Time Series

Overview of MA(p) and AR(q) Time Series Models

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.

MA(q): Moving Average of order q

- A process is said to follow an MA(q) model if its current value depends on the current random shock plus q previous shocks.
- **Formally**,

$$\text{MA}(q): \quad y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- μ is the mean of the series (possibly zero for simplicity).
- ε_t are zero-mean, uncorrelated white noise innovations with variance σ_ε^2.

MA(q): Moving Average of order q

- The “moving average” name arises because y_t is a weighted sum of the last q noise terms.
- MA models capture how present values depend on current and past “shocks” (error terms). They fade out if q is finite.
- They are often combined with AR models to form ARMA/ARIMA, but here we focus on pure MA behavior.

Stationarity of MA(q)

Stationarity Conditions:

- An MA(q) process (finite order) is automatically covariance stationary if the noise terms are stationary white noise.
- Because there are a finite number of past error terms, the necessary sum-of-squared coefficients is finite $\rightarrow$ no infinite memory.

Implication:

- No complicated root conditions for stationarity (as opposed to AR).
- In practice: As soon as q is finite, the process is stationary.

Autocovariance & Autocorrelation

Autocovariance function of MA(q):

- For $k > q$, $\gamma_k = 0$.
- For $k \leq q$, γ_k has a formula involving sums of products of θ-coefficients. One version is:

$$\gamma_k = \sigma_\varepsilon^2 \Big(-\theta_k + \sum_{j=1}^{q-k} \theta_j \, \theta_{j+k}\Big),$$

Autocorrelation function (ACF):

- $\rho_k = 0$ for $k > q$.
- ρ_k remains nonzero up to lag q.
- This "cut-off" in the ACF after lag q is a defining signature of an MA(q) process.

MA(1) Example

Model Form:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

Autocorrelations:

- $\rho_0 = 1$ (by definition).
- $\rho_1 = \frac{\theta_1}{1+\theta_1^2}$ (up to sign convention).
- $\rho_k = 0$ for $k > 1$. Hence a simple "cut-off" at lag 1.

Key Observations:

- The largest magnitude of ρ_1 can be 0.5 in absolute value.
- As $\theta_1 \to \pm 1$, $\rho_1 \to \pm 0.5$.
- MA(1) processes can show negative or positive short-term correlations.

MA(2) Example

Model Form:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}.$$

Autocorrelations:

- ρ_1 depends on θ_1 and θ_2.
- ρ_2 depends primarily on θ_2.
- $\rho_k = 0$ for $k > 2$.

Interpretation:

- The current value depends on the current shock and the last two shocks.

Why Do We See a Sharp ACF Cut-Off?

MA(q) as a finite sum of q+1 noise terms:

- The correlation between y_t and y_{t+k} only stems from overlapping noise terms.
- Once $k > q$, there is no overlap in those noise terms.
- If $q = 2$, y_t depends on $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}$ while y_{t+3} depends on $\varepsilon_{t+3}, \varepsilon_{t+2}, \varepsilon_{t+1}$.
- No shared noises if $k = 3$.
- Memory is only from q recent "shock" values. After q lags, no shared random shock remains $\rightarrow \rho_k = 0$.

Identification Strategy

Steps:

1. Check for Stationarity:
   - If data is not stationary, difference or transform.
2. Look at ACF:
   - For a pure MA(q), we expect a sharp cut-off at lag q.
3. If cut-off is abrupt at some lag q (and partial autocorrelation does not show "tail"), an MA(q) is suspect.
4. If no cut-off or if we see sinusoidal damping in ACF, suspect AR... or ARMA.

Estimation Approaches

Once q is tentatively identified from ACF:

- Method of Moments (matching sample autocorrelations).
- Maximum Likelihood Estimation (MLE) often used in software (e.g., statsmodels in Python).
- Innovations (recursive) algorithms are also possible.

Software Tools:

- In R: `arima()` with `order=c(0,0,q)`.
- In Python's statsmodels: `SARIMAX(..., order=(0,0,q))`.

Interpretation of Estimated θ_j:

- Sign & magnitude show which past shocks push the series up or down.
- Model is typically straightforward to interpret if q is small.

Forecasting MA(q)

Forecasting Logic:

- In an MA(q), the future value depends on future shocks (not yet realized) + known past shocks.
- Beyond q steps ahead:
  - The forecast inevitably reverts to the mean (or only slight corrections if partial knowledge of new shocks is present).

Multistep Forecast:

- For horizon $h > q$, the predicted value simplifies to μ because older shocks vanish & future shocks have expectation zero.
- For short horizons up to q, need to carefully handle the "not-yet-observed" error terms.
- In practice, rolling forecasts are used for multi-step if q is less than the forecast horizon.

Rolling Forecasting Demonstration

Rolling or Sliding Window Forecast:

- Fit an MA(q) on data up to time t.
- Forecast up to q steps.
- Shift the window by q steps, incorporate newly observed data, and repeat.

Practical Tips & Pitfalls

- **Don't forget stationarity:**
  - Data with trend or seasonality must be differenced or detrended first.
  - Otherwise, an MA might appear with spurious patterns.
- **Overfitting hazard:**
  - If the ACF does not neatly cut off, do not force an MA(q). Possibly it's an AR, ARMA, or something else.
- **Residual diagnostics:**
  - After fitting an MA(q), check residuals $\rightarrow$ ideally white noise, no further correlation.
- **Check invertibility:**
  - For a later course.

AR(p): Autoregressive of order p

- A process is said to follow an AR(p) model if its current value depends on p of its own past values plus a current white noise term.
- **Formally**,

$$\text{AR}(p): \quad y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

  where
  - c is a constant term (often denoted as μ or C).
  - ϕ_i are parameters defining how strongly past values influence the present.
  - ε_t are zero-mean, uncorrelated white noise innovations with variance σ_ε^2.
- The name "autoregressive" arises because y_t is regressed on its own previous values.

Stationarity of AR(p)

Stationarity Conditions:

- Unlike a finite MA(q), an AR(p) process *is not automatically* covariance stationary.
- The characteristic equation $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ must have **all roots** lying **outside** the unit circle for the process to be stationary.
- Equivalently, $|\phi_i| < 1$ is not enough by itself if $p > 1$; you must check the overall polynomial's roots.

Implications:

- If the stationarity condition is not satisfied (e.g., a root on or inside the unit circle), the process can exhibit non-stationary behavior (like a random walk, if $\phi_1 = 1$ for AR(1)).
- In practice, you often difference non-stationary data or fit an ARIMA model to handle trends.

Autocovariance & Partial Autocorrelation for AR(p)

- **Autocovariance function (ACF):**
  - For a stationary AR(p), the ACF typically shows a *gradual* (often exponential or damped sinusoidal) decay rather than a strict cut-off.
  - The exact form of γ_k depends on solving a system of Yule-Walker equations.
- **Partial Autocorrelation Function (PACF):**
  - Measures the direct correlation of y_t with y_{t-k}, *excluding* indirect correlations via intervening lags.
  - For an AR(p), the PACF **cuts off** after lag p. That is, $\text{PACF}(k) = 0$ for $k > p$.
  - This "cut-off" in the PACF at lag p is a hallmark of a pure AR(p) model.

AR(1) Example

Model Form:

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t.$$

Stationarity Condition:

- $|\phi_1| < 1$ for stationarity in the AR(1) case.
- If $\phi_1 = 1$, it becomes a *random walk* (non-stationary).
- If $c \neq 0$ and $\phi_1 < 1$, it's an AR(1) with drift.

Autocorrelation Patterns:

- ACF declines exponentially (or oscillates if $\phi_1 < 0$) but no abrupt cut-off.
- PACF typically shows a significant spike at lag 1, then no large spikes beyond that.

AR(2) Example

Model Form:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t.$$

Stationarity Condition:

- The polynomial $1 - \phi_1 z - \phi_2 z^2 = 0$ must have solutions with $|z| > 1$.
- If this condition fails, the process may be non-stationary (or even explosive).

ACF & PACF:

- ACF: Exponential or damped oscillatory decay.
- PACF: Sharp cut-off at lag 2 (i.e., no significant PACF beyond lag 2).

Interpretation:

- y_t depends on the previous two values y_{t-1}, y_{t-2}.
- Coefficients ϕ_1, ϕ_2 indicate how each lag contributes positively or negatively.

Identification Strategy for AR(p)

Steps:

1. **Check for Stationarity:**
   - If data is not stationary, difference or transform (like taking $\nabla y_t = y_t - y_{t-1}$).
   - Use ADF test or KPSS test.
2. **Look at ACF and PACF:**
   - An AR(p) often shows a **gradually** decreasing ACF (no abrupt cut-off).
   - The PACF **cuts off** sharply after lag p.
3. **Tentative identification:**
   - If the PACF "dies out" after lag p, suspect an AR(p).
   - If both ACF and PACF show slow decay or complex patterns, suspect ARMA or ARIMA.

Estimation Approaches for AR(p)

Once p is tentatively identified from PACF:

- **Yule-Walker estimation** (for pure AR processes):
  - Method of moments matching the autocovariances.
- **Least Squares or Maximum Likelihood Estimation (MLE):**
  - In practice, MLE or conditional least squares is common (e.g., in `statsmodels`).
- **Software Tools:**
  - R: `arima()` with `order = c(p,0,0)`.
  - Python `statsmodels`: `SARIMAX(..., order=(p,0,0))`.

Interpretation of Estimated ϕ_i:

- Sign and magnitude show which past values push the series up or down.
- Large $|\phi_i|$ means strong dependency on lag i.

Forecasting AR(p)

Forecasting Logic:

- In AR(p), future values depend on known past values up to y_{t-p} plus unobserved future shocks.
- Multi-step predictions require iterating the AR recursion:

$$\hat{y}_{t+h} = c + \phi_1 \hat{y}_{t+h-1} + \phi_2 \hat{y}_{t+h-2} + \cdots + \phi_p \hat{y}_{t+h-p}.$$

- As h grows large, forecasts often revert gradually toward a long-run mean (depending on ϕ_i).

Multistep Forecast:

- For horizon h, if $h \leq p$, the past p observations are partly known.
- For $h > p$, the forecast uses previously forecasted points as inputs (iterative approach).
- ε_{t+h} has expected value zero, so it does not contribute directly to the forecast's mean.

Practical Tips & Pitfalls for AR Models

- **Check Stationarity Carefully:**
  - AR processes require stricter conditions than finite MA for stationarity. If there is a trend, difference or detrend the data.
  - If $\phi_1 \approx 1$ in an AR(1), the series might be nearly non-stationary and can behave like a random walk.
- **Overfitting Hazard:**
  - Selecting p too large can lead to over-parameterization and poor forecasts out of sample. Use criteria such as AIC, BIC to find a suitable p.
- **Residual Diagnostics:**
  - After fitting AR(p), residuals should be white noise (check with ACF/PACF of residuals. If residuals still show structure, consider higher p or an ARMA model.
- **Roots and Invertibility:**
  - For *stationarity* of AR, ensure roots of the characteristic polynomial lie outside the unit circle.

Conclusion

Key Takeaways:

- Moving Average (MA) models capture dependencies on past shocks, with an ACF that cuts off after lag q.
- Autoregressive (AR) models capture dependencies on past values, with a PACF that cuts off after lag p.
- Stationarity is crucial—MA(q) is always stationary for finite q, while AR(p) requires checking roots of its characteristic polynomial.
- Use ACF and PACF for model identification; then use appropriate estimation methods (Yule-Walker, MLE, etc.).
- Always validate the final model via diagnostic checks on residuals.

Next Steps?

- ARMA and ARIMA models for more complex data.
- Incorporate seasonality / exogenous variables where relevant.

Thank you!