Time Series

Modeling complex time series

Echcharif EL JAZOULI
Yakine TAHTAH

Sia Partners

27 janvier 2025

Table of Contents

Note

Always verify what are the exact equations of the implementations that you're using. Specifically the sign convention of the parameters of the model.

From Moving Average to ARMA Models

- In last chapter, we covered the Moving Average (MA) process, denoted as MA(q), where the present value depends on past error terms.
- The order q can be identified using the ACF plot, which exhibits significant autocorrelations up to lag q.
- We studied the Autoregressive (AR) process, denoted as AR(p), where the present value depends on past values of itself.
- The order p can be identified using the PACF plot, where partial autocorrelations are significant up to lag p.
- However, when both ACF and PACF plots exhibit slow decay or sinusoidal patterns, we are in the presence of an Autoregressive Moving Average (ARMA) process.

Objectives of this Chapter

- Understand the ARMA(p,q) model and its components.
- Recognize the limitations of ACF and PACF plots in model identification.
- Learn how to select the best model using the Akaike Information Criterion (AIC).
- Analyze time series models using residual analysis.
- Establish a general modeling procedure for complex time series.
- Forecast time series using the ARMA(p,q) model.

Introduction to ARMA Models

- The ARMA(p,q) model combines both autoregressive and moving average components:
  - AR(p): Models dependency on past values.
  - MA(q): Models dependency on past forecast errors.
- Unlike pure AR or MA models, identifying p and q from ACF and PACF plots becomes more difficult.
- A structured approach is needed to select the best combination of p and q.

General Modeling Procedure

- Use AIC to determine the optimal values of p and q.
- Evaluate model validity using residual analysis:
  - Correlogram of residuals
  - Q-Q plot
  - Density plot
- If residuals resemble white noise, the model is valid for forecasting.

Examining the ARMA Process

- The ARMA(p,q) model combines autoregressive (AR) and moving average (MA) components.
- The present value is linearly dependent on:
  - Past values of the series (AR component)
  - Mean of the series, current error, and past errors (MA component)
- Mathematically, an ARMA(p,q) model is given by:

$$y_t = C + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{1}$$

Understanding ARMA Orders

- The order p determines the number of past values affecting the present value.
- The order q determines the number of past errors affecting the present value.
- Special cases:
  - ARMA(0,q) = MA(q) (pure moving average process)
  - ARMA(p,0) = AR(p) (pure autoregressive process)
- Example: An ARMA(1,1) model:

$$y_t = C + \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} \tag{2}$$

Higher-Order ARMA Models

- An ARMA(2,1) model combines an AR(2) and an MA(1):

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} \tag{3}$$

- The complexity increases with higher orders of p and q.
- Determining the optimal values of p and q is crucial for accurate modeling and forecasting.

Steps to Identify an ARMA Process

- Extend previous identification steps to include ARMA(p,q).
- If neither ACF nor PACF shows a clear cutoff between significant and non-significant coefficients, we likely have an ARMA process.
- Steps for identification:
    1. Check stationarity using ADF test.
    2. Examine ACF plot: No clear cutoff suggests ARMA.
    3. Examine PACF plot: No clear cutoff confirms ARMA.
    4. Use model selection criteria (e.g., AIC) to determine optimal p and q.

Simulating an ARMA(1,1) Process

- The ARMA(1,1) process is given by:

$$y_t = 0.33y_{t-1} + 0.9\epsilon_{t-1} + \epsilon_t \tag{4}$$

- We simulate using Python:

Analyzing ACF and PACF Plots

- ACF plot:
  - Shows a sinusoidal pattern, indicating autoregressive behavior.
  - No clear cutoff at lag q $\rightarrow$ Cannot determine q.
- PACF plot:
  - Also shows a sinusoidal pattern with no clear cutoff.
  - Cannot determine order p from PACF.
- Conclusion: When ACF and PACF show no clear cutoff, we conclude an ARMA(p,q) process.

Key Takeaways

- If ACF and PACF both show sinusoidal patterns or decaying, the process is likely ARMA.
- Traditional methods (ACF/PACF) fail to determine orders p and q.
- Model selection techniques such as AIC/BIC should be used.
- Residual analysis ensures a valid model.

Need for a General Modeling Procedure

- Identifying ARMA(p,q) orders using ACF and PACF is unreliable.
- A more structured approach is needed to determine the best p and q.
- This method extends to non-stationary and seasonal data.
- The new approach removes qualitative reliance on ACF/PACF and depends entirely on statistical tests.

Steps of the General Modeling Procedure

- Gather data
- Test for stationarity and apply transformations if necessary
- Define a list of possible values for p and q
- Fit every ARMA(p,q) model combination
- Compute the Akaike Information Criterion (AIC) for each model
- Select the model with the lowest AIC
- Analyze model residuals:
  - Use Q-Q plot for normality assessment (see later)
  - Check for uncorrelated residuals using the Ljung-Box test (see later)
- If residuals resemble white noise, the model is ready for forecasting.
- Otherwise, adjust p and q and restart the process.

Advantages of this Approach

- Eliminates reliance on subjective ACF/PACF interpretations.
- Provides a consistent framework applicable to different datasets.
- Ensures optimal model selection based on statistical validation.
- Incorporates residual analysis for validation before forecasting.

Next Steps

- Explore Akaike Information Criterion (AIC) in detail.
- Perform residual analysis with Q-Q plots and correlograms.
- Apply the procedure to our simulated ARMA(1,1) process.
- Extend the methodology to real-world data, such as bandwidth usage modeling.

What is the Akaike Information Criterion (AIC)?

- AIC estimates the quality of a model relative to other models.
- It quantifies the amount of information lost when a model is fitted to data.
- Lower AIC values indicate better models.

AIC Formula

- AIC is a function of:
  - k: Number of estimated parameters.
  - $\hat{L}$: Maximum value of the likelihood function.
- The formula for AIC:

$$AIC = 2k - 2\ln(\hat{L}) \tag{5}$$

- A balance between model complexity and goodness of fit.

Effect of Model Complexity on AIC

- More parameters (k) increase model complexity and penalize AIC.
- Higher likelihood ($\hat{L}$) improves fit and decreases AIC.
- AIC helps avoid:
  - Overfitting: Too many parameters make the model too specific.
  - Underfitting: Too few parameters lead to poor fit.

Likelihood Function and AIC

- The likelihood function measures how well the model fits the data.
- Given observed data, it estimates how likely different parameters generate it.
- A high likelihood value means the model fits well.
- AIC penalizes complex models but rewards good fit.

Using AIC for Model Selection

- Compute AIC for different ARMA(p,q) models.
- Select the model with the lowest AIC.
- Compare models relative to each other, as AIC is not an absolute metric.
- Apply this method to our simulated ARMA(1,1) process.

Steps to Select the Best Model

- Test for stationarity.
- Define a range of possible values for p and q.
- Fit all unique ARMA(p,q) models.
- Compute the Akaike Information Criterion (AIC) for each model.
- Select the model with the lowest AIC.

Optimizing ARMA Model Selection

- Generate all combinations of (p,q):
- Fit all models and compute AIC:

Result Interpretation

- The model with the lowest AIC is the best relative to other candidates.
- Example output:
- Next step: Perform residual analysis to validate model quality.

Understanding Residual Analysis

- Residuals are the difference between observed and predicted values.
- Good models have residuals that resemble white noise.
- Residual analysis includes:
  - Q-Q plot analysis.
  - Ljung-Box test for autocorrelation.

Quantile-Quantile (Q-Q) Plot

- A Q-Q plot compares residual quantiles to a normal distribution.
- If the residuals follow a normal distribution, the Q-Q plot forms a straight line.
- If residuals deviate from normality, the model may not be adequate.

Ljung-Box Test

- The Ljung-Box test checks whether residuals are uncorrelated.
- Null hypothesis: Residuals are independently distributed.
- If p-value > 0.05, we fail to reject the null hypothesis, meaning residuals resemble white noise.
- If p-value < 0.05, residuals are correlated, requiring a model adjustment.