Business Forecasting

ADIA Course

Day 3 – Session 1

Dr. Tanujit Chakraborty

Sorbonne University, Abu Dhabi, UAE

Sorbonne Centre for Artificial Intelligence, Paris, France

Research Areas: Time Series Forecasting, Machine Learning, Econometrics, Health Data Science

July 2, 2025

Auto-Regression Model

Auto-Regression Model

An autoregressive model (also called AR model) is used to model a future behavior for a time-ordered data, using data from past behaviors.

- Essentially, it is a linear regression analysis of a dependent variable using one or more variables(s) in a given time-series data: $Y_t = f(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$
- A natural starting point for forecasting model is to use past values of Y, that is, $Y_{t-1}, Y_{t-2}, \dots$ to predict Y_t.
- An auto-regression is a regression model in which Y_t is regressed against its own lagged values.
- The number of lags used as regressors is called the **order of auto-regression.**
  - In first order auto-regression (denoted as AR(*1*)), Y_t is regressed against Y_{t-1}
  - In p^{th} order auto-regression (denoted as AR(p)), Y_t is regressed against, $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$

Auto-Regression Model for Forecasting

SORBONNE UNIVERSITÉ

An **AutoRegressive model of order 1** (AR(1)) says: Today's value = ϕ×Yesterday's value + shock

More formally:

$$Y_t = \phi Y_{t-1} + \epsilon_t$$

Where:

- Y_t: value of the variable at time t (e.g., stock return, inflation, sales).
- ϕ: a parameter that tells us how much of yesterday's value spills into today.
- ϵ_t: a random shock (news, surprises, etc.) $\epsilon_t \sim \text{iid } N(0, \sigma^2)$ is white noise.

Example:

Suppose a stock's daily return depends slightly on the previous day:

$$R_t = 0.4 R_{t-1} + \epsilon_t$$

If yesterday's return was **2%**, then **today's expected return** is:

$$\mathbb{E}[R_t] = 0.4 \times 2\% = 0.8\%$$

p^{th} order Auto-regression Model

Formula: p^{th} Order Auto-regression Model

In general, the p^{th} order auto-regression model is defined as

$$Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + \varepsilon_t$$

where, $\beta_0, \beta_1, \ldots, \beta_p$ is called auto-regression coefficients and ε_t is the noise term or residue and in practice it is assumed to Gaussian white noise

Computing AR Coefficients

Method 1: Ordinary Least Squares (OLS)

$$Y_t = \phi Y_{t-1} + \epsilon_t, \quad t = 2, \ldots, T$$

Step-by-step derivation

1. Define the sum of squared residuals (SSR):

$$S(\phi) = \sum_{t=2}^{T} (Y_t - \phi Y_{t-1})^2$$

2. Differentiate $S(\phi)$ with respect to ϕ:

$$\frac{dS(\phi)}{d\phi} = \sum_{t=2}^{T} 2(Y_t - \phi Y_{t-1})(-Y_{t-1}) = -2 \sum_{t=2}^{T} Y_{t-1}(Y_t - \phi Y_{t-1})$$

Computing AR Coefficients

Method 1: Ordinary Least Squares (OLS)

3. Set the derivative to zero for minimization:

$$\sum_{t=2}^{T} Y_{t-1}(Y_t - \phi Y_{t-1}) = 0$$

4. Solve for ϕ:

$$\sum_{t=2}^{T} Y_{t-1} Y_t = \phi \sum_{t=2}^{T} Y_{t-1}^2 \quad \Rightarrow \quad \hat{\phi} = \frac{\sum_{t=2}^{T} Y_{t-1} Y_t}{\sum_{t=2}^{T} Y_{t-1}^2}$$

5. Estimate residual variance:

$$\hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=2}^{T} (Y_t - \hat{\phi} Y_{t-1})^2$$

Computing AR Coefficients

Computing AR(p) model

- A number of techniques known for computing the AR coefficients
- The most common method is called Least Squares Method (LSM)
- The LSM is based upon the **Yule-Walker equations**

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 & \cdots & \cdots & \rho_{p-2} & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \rho_2 & \rho_3 & \cdots & \cdots & \rho_{p-3} & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 & \cdots & \cdots & \rho_{p-4} & \rho_{p-3} \\ \rho_3 & \rho_2 & \rho_1 & 1 & \rho_1 & \cdots & \cdots & \rho_{p-5} & \rho_{p-4} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \rho_{p-4} & \rho_{p-5} & \cdots & \cdots & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \vdots \\ \beta_{p-1} \\ \beta_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \vdots \\ \rho_{p-1} \\ \rho_p \end{bmatrix}$$

AR Model by Hand

Here is a small dataset of 5 time points used for manually fitting the AR(1) model.

	Time	Y
1	1	1.2
2	2	1.8
3	3	2.4
4	4	2.1
5	5	1.9

AR Model by Hand

1. **"Step-by-step AR(1) Estimation (Small Data)"** – This table shows:
   - Y_{t-1} (lagged values)
   - Y_t (current values)
   - Predicted Y_t using $\hat{\phi}Y_{t-1}$
   - Residuals $= Y_t - \hat{\phi}Y_{t-1}$
2. **"Estimated Parameters (Small Data)"** – This table displays:
   - Estimated $\hat{\phi} \approx 1.044$
   - Estimated $\hat{\sigma} \approx 0.453$

Y_{t-1}	Y_t	Predicted_	Residual
1.2	1.8	1.25333	0.54667
1.8	2.4	1.88	0.52
2.4	2.1	2.50667	-0.40667
2.1	1.9	2.19333	-0.29333

These values are derived manually (without libraries) using the formulas:

- $\hat{\phi} = \frac{\sum Y_{t-1}Y_t}{\sum Y_{t-1}^2}$
- $\hat{\sigma}^2 = \frac{1}{T-1}\sum(Y_t - \hat{\phi}Y_{t-1})^2$

Applications in Business and Finance

In Finance:

- **Interest rates** tend to follow an AR(1) pattern — central banks move gradually.
- **Volatility forecasting** uses AR-type models in GARCH models.
- **Stock returns** may have slight autocorrelation due to investor herding.

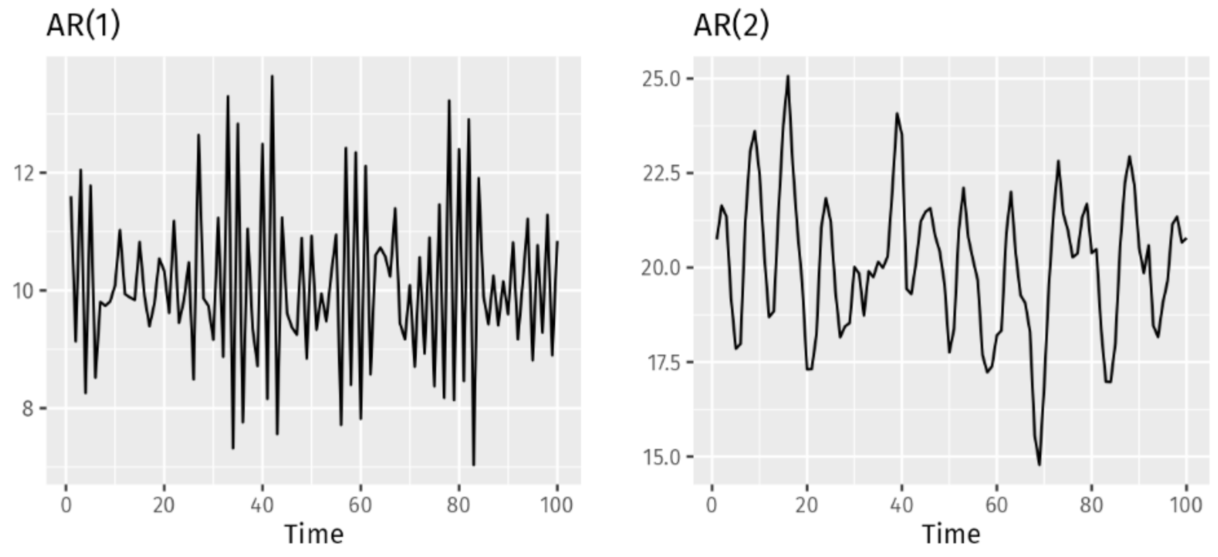


Figure 8.5: Two examples of data from autoregressive models with different parameters. Left: AR(1) with $y_t = 18 - 0.8y_{t-1} + \varepsilon_t$. Right: AR(2) with $y_t = 8 + 1.3y_{t-1} - 0.7y_{t-2} + \varepsilon_t$. In both cases, ε_t is normally distributed white noise with mean zero and variance one.

In Business:

- **Monthly sales**, if not seasonally adjusted, often follow AR(1)-like behavior.
- **Customer demand** today may depend on yesterday's trend.

Applications in Business and Finance

Application	What AR Model Helps With
Inventory management	Forecast next week's demand
Revenue forecasting	Predict next quarter's sales
Risk management	Model interest rate paths
Trading strategies	Test autocorrelation of returns

Key Takeaways:

- AR models use past information — simple yet powerful.
- AR(1) means "today is a shadow of yesterday, plus some new noise."
- Widely used in finance (returns, rates) and business (sales, demand).
- Easy to implement, interpret, and extend.

ARIMA Model

ARIMA

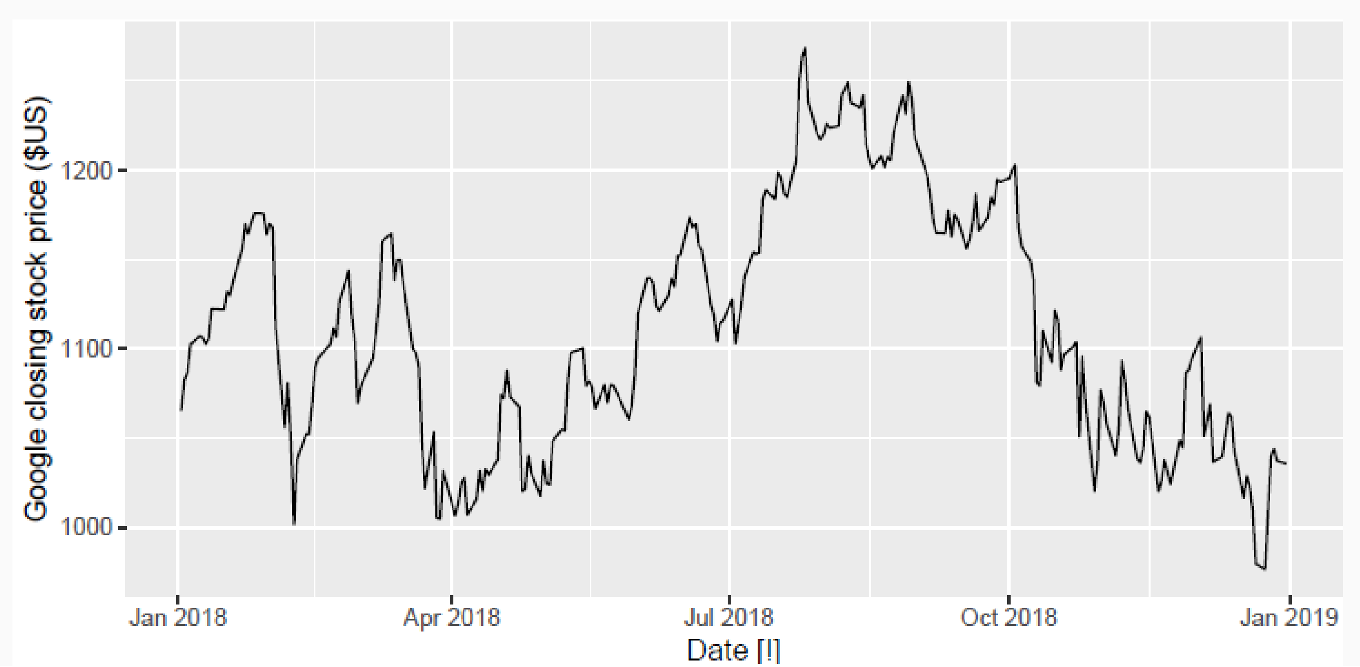
- **AR:** autoregressive (lagged observations as inputs)
- **I:** integrated (differencing to make series stationary)
- **MA:** moving average (lagged errors as inputs)

An ARIMA model is rarely interpretable in terms of visible data structures like trend and seasonality. But it can capture a huge range of time series patterns.

Stationary

Definition: If $\{y_t\}$ is a stationary time series, then for all s, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on t.

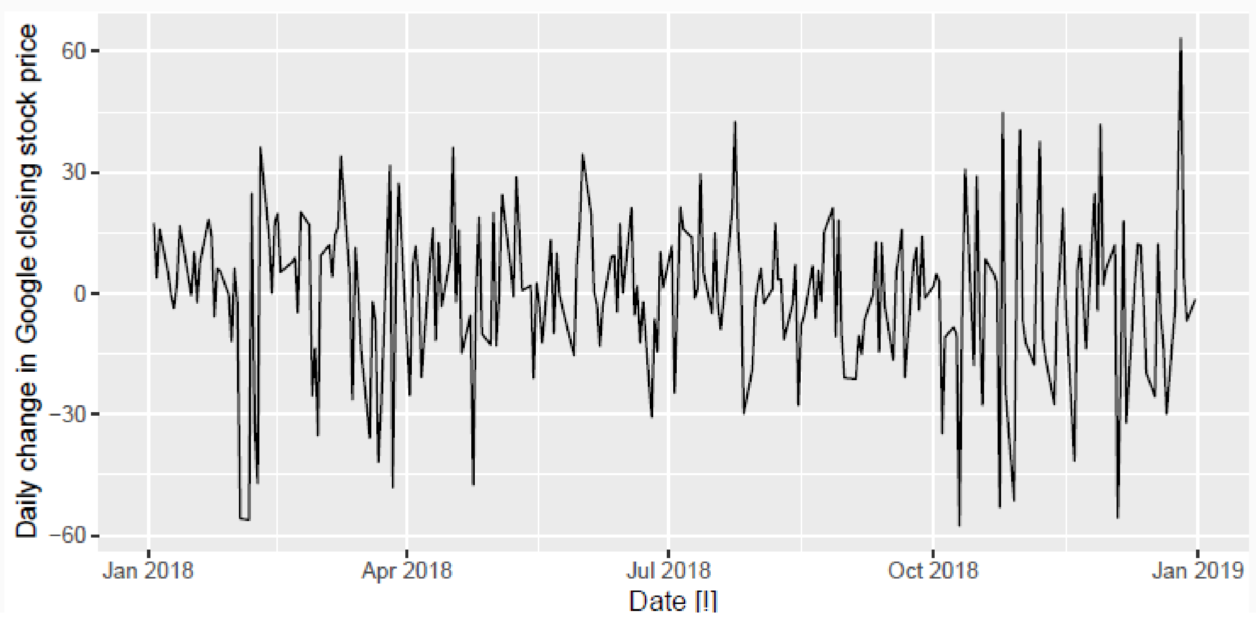
- A stationary series is:
- roughly horizontal
- constant variance
- no patterns predictable in the long-term



Daily close price of Google Stock during Jan – Dec 2018.

Differencing

- Differencing helps to stabilize the mean.
- The differenced series is the change between each observation in the original series.
- Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time.
- In practice, it is almost never necessary to go beyond second-order differences.



Daily change in close price of Google Stock during Jan – Dec 2018.

ARIMA models

Autoregressive Moving Average models:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

- Predictors include both lagged values of y_t and lagged errors.
- Combine ARMA model with differencing.
- d-differenced series follows an ARMA model.
- Need to choose p, d, q and whether or not to include c.

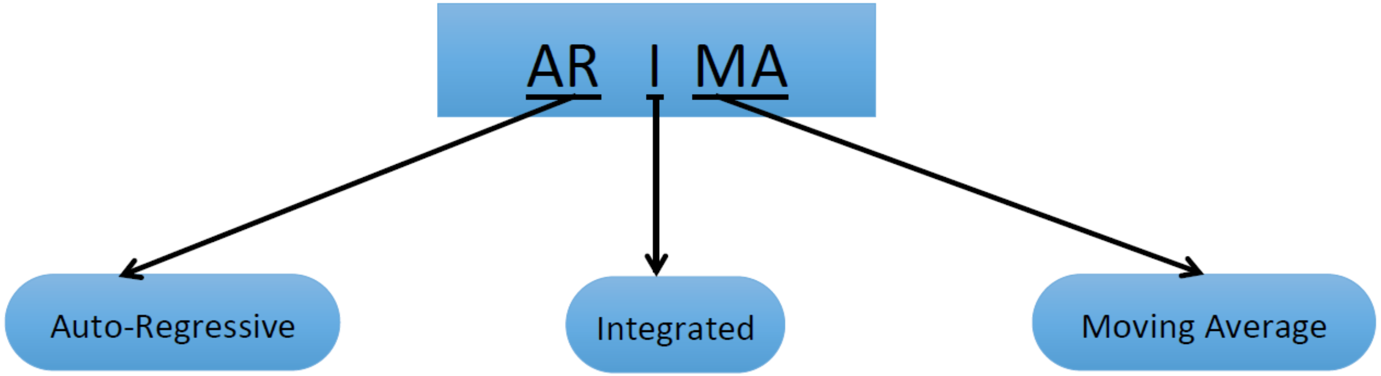
ARIMA models

ARIMA(p, d, q) model

AR: p = order of the autoregressive part

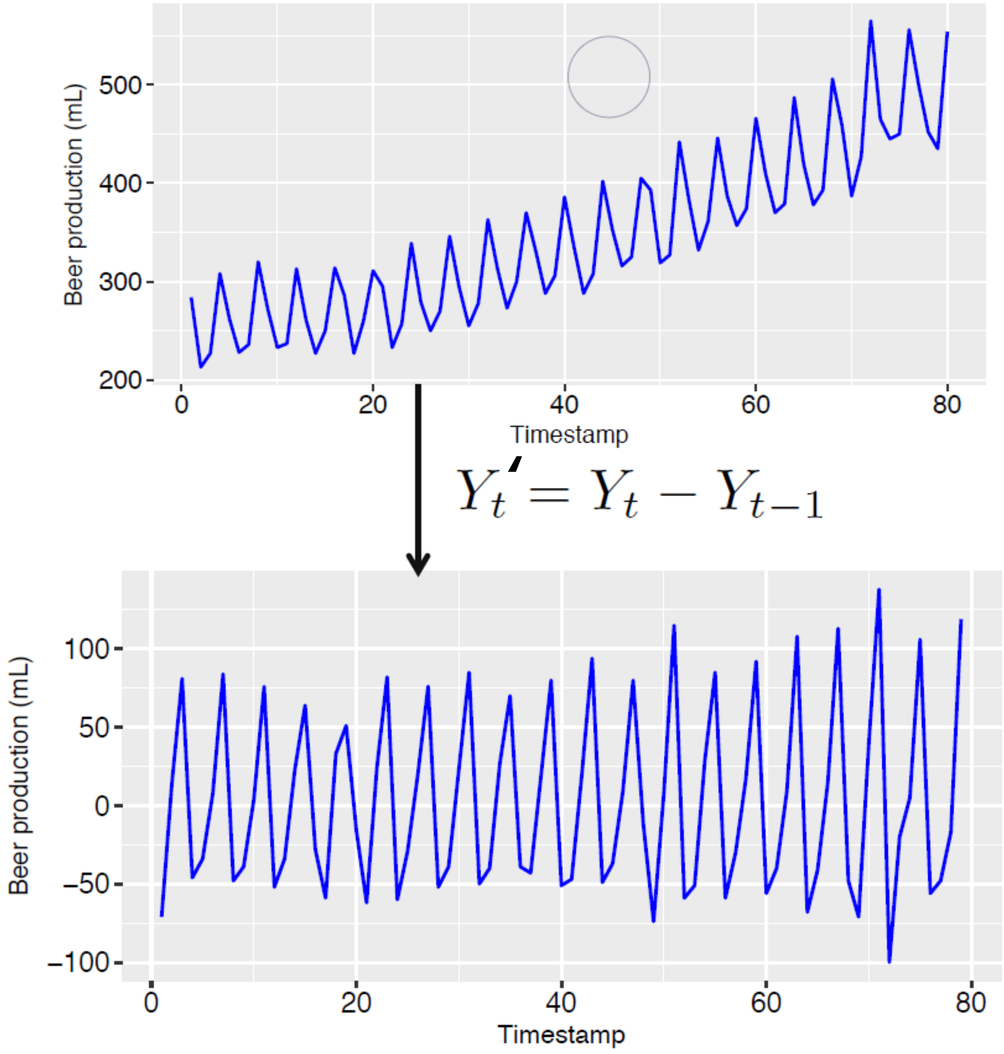
I: d = degree of first differencing involved

MA: q = order of the moving average part.



- White noise model: ARIMA(0,0,0)
- Random walk: ARIMA(0,1,0) with no constant
- Random walk with drift: ARIMA(0,1,0) with const.
- AR(p): ARIMA(p,0,0)
- MA(q): ARIMA(0,0, q)

How to select d?

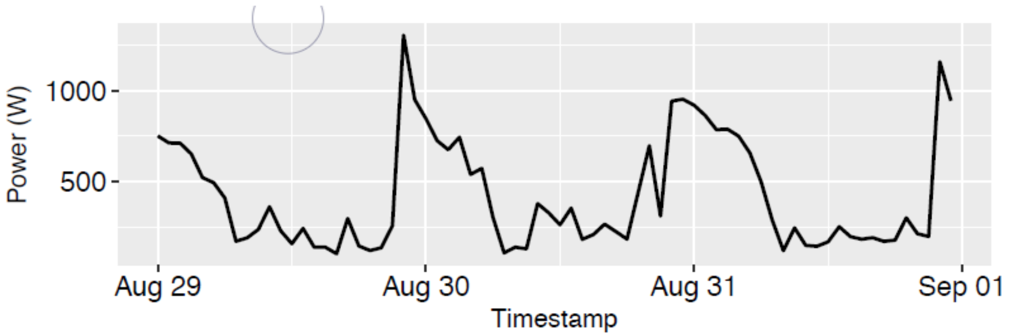


Differencing order (*d*):
Number of times differencing is done

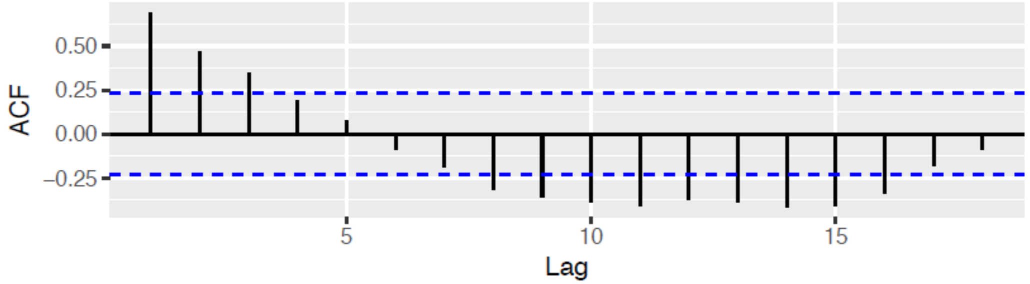
Selecting p and q manually

1. Auto-Correlation Function (ACF) Plot:
    - Correlation coefficients of time-series at different lags
    - Defines q order of MA model
2. Partial Auto-correlation Function (PACF) Plot:
    - Partial correlation coefficients of time series at different lags
    - Defines p order of AR model

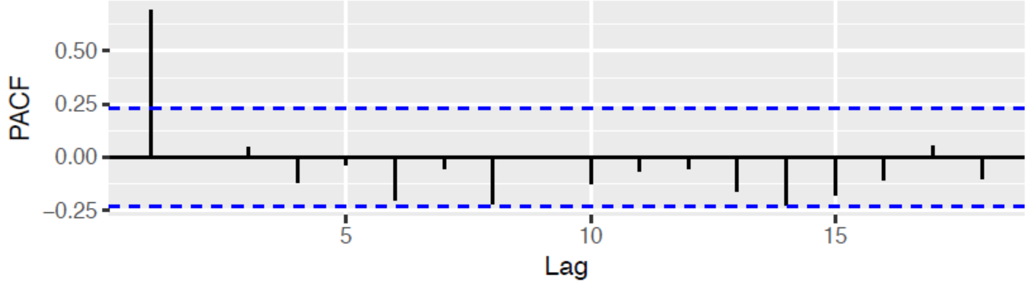
ACF / PACF Plots : Example



Data



ACF Plot

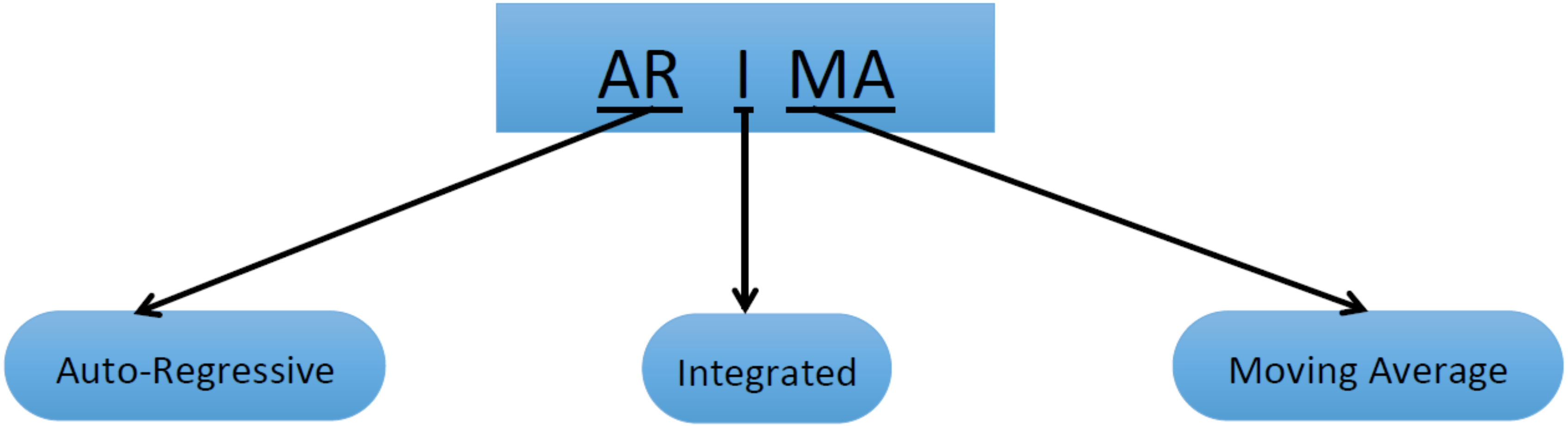


PACF Plot

Understanding ARIMA models

- If $c = 0$ and $d = 0$, the long-term forecasts will go to zero.
- If $c = 0$ and $d = 1$, the long-term forecasts will go to a non-zero constant.
- If $c = 0$ and $d = 2$, the long-term forecasts will follow a straight line.
- If $c = 0$ and $d = 0$, the long-term forecasts will go to the mean of the data.
- If $c = 0$ and $d = 1$, the long-term forecasts will follow a straight line.
- If $c = 0$ and $d = 2$, the long-term forecasts will follow a quadratic trend.

ARIMA models



ARIMA(p, d, q) model

AR: p = order of the autoregressive part

I: d = degree of first differencing involved

MA: q = order of the moving average part.

How does ARIMA select parameters?

- Select no. differences d via differencing tests, such as Augmented Dickey Fuller (ADF) test.
- Select p, q and inclusion of c by minimizing Corrected Akaike's Information Criterion (AICc).
- Use stepwise search to traverse model space.

$$\text{AICc} = -2\log(L) + 2(p+q+k+1)\left[1 + \frac{(p+q+k+2)}{T-p-q-k-2}\right]$$

where L is the maximised likelihood fitted to the differenced data, $k = 1$ if $c \neq 0$ and $k = 0$ otherwise.

Note: Can't compare AICc for different values of d.

How does ARIMA work?

Step1: Select current model (with smallest AICc) from:

$\text{ARIMA}(2, d, 2)$

$\text{ARIMA}(0, d, 0)$

$\text{ARIMA}(1, d, 0)$

$\text{ARIMA}(0, d, 1)$

Step 2: Consider variations of current model:

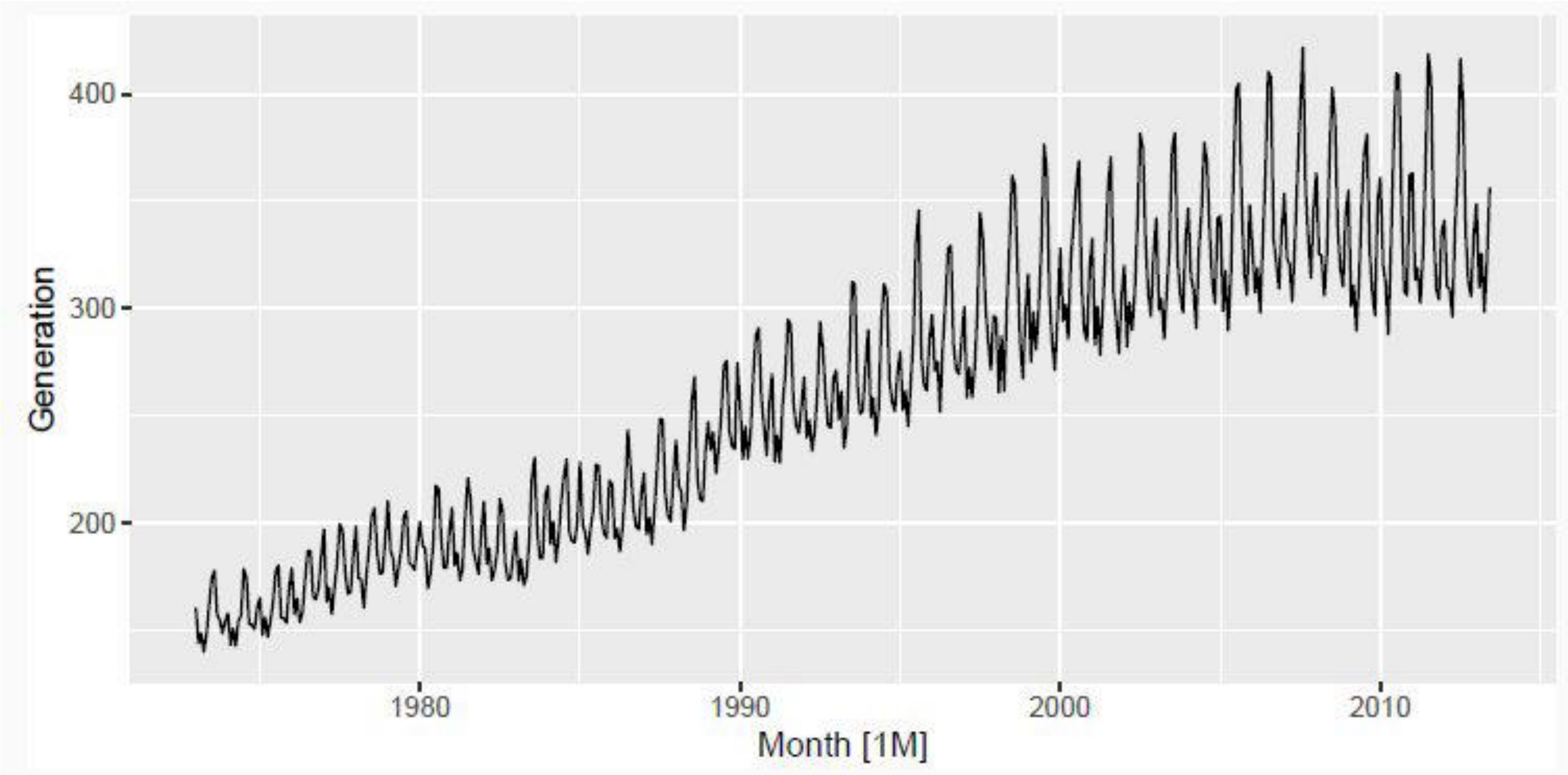
- vary one of p, q, from current model by ± 1;
- p, q both vary from current model by ± 1;
- Include/exclude c from current model.

Model with lowest AICc becomes current model.

Repeat Step 2 until no lower AICc can be found.

Seasonal ARIMA

- Many real-world time series, like temperature, sales, rainfall, electricity consumption, exhibit seasonal patterns that repeat at regular intervals. ARIMA alone cannot handle such seasonal effects effectively.



Monthly US electricity generation data from 1972 to 2013

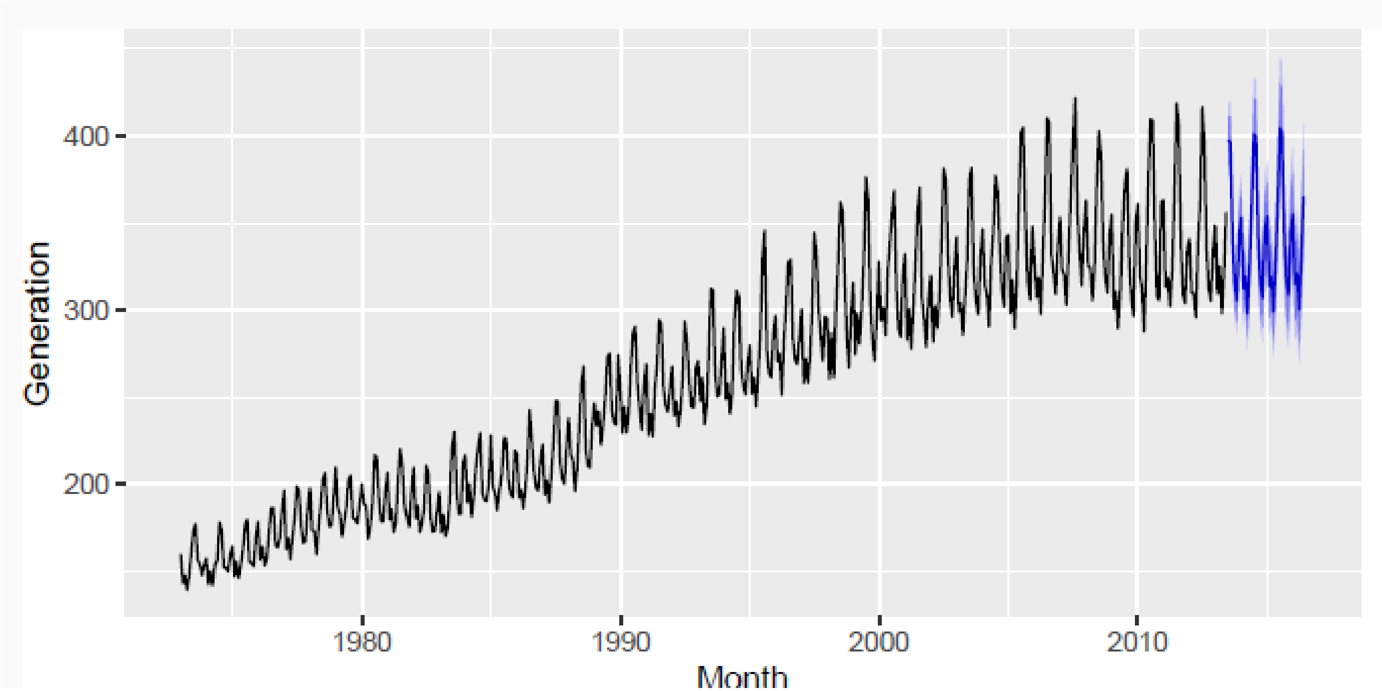
Seasonal ARIMA

ARIMA	$\underbrace{(p, d, q)}$	$\underbrace{(P, D, Q)_m}$
	$\uparrow$ Non-seasonal part of the model	$\uparrow$ Seasonal part of of the model

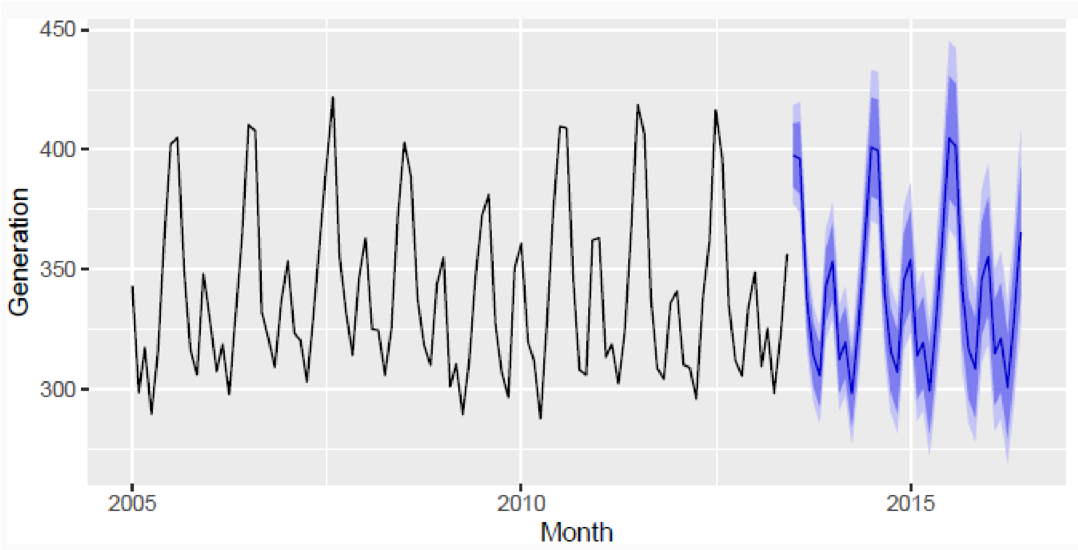
- m = number of observations per year.
- d first differences, D seasonal differences
- p AR lags, q MA lags
- P seasonal AR lags, Q seasonal MA lags

Seasonal and non-seasonal terms combine multiplicatively

Seasonal ARIMA Model Forecast



Monthly US electricity generation forecast for 2014 - 2016

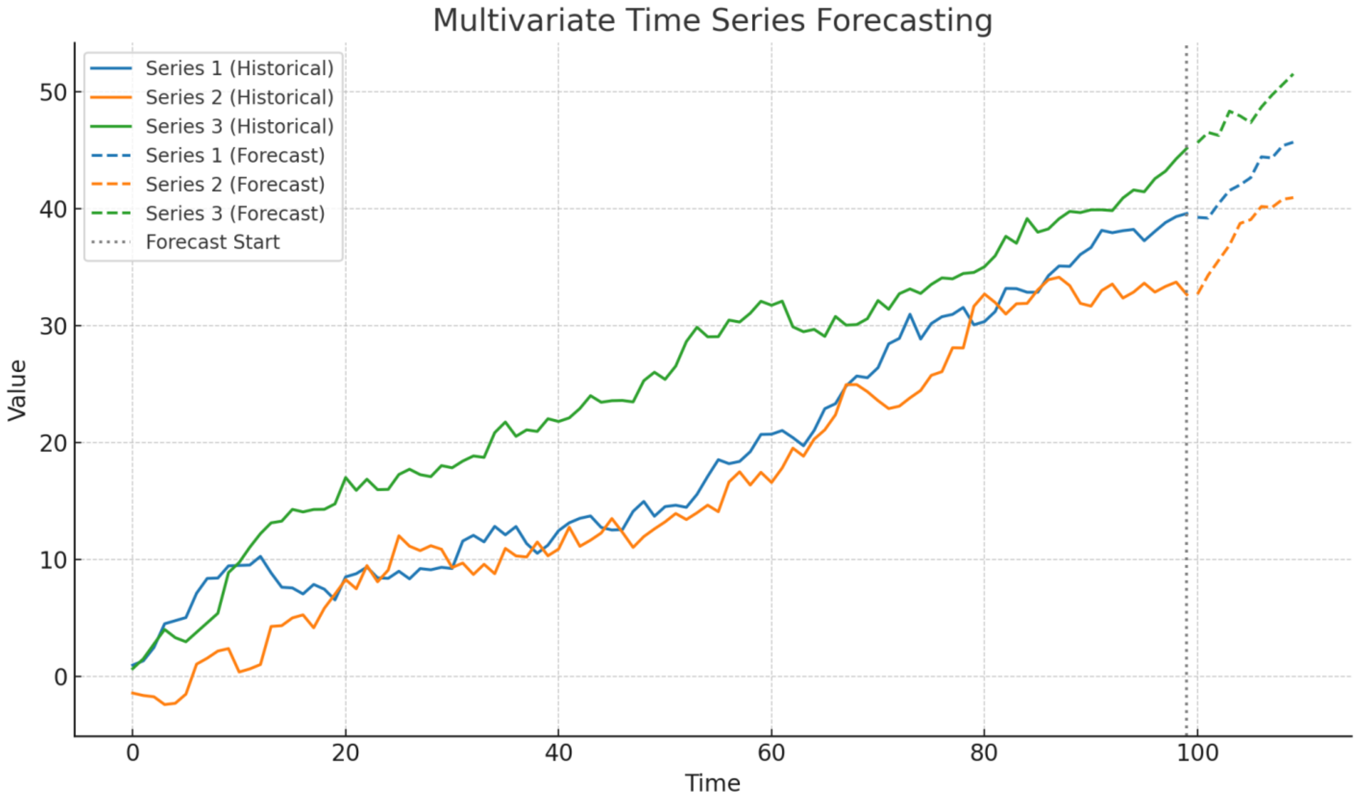


Forecast horizon zoomed in

Multivariate Forecasting

Multivariate Forecasting

- Multivariate Forecasting involves predicting future values of multiple interrelated time series simultaneously, capturing both temporal dynamics and cross-variable dependencies.



Multivariate Time Series Forecasting with three interdependent variables and their future projections

Multivariate Forecasting

- Key Features:
  a) *Joint Modeling:* Considers interactions among variables (e.g., temperature ↔ humidity ↔ rainfall)
  b) *More Informative:* Leverages additional contextual data for improved accuracy
  c) *Causal Relationships:* Helps uncover how one variable affects another over time
- Applications:
  a) Finance: Forecasting multiple asset prices
  b) Climate Science: Modeling interrelated climate indicators
  c) Healthcare: Predicting vitals in ICU patients
  d) Industry: Monitoring sensor networks in manufacturing

VAR

Vector AutoRegression (VAR) is a powerful statistical model used to capture the linear interdependencies among multiple time series.

- Developed by: Christopher Sims (1980)
- Used in: Macroeconomics, finance, climate modeling, and more
- Key Idea: Each variable is modeled as a function of its own past values and the past values of all other variables in the system.

Why to use VAR?

- Handles Multivariate Time Series: Captures dynamic relationships across several variables.
- No Need for Strong Theoretical Structure: Unlike structural models, VAR is largely data-driven.
- Great for Forecasting: Effective in short-term prediction of systems with interacting variables.
- Useful in Causality Analysis: Enables tools like Granger causality tests and impulse response functions (IRFs).

VAR(p)

- VAR models (vector autoregressive models) are used for multivariate time series. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables.
- As an example suppose that we measure three different time series variables, denoted by $x_{(t,1)}$, $x_{(t,2)}$, and $x_{(t,3)}$.
- The vector autoregressive model of order 1, denoted as VAR(1), is as follows:

$$x_{t,1} = \alpha_1 + \phi_{11}x_{t-1,1} + \phi_{12}x_{t-1,2} + \phi_{13}x_{t-1,3} + w_{t,1}$$

$$x_{t,2} = \alpha_2 + \phi_{21}x_{t-1,1} + \phi_{22}x_{t-1,2} + \phi_{23}x_{t-1,3} + w_{t,2}$$

$$x_{t,3} = \alpha_3 + \phi_{31}x_{t-1,1} + \phi_{32}x_{t-1,2} + \phi_{33}x_{t-1,3} + w_{t,3}$$

- Each variable is a linear function of the lag 1 values for all variables in the set.

VAR(p)

- In a VAR(2) model, the lag 2 values for all variables are added to the right sides of the equations, In the case of three x-variables (or time series) there would be six predictors on the right side of each equation, three lag 1 terms and three lag 2 terms.
- In general, for a VAR(p) model, the first p lags of each variable in the system would be used as regression predictors for each variable.
- VAR models are a specific case of more general VARMA models. VARMA models for multivariate time series include the VAR structure above along with moving average terms for each variable. More generally yet, these are special cases of ARMAX models that allow for the addition of other predictors that are outside the multivariate set of principal interest.

  It arose from macroeconomic data where large changes in the data permanently affect the level of the series. $\mathbf{x}_t = \phi \mathbf{x}_{t-1} + \mathbf{w}_t$