ARMA and ARIMA

LECTURE OUTLINE

Last time we studied the following topics:

- Auto Regressive Models (AR)
- Moving Average Models (AR)

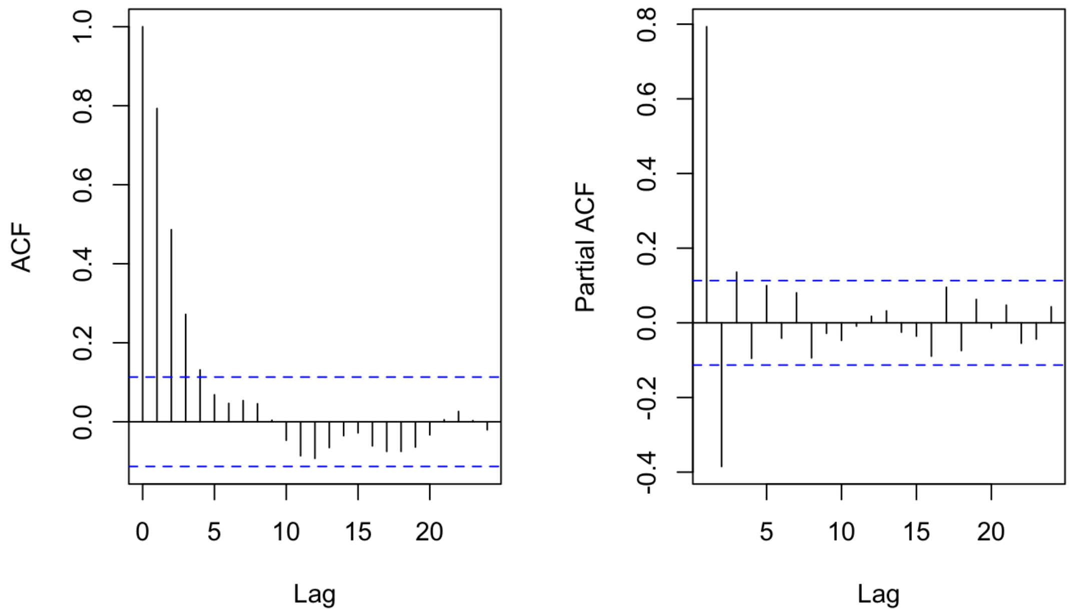
In this lecture we will study:

- ARMA, ARIMA
- Impulse Response Functions

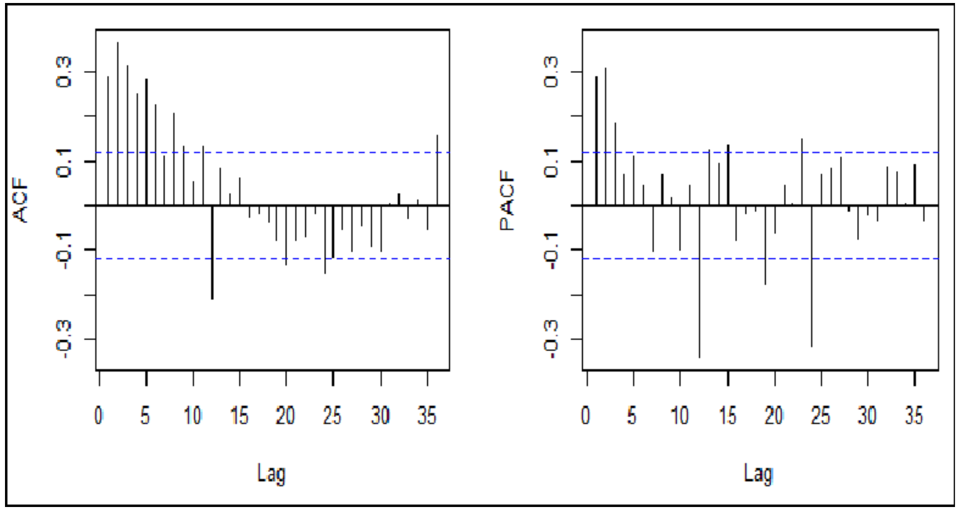
ARMA Models

When to use AR or MA?

- We used PACF to decide the order for AR
- We used ACF instead for MA



Easy



Harder

Kind of plot	AR(p)	MA(q)	ARMA
ACF behavior	Falls off slowly	Sharp drop after lag $= q$	No sharp cutoff
PACF behavior	Sharp drop after lag $= p$	Falls off slowly	No sharp cutoff

Introduction to ARMA

- Applied in the case that neither AR nor MA terms alone sufficiently describe the empirical dynamics

- An ARMA model can be written this way:

Number of autoregressive terms

ARMA(p,q) $\leftarrow$ $\boxed{y_t = \phi_0 + \sum_i^p \phi_i y_{t-i} + \epsilon_t + \sum_j^q \theta_j \epsilon_{t-j}}$

Number of error terms

So an ARMA(1,1) model is written: $y_t = \phi_0 + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$

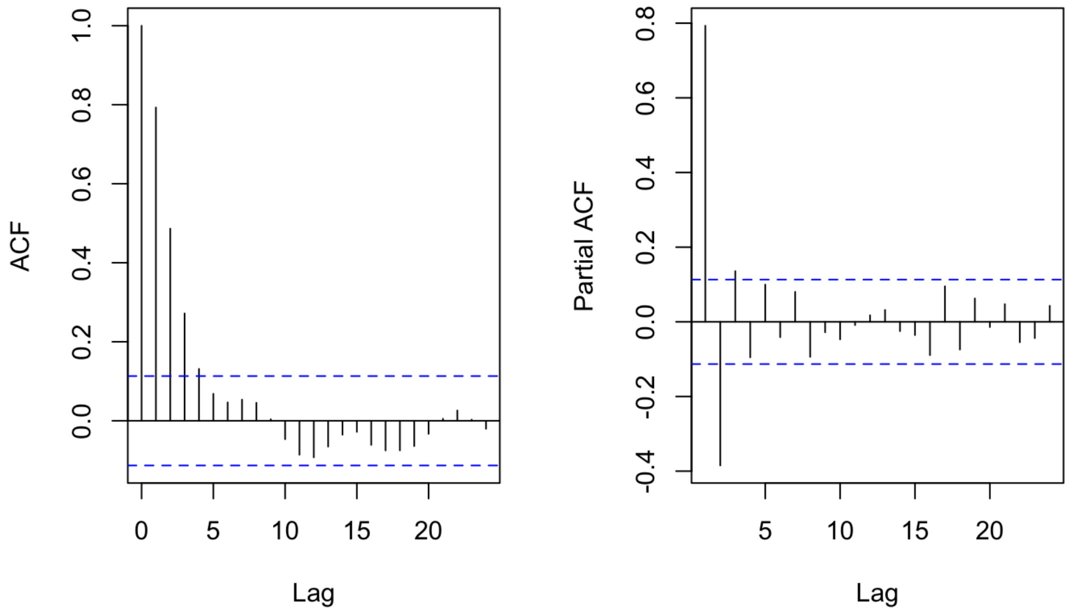
The prediction is written: $\hat{y}_t = \phi_0 + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1}$

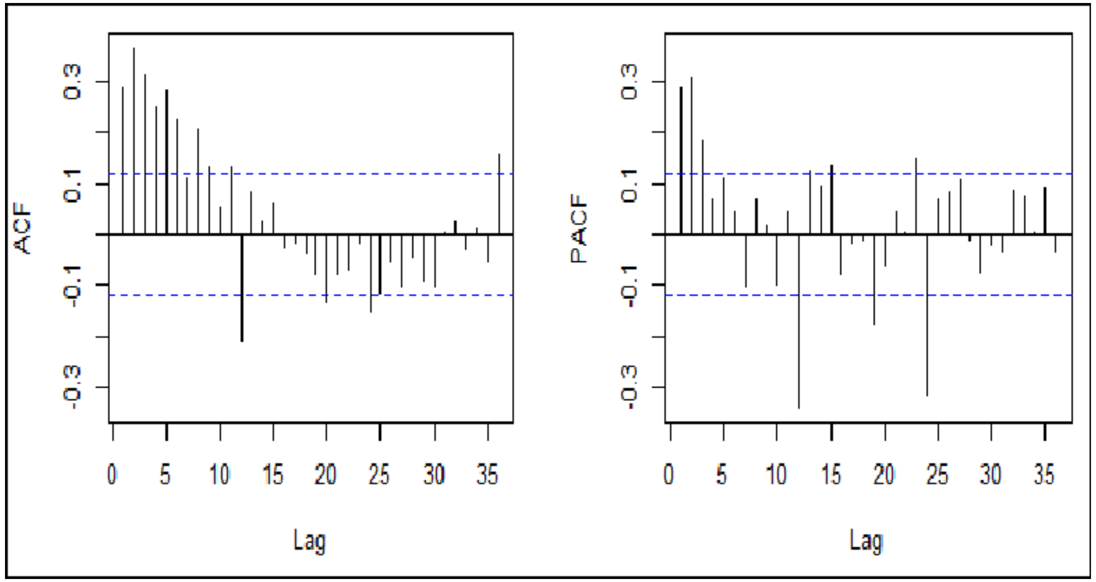
Where is ϵ_t ?

How to choose ARMA order

- Use PACF to decide the order for the AR part
- and ACF to decide the order of MA part

What would be the orders in these examples?





The orders seem quite high, can we simplify it?

Backshift Operator

The backshift operator, a.k.a. the lag operator, operates on time series points and moves them back by one time step each time it is applied:

$$L^k y_t = y_{t-k}$$

Helps to simplify the expression of time series models.
→ For example, an MA model can be rewritten as:

$$y_y = \mu + (1 + \theta_1 \times L + \theta_2 \times L^2 + \cdots + \theta_q \times L^q)e_t$$

Example of Backshift Operator on ARMA

One way of expressing an ARMA model is to put the AR components on one side and the MA components on another side. Suppose we have the following ARMA model:

$$y_t = 0.5 \times y_{t-1} + 0.24 \times y_{t-2} + e_t + 0.6 \times e_{t-1} + 0.09 \times e_{t-2}$$

$$y_t - 0.5 \times y_{t-1} - 0.24 \times y_{t-2} = e_t + 0.6 \times e_{t-1} + 0.09 \times e_{t-2}$$

We then re-express this using the lag operator. So, for example, y_{t-2} is re- expressed as $L^2 y$

$$y_t - 0.5 \times L \times y_t - 0.24 \times L^2 \times y_t = e_t + 0.6 \times L \times e_t + 0.09 \times L^2 \times e_t$$

So let's factorize these equations. for the left one, we have $\Delta = 1.21; x_1 = 1.25; x_2 = -3.33$
For the right one we have $\Delta = 0; x = -3.33$

$$0.24(L + 3.33)(L - 1.25) = 0.09(L + 3.33)^2$$

$$\longrightarrow \quad 0.24(L - 1.25)y_t = 0.09(L + 3.33)e_t$$

Post factoring, the order of the overall model has been reduced and now involves only values at time t − 1 rather than also values at t − 2.

Real world application

In real-world data sets you may end up with something such that the two sides are quite similar numerically, but not exact.

$$0.24(L+3.33)(L-1.25) = 0.09(L+3.33)^2$$

L+3.28

In such cases it is worth considering that this is close enough to a common factor to drop for model complexity.

ARMA stationarity

For an ARMA(p, q) model, stationarity is primarily determined by the autoregressive (AR) component.

Recall that in the last lecture we said that MA is stationary by design

So it is only necessary to check whether the AR part of the ARMA process is stationary.

Reminder: The characteristic polynomial associated with the AR(p) process is:

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

For the AR(p) process to be weakly stationary, the following condition must hold:

- All roots of this polynomial must lie outside the unit circle in the complex plane.

→ This means the magnitude of each root z_i must satisfy $|z_i| > 1$.

Impulse Response Functions

Simo Alami

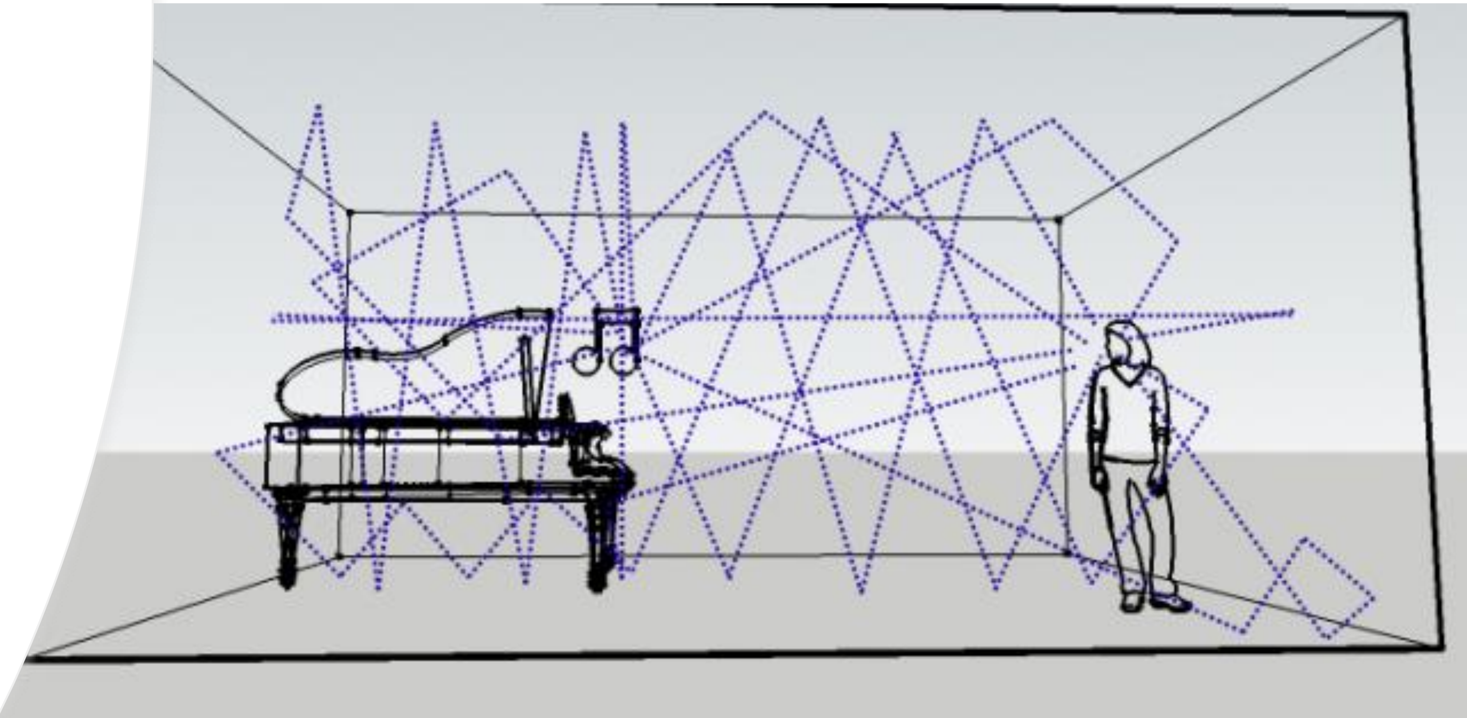
Definition

- The Impulse Response Function (IRF) quantifies how a one time shock $e_t = 1$ affects future values of y_t over time.

Steps to Analyze Dynamic Response:

1. Apply a one-time shock (e.g., set et = 1, et−1 = 0, et−2 = 0,. . .).

2. Observe how the series yt evolves using the model equations.



Equilibrium

The IRF describes how the output y_t of the system responds over time to a one-time shock of size 1 applied to the innovation e_t, assuming no other shocks occur (i.e., $e_{t+1} = 0, \mathrm{e}_{\mathrm{t}+2} = 0$, ...) and that the system starts at equilibrium.

What does equilibrium means?

1. **Constant Mean**: The process y_t fluctuates around a fixed mean value, which doesn't change over time. For example, if $E[y_t] = \mu$ for all t, the process is said to be in equilibrium with a mean of μ.
2. **Stationary Variance**: The variance of the process is constant over time, meaning Var(y_t) does not depend on t.
3. **No Unresolved Shocks**: Any shocks introduced in the past have already "died out," and the system has returned to its typical behavior.
4. **Time-Invariant Dynamics**: The statistical dependence of y_t on its past values is the same over time. For example, in an AR(1) process:

$$y_t = \phi_1 y_{t-1} + e_t,$$

the equilibrium is maintained as long as $|\phi_1| < 1$, ensuring stationarity.

IRF for AR(1)

$$y_t = \phi_1 y_{t-1} + e_t,$$

Suppose $y_0 = 0$ and $e_t = 0$ for all t. When a shock $e_0 = 1$ is applied at $t = 0$, we have: $y_0 = e_0 = 1$

After the shock: At t = 1: The output is influenced by y_0 and no new shocks ($e_1 = 0$):

$$y_1 = \phi_1 y_0 + e_1 = \phi_1 \cdot 1 + 0 = \phi_1$$

At t=2: $$y_2 = \phi_1 y_1 + e_2 = \phi_1(\phi_1) + 0 = \phi_1^2$$

At t=3: $$y_3 = \phi_1 y_2 + e_3 = \phi_1(\phi_1^2) + 0 = \phi_1^3$$

This recursive pattern continues, with $y_t = \phi_1^t$

- the system's response to the shock gradually diminishes over time unless $|\phi_1| = 1$, in which case the shock's effect would persist indefinitely.
- The rate at which the response decays depends on the value of ϕ_1
- If $|\phi_1| < 1$, the shock will eventually dissipate.

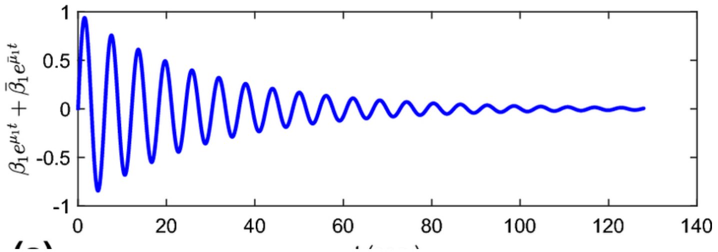
→ The value of the autoregressive coefficient ϕ_1 determines how quickly or slowly the system "forgets" the shock.

IRF for AR(2)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$

- At $t = 0$: $y_0 = 1$ (shock applied).
- At $t = 1$: $y_1 = \phi_1 y_0 + e_1 = \phi_1$.
- At $t = 2$: $y_2 = \phi_1 y_1 + \phi_2 y_0 = \phi_1^2 + \phi_2$.
- At $t = 3$: $y_3 = \phi_1 y_2 + \phi_2 y_1 = \phi_1^3 + \phi_1 \phi_2 + \phi_2 \phi_1$.

→ The dynamic response has a more complex behavior, potentially including oscillations depending on the roots of the characteristic equation



IRF for MA(1)

$$y_t = e_t + \theta_1 e_{t-1}$$

- At $t = 0$: $y_0 = 1$ (shock applied).
- At $t = 1$: $y_1 = \theta_1 e_0 = \theta_1 \times 1 = \theta_1$.
- At $t = 2$: $y_2 = \theta_1 e_1 = 0$ (because $e_1 = 0$)

General form:

$$\text{IRF}_t = \begin{cases} 1 & \text{if } t = 0, \\ \theta_1 & \text{if } t = 1, \\ 0 & \text{if } t \geq 2. \end{cases}$$

IRF for MA(2)

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$

- At $t = 0$: $y_0 = 1$ (shock applied).
- At $t = 1$: $y_1 = \theta_1 e_0 = \theta_1$.
- At $t = 2$: $y_2 = \theta_2 e_0 = \theta_2$.
- At $t = 3$: $y_3 = 0$ (shock has no further effect).

$$\mathrm{IRF}_t = \begin{cases} 1 & \text{if } t = 0, \\ \theta_1 & \text{if } t = 1, \\ \theta_2 & \text{if } t = 2, \\ 0 & \text{if } t \geq 3. \end{cases}$$

Finite Duration of Impact: impact of a shock is limited to a finite number of periods

→ If a shock e_0 occurs at t = 0, its effect will be felt at t = 0, 1, 2, . . . , q and will not persist beyond this time horizon

→ the IRF for an MA process has finite length (at most q +1 periods), unlike AR processes where the shock can have a persistent effect

Relationship Between IRF and MA Coefficients

The IRF at time t in an MA process is directly determined by the coefficients θ_i for i = 0, 1, . . . , t. For example:
• At t = 0, the response is $y_0 = e_0$.
• At t = 1, the response is $y_1 = \theta_1 e_0$.
• At t = 2, the response is $y_2 = \theta_2 e_0$, and so on.

→ The response to a shock diminishes (or increases) depending on the magnitude of the θ_i coefficients

Time-Limited Response:
Once the shock has passed through all its specified lags (from e_0 to e_q), the system is ”reset” and no further impact is observed.

IRF for ARMA

For an ARMA(p, q) process:

- **AR Component**: The AR component governs the **persistence** of shocks, creating longer-lasting effects through feedback.
- **MA Component**: The MA component determines the immediate response to shocks, capturing the direct influence of past error terms

- **Combined Behavior**: ARMA models combine the short-term effects of the MA component with the long-term effects of the AR component. The response to a shock typically decays over time, provided the AR roots lie outside the unit circle (stationarity)

Practical Implications:

- Short-Term Dynamics: MA terms capture short-term shock effects (e.g., sudden economic events).
- Long-Term Dynamics: AR terms describe how the system reverts to its mean or equilibrium after a shock.

→ **Forecasting**: ARMA models incorporate these dynamic responses to forecast future values based on observed data and shocks.

IRF for ARMA(1,1)

$$y_t = \phi_1 y_{t-1} + e_t + \theta_1 e_{t-1}$$

Impulse Response:

- At $t = 0$, $y_0 = 1$ (shock applied).
- At $t = 1$, $y_1 = \phi_1 y_0 + \theta_1 e_0 = \phi_1 + \theta_1$.
- At $t = 2$, $y_2 = \phi_1 y_1 + \theta_1 e_1 = \phi_1^2 + \phi_1\theta_1$.
- At $t = 3$, $y_3 = \phi_1 y_2 + \theta_1 e_2 = \phi_1^3 + \phi_1^2\theta_1$.

General Form:

$$\boxed{\text{IRF}_t = \phi_1^t + (\text{decay due to MA terms}).}$$

Key Points:

- The dynamic response is a combination of geometric decay (from AR) and short-term effects (from MA).
- The exact pattern depends on ϕ_1 and θ_1.

The AR component contributes a persistent effect, while the MA component adds a finite-duration influence

IRF for ARMA(2,2)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$

Impulse Response:

- At $t = 0$, $y_0 = 1$ (shock applied).
- At $t = 1$, $y_1 = \phi_1 y_0 + \theta_1 e_0 = \phi_1 + \theta_1$.
- At $t = 2$, $y_2 = \phi_1 y_1 + \phi_2 y_0 + \theta_2 e_0 = \phi_1^2 + \phi_1\theta_1 + \phi_2 + \theta_2$.
- At $t = 3$, $y_3 = \phi_1 y_2 + \phi_2 y_1 + \ldots$.

- The response behavior becomes more complex as AR and MA orders increase.
- Solving the characteristic equation of the AR polynomial and computing the weights of MA terms can provide exact IRF values.

Notes on the IRF for ARMA

1. Combination of AR and MA Effects:.

- The AR component provides a long-lasting effect that decays over time, as the current value depends on past values of the process.
- The MA component captures the short-term effect of past shock terms, which disappears after a finite number of periods (q).

2. Dynamically Changing IRF: The IRF is not fixed but varies based on the coefficients of the AR and MA components. A larger AR coefficient means a more persistent response to shocks, whereas a larger MA coefficient means a stronger immediate response to the shock.

3. Stationarity and Invertibility Considerations:

- The behavior of the IRF also depends on the stationarity of the AR part (i.e., whether the roots of the characteristic equation lie outside the unit circle)
- For a stationary ARMA model, the IRF will eventually decay to zero, but for a non-stationary process, the IRF can exhibit non-decaying, persistent behavior.

Evaluating ARMA

Evaluating an ARMA Model

As for AR and MA, you can use the AIC and BIC to evaluate them. However, it is also necessary to make some additional tests on the residuals to assess if the ARMA model is appropriated for the considered dataset

→ **Residuals diagnostic**

- Residuals are the differences between the observed values and the values predicted by the model.

- Residual diagnostics refers to the process of evaluating the residuals (errors) from a fitted time series model to check if the model has adequately captured the underlying data structure.

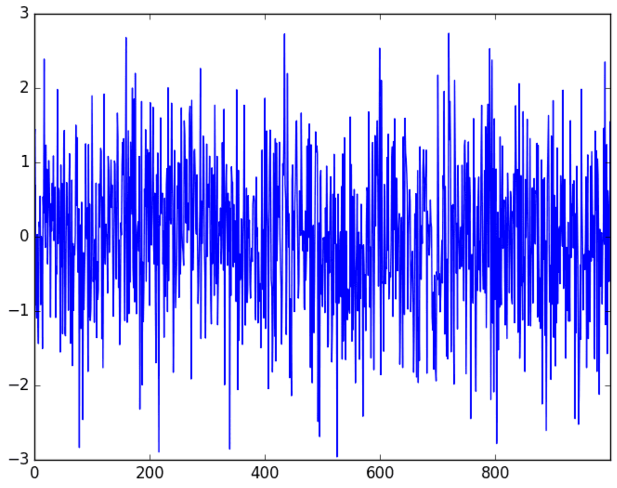
- Proper residual diagnostics are important because they help to ensure that the model assumptions hold and that the model is suitable for forecasting and inference.

A Focus on Residual Diagnostics

- **White Noise Assumption:** One of the key assumptions in time series modeling (especially for ARMA models) is that the residuals should behave like white noise. This means:
    - **Zero mean**: The residuals should have an average of zero.
    - **Constant variance**: The variance of the residuals should be constant over time (i.e., no heteroscedasticity).
    - **No autocorrelation**: There should be no significant correlations between the residuals at different lags.

Why do we want the errors to be white noise?

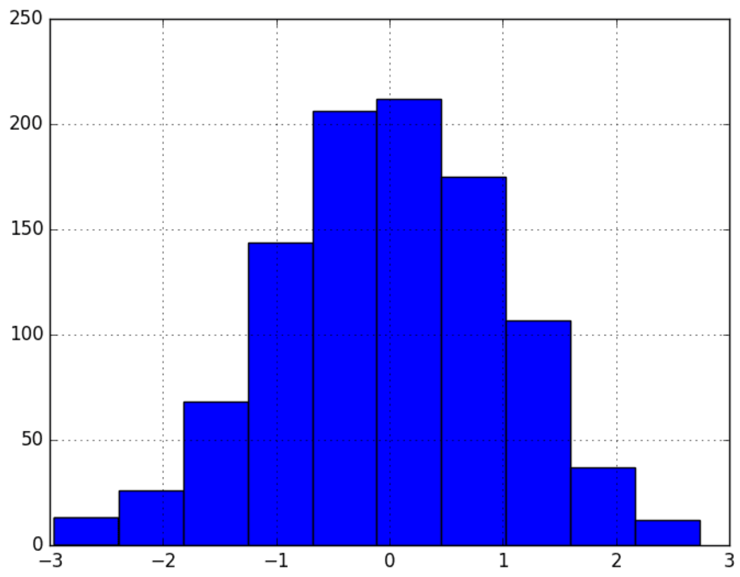
If a time series is white noise, it is a sequence of random numbers and cannot be predicted. If the series of forecast errors are not white noise, it suggests improvements could be made to the predictive model.

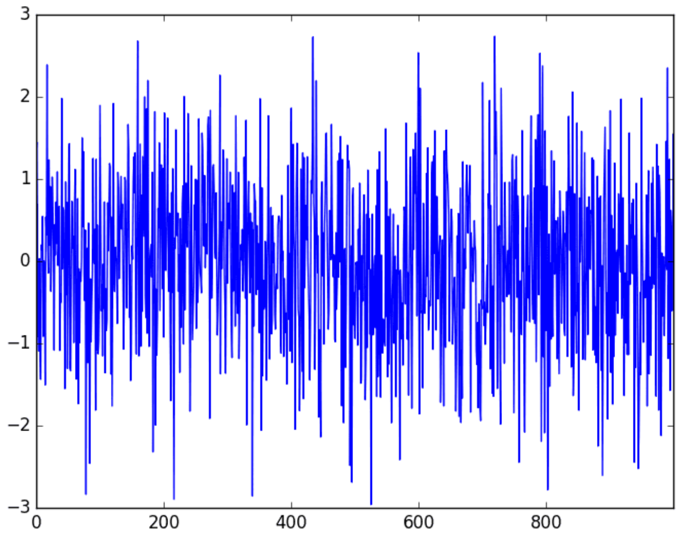


Methods for Residual Diagnostics (1/2)

- **Plotting Residuals**:
  - Time Series Plot: A simple plot of residuals over time to check for any patterns or trends. If there is a clear pattern or structure, the model might not have adequately captured the data's dynamics.

  - Histogram: A histogram of residuals can help to check for normality





Methods for Residual Diagnostics (2/2)

- **Autocorrelation of Residuals**:
  - ACF : The autocorrelation function of the residuals helps to check if there are any significant correlations between residuals at different lags. If the residuals are white noise, the ACF should show values close to zero for all lags.

  - Ljung-Box Test: A statistical test used to check if there is significant autocorrelation in the residuals. The null hypothesis is that the residuals are white noise (i.e., no autocorrelation)

- **Normality Test:** Shapiro-Wilk Test or Jarque-Bera Test can be used to check if the residuals are normally distributed.

- **Heteroscedasticity Test**: Tests like the Breusch-Pagan test or White test can be used to check for changing variance over time.

Interpretation for Residual Diagnostics

- **No Significant Patterns**: If the residuals appear as white noise (i.e., they show no significant autocorrelation, are normally distributed, and have constant variance), then the model is considered appropriate, and no further adjustments are needed.

- **Patterns or Autocorrelation**: If the residuals show patterns, significant autocorrelation, or heteroscedasticity, this suggests that the model is misspecified or that additional factors (e.g., non-linearity, non-stationarity, or seasonality) should be considered in the model.

- **Non-Normality**: If residuals are not normally distributed and the model assumes normal errors, this may affect confidence intervals and hypothesis tests. It could signal the need for a transformation or a more flexible model.

Limitations of ARMA Models (1/2)

- **Stationarity Assumption**: ARMA models are inherently designed for stationary time series, meaning they are unsuitable for modeling series with trends, changing variance, or seasonal behavior without transformation (e.g., differencing).

- **Linear Relationships**:
  - ARMA models assume that the relationships between past values and the noise are linear. This can limit their ability to model more complex, non-linear dependencies or interactions.

  - For non-linear patterns, more advanced models like GARCH, ARCH, or nonlinear ARMA might be needed.

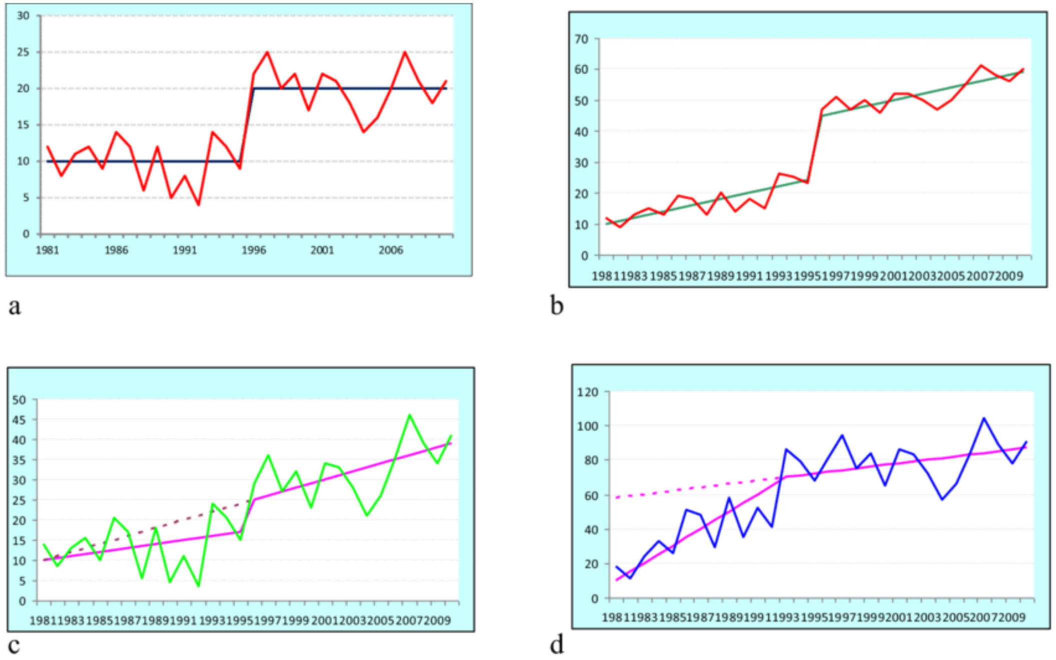
- **No Seasonality**: Standard ARMA models do not directly account for seasonality or periodic effects. If seasonality is present in the data, one needs to use SARMA (Seasonal ARMA) or SARIMA models to capture these effects.

Limitations of ARMA Models (2/2)

- **Overfitting**: There is a risk of overfitting the model by using too many AR or MA terms, especially when the data set is small or has a lot of noise.

- **Assumption of White Noise**: ARMA models assume that the error terms are white noise. In practice, this assumption may not always hold, leading to poor model performance.

- **Difficulty with Structural Breaks:** ARMA models struggle with structural breaks or sudden changes in the time series process. If there are abrupt shifts in the underlying data generating process (e.g., due to policy changes, economic shocks), ARMA models may not adequately capture the new dynamics.



ARIMA:
Autoregressive Integrated Moving Mverage

ARIMA vs ARMA

ARIMA is a simple extension of ARMA:

- **Similarity:** Their elements are identical, in the sense that both of them leverage a general autoregression AR(p) and general moving average model MA(q).

- **Difference:** The main differences between ARMA and ARIMA methods are the notions of integration and differencing.
  → ARMA Works for stationary series
  → ARIMA first makes a series stationary then performs ARMA on the resulting series

How does it make a series stationary ? It is possible to stationarize a time series through differencing

- The process of estimating how many nonseasonal differences are needed to make a time series stationarity is called integration (I)

ARIMA Parameters

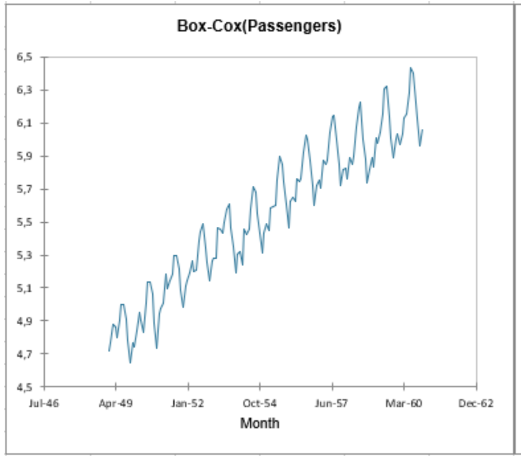
Process:

1. Take non stationary series
2. Make it stationary using differencing
3. Use ARMA as usual

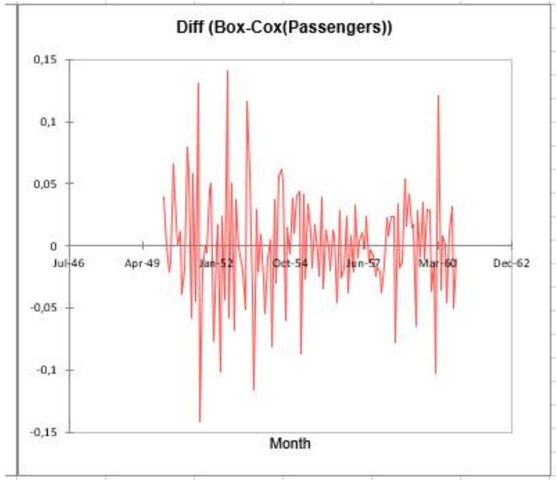
All in all, in the preceding exercise, you already applied ARIMA without realizing.

ARIMA models have three main components, denoted as p, d, q; these parameters are defined as follows:

- p stands for the number of lag variables included in the AR model, also called the lag order.
- d stands for the number of times that the raw values in a time series data set are differenced, also called the degree of differencing.
- q denotes the magnitude of the moving average window, also called the order of moving average.



Differencing



ARMA

Forecast