parameters.

Example - break: A sales series jumps permanently after a 2023 rebranding; fitting one ARIMA across the break mis-estimates everything - model the two regimes separately or add a level-shift term.

## Summary

- ✓ SARIMA adds seasonal (P,D,Q) at lag  $s$ ; the airline model is the classic starter.
- ✓ Look for ACF/PACF spikes at seasonal lags  $s, 2s, \dots$
- ✓ Used in economics, business planning, finance, environment.
- ✓ Avoid over-differencing, untreated outliers, structural breaks, ignored seasonality.

## Exam Pointers

- ★ Write the SARIMA(0,1,1)(0,1,1)<sub>12</sub> model in operator and expanded form.
- ★ How do you detect seasonality from a correlogram?
- ★ Name three pitfalls in applied time series modeling and how to handle each.