Linear Time Series Models: ARMA Processes

From Stationary Noise to Structured Dynamics

Diogo Ribeiro

Data Science and Applied Mathematics

November 11, 2025

Outline

Where We Are Coming From

- We have seen:
  - Stochastic processes and time-ordered data.
  - Stationarity (strict and weak).
  - Ergodicity: time averages $\approx$ ensemble averages.
- For modelling and forecasting, we often assume:
  - Weak stationarity: mean, variance, and autocovariance are time-invariant.
  - Some form of ergodicity, so sample moments are informative.

Next step: build concrete models for stationary series that explain temporal dependence.

Goal of This Talk

- Introduce **linear time series models**, especially ARMA:
  - Autoregressive (AR) processes.
  - Moving average (MA) processes.
  - Combined ARMA processes.
- Show:
  - How these models are defined.
  - Conditions for stationarity and invertibility.
  - Behaviour of autocorrelation and partial autocorrelation.
  - Basic identification and estimation ideas.

These models form the basis for ARIMA, SARIMA, VAR, and many other extensions.

Linear Time Series Models

- Many useful time series can be written as **linear** functions of white noise:
$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j},$$
with $\{\varepsilon_t\}$ white noise and $\{\psi_j\}$ real coefficients.
- Intuition:
  - The current value X_t is a filtered version of past shocks.
  - The filter $\{\psi_j\}$ controls dependence over time.
- AR, MA, and ARMA models are finite or rational forms of such filters.

Wold Decomposition (Informal)

Wold theorem (informal):

- Any purely non-deterministic, weakly stationary process can be expressed as
$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$
where $\{\varepsilon_t\}$ is white noise and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.
- There may be an additional deterministic part, e.g. a periodic component, but we focus on the random part.

ARMA models are finite-order approximations of this infinite linear representation.

Definition of an AR(p) Process

Autoregressive process of order p (AR(p)):

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t,$$

where:

- $\{\varepsilon_t\}$ is white noise with mean 0 and variance σ_ε^2.
- $\phi_1, \ldots, \phi_p$ are real coefficients.

Current value depends linearly on previous p values plus a random innovation.

Backshift Operator and Characteristic Polynomial

- Define the backshift operator B by $BX_t = X_{t-1}$.
- AR(p) model can be written as:

$$\phi(B)X_t = \varepsilon_t,$$

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p.$$

- The polynomial

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

is the **characteristic polynomial** of the AR part.

Stationarity Condition for AR(p)

- A fundamental result:
  - An AR(p) process has a unique weakly stationary solution if and only if all roots of

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0$$

    lie **outside** the unit circle, i.e. $|z| > 1$.
- If this holds:

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

  for some absolutely summable sequence $\{\psi_j\}$.
- If the condition fails:
  - The process is non-stationary.
  - Variance typically grows without bound or is undefined.

Example: AR(1) Model

AR(1):

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2).$$

- Stationarity condition: $|\phi| < 1$.
- Mean:

$$E[X_t] = 0 \quad \text{(assuming zero mean innovations)}.$$

- Variance:

$$\mathrm{Var}(X_t) = \frac{\sigma_\varepsilon^2}{1 - \phi^2}.$$

- Autocorrelation:

$$\rho(h) = \phi^{|h|}, \quad h \in \mathbb{Z}.$$

The ACF of a stationary AR(1) decays geometrically.

ACF and PACF of AR(p)

- For an AR(p) process:
  - The **ACF** decays (typically exponentially or as a damped sine wave).
  - The **PACF** (partial autocorrelation function) has a **cut-off** at lag p:

$$\alpha(h) = \begin{cases} \neq 0, & h \leq p, \\ 0, & h > p, \end{cases}$$

  in the ideal infinite-sample case.
- In finite samples, the cut-off is approximate:
  - PACF values after lag p fluctuate around zero within significance bounds.

This pattern is often used for **model identification**.

Definition of an MA(q) Process

Moving average process of order q (MA(q)):

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

where:

- $\{\varepsilon_t\}$ is white noise with mean 0 and variance σ_ε^2.
- $\theta_1, \ldots, \theta_q$ are real coefficients.

The current value is a finite linear combination of current and past innovations.

Stationarity of MA(q)

- Any MA(q) process with finite q and finite variance innovations is **weakly stationary**:
  - Mean:
  $$E[X_t] = 0.$$
  - Variance:
  $$\mathrm{Var}(X_t) = \sigma_\varepsilon^2 \left(1 + \theta_1^2 + \cdots + \theta_q^2\right).$$
  - Autocovariance:
  $$\gamma(h) = \sigma_\varepsilon^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}, \quad h = 0, 1, \ldots, q,$$

  and $\gamma(h) = 0$ for $h > q$. (With $\theta_0 = 1$.)
- No additional stationarity condition required.

ACF of MA(q) and Invertibility

- For an MA(q) process:
  - The **ACF** has a **cut-off** at lag q:

$$\rho(h) = 0 \quad \text{for } h > q.$$

  - The **PACF** decays gradually.
- **Invertibility:**
  - We can sometimes write

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad \Longleftrightarrow \quad \varepsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

    if the MA polynomial has roots outside the unit circle.
  - Invertibility ensures a unique representation in terms of X_t.

MA(1) Example

MA(1):

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}.$$

- Mean: $E[X_t] = 0$.
- Variance:

$$\mathrm{Var}(X_t) = \sigma_\varepsilon^2(1+\theta^2).$$

- Autocorrelation:

$$\rho(0) = 1, \quad \rho(1) = \frac{\theta}{1+\theta^2}, \quad \rho(h) = 0 \text{ for } h \geq 2.$$

- Invertibility condition:

$$|\theta| < 1.$$

The ACF cuts off after lag 1, which is a clear signature of an MA(1) structure.

Definition of an ARMA(p,q) Process

ARMA(p,q) model:

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

or in compact form:

$$\phi(B) X_t = \theta(B) \varepsilon_t,$$

where:

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p, \\ \theta(B) &= 1 + \theta_1 B + \cdots + \theta_q B^q. \end{aligned}$$

This combines autoregressive and moving average parts.

Stationarity and Invertibility in ARMA(p,q)

- **Stationarity:**
  - AR part must satisfy: all roots of
    $$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0$$
    lie outside the unit circle ($|z| > 1$).
- **Invertibility:**
  - MA part must satisfy: all roots of
    $$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q = 0$$
    lie outside the unit circle.
  - This ensures a unique representation in terms of the innovations.
- Under these conditions:
  $$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$
  with $\{\psi_j\}$ absolutely summable.

ACF and PACF Patterns in ARMA

- **AR(p):**
  - ACF: decays slowly.
  - PACF: cut-off after lag p.
- **MA(q):**
  - ACF: cut-off after lag q.
  - PACF: decays slowly.
- **ARMA(p,q):**
  - Both ACF and PACF decay.
  - No sharp cut-off in either.

These heuristics guide the order selection but are not foolproof.

Interpretation of ARMA Components

- **AR part:**
  - Captures persistence in the levels of the series.
  - Similar to regressing X_t on its own past.
- **MA part:**
  - Captures short-term shocks that persist for a few periods.
  - Models correlations in the residuals.
- ARMA models:
  - Flexible yet relatively parsimonious.
  - Good baseline for many stationary time series.

Model Identification

- **Step 1: Check stationarity.**
  - Visual inspection, unit root tests.
  - Difference or transform the series if needed.
- **Step 2: Inspect ACF and PACF.**
  - Look for cut-offs or slow decays.
  - Suggest ranges for (p, q).
- **Step 3: Compare candidate models.**
  - Fit several ARMA(p,q) models.
  - Use information criteria: AIC, BIC, HQ.

Always balance goodness of fit and parsimony.

Information Criteria

For a fitted model with log-likelihood ℓ and k parameters on T observations:

- **Akaike Information Criterion (AIC):**

$$\mathrm{AIC} = -2\ell + 2k.$$

- **Bayesian Information Criterion (BIC):**

$$\mathrm{BIC} = -2\ell + k \log T.$$

- Lower values indicate a better trade-off between fit and complexity.

In practice, BIC tends to favour more parsimonious models than AIC.

Estimation Methods

- **Maximum Likelihood (ML):**
  - Assume Gaussian innovations.
  - Numerically maximize the likelihood function.
  - Provides parameter estimates and approximate standard errors.
- **Conditional least squares:**
  - Minimize squared prediction errors, conditional on initial values.
  - Often used as starting values for ML.
- **Software:**
  - R: `arima`, `Arima`, `arima_order`.
  - Python: `statsmodels.tsa.ARIMA`, `SARIMAX`.

Residual Diagnostics

After fitting an ARMA model:

- Compute residuals:
$$\hat{\varepsilon}_t = X_t - \hat{X}_t.$$
- Check:
  - Residual series plot: look for structure or regime changes.
  - ACF/PACF of residuals: should resemble white noise.
  - Portmanteau tests (Ljung–Box, Box–Pierce).
  - Normal Q-Q plot (if Gaussian assumption is important).
- If residuals show remaining dependence:
  - Increase model order, or
  - Change model structure.

Example ARMA Modelling Workflow

1. **Plot the series.**
   - Identify trends, cycles, structural breaks.
2. **Stabilise the series.**
   - Difference to remove trends or unit roots.
   - Transform (log, Box–Cox) to stabilise variance.
3. **Inspect ACF/PACF.**
   - Suggest candidate (p, q).
4. **Fit several ARMA(p,q) candidates.**
   - Use ML or conditional least squares.
5. **Compare and refine.**
   - Use AIC/BIC and residual diagnostics.
6. **Forecast and evaluate.**
   - Create forecasts.
   - Compare against hold-out data if possible.

Simple Conceptual Example

Consider a stationary series after differencing and log-transform:

- ACF shows geometric decay.
- PACF cuts off after lag 2.

Steps:

1. Propose ARMA(2,0) or AR(2).
2. Fit AR(2), check residuals:
   - If residual ACF is near zero, model may be adequate.
   - If residual ACF shows a spike at lag 1, consider ARMA(2,1).
3. Compare AIC/BIC for AR(2) vs ARMA(2,1).
4. Choose final model and compute forecasts.

Summary

- Linear time series models represent X_t as a filtered version of white noise.
- AR(p) models use past values of X_t.
- MA(q) models use past innovations.
- ARMA(p,q) models combine both to capture richer dynamics.
- Stationarity and invertibility are expressed through root conditions on $\phi(z)$ and $\theta(z)$.
- ACF and PACF provide useful patterns for order identification.

What Comes Next?

- Non-stationary series:
  - ARIMA models (integrated ARMA).
  - Seasonal ARIMA (SARIMA).
- Conditional heteroscedasticity:
  - ARCH, GARCH and related models.
- Multivariate extensions:
  - Vector autoregressions (VAR).
  - VARMA models.
- State-space formulations and Kalman filtering.

ARMA forms the core building block for many of these generalizations.

Further Reading

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control*. Wiley.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Springer.
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications*. Springer.
- Tsay, R. S. (2010). *Analysis of Financial Time Series*. Wiley.

AR(p) Yule–Walker Equations

For a zero-mean AR(p) process:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t,$$

with ε_t white noise, the autocovariances satisfy:

$$\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p), \quad h \geq 1,$$

and

$$\gamma(0) = \phi_1 \gamma(1) + \cdots + \phi_p \gamma(p) + \sigma_\varepsilon^2.$$

These are the **Yule–Walker equations**, used for parameter estimation.

MA(q) Spectrum

For an MA(q) process:

$$X_t = \sum_{j=0}^{q} \theta_j \varepsilon_{t-j}, \quad \theta_0 = 1,$$

the spectral density is:

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \sum_{j=0}^{q} \theta_j e^{-i\lambda j} \right|^2, \quad \lambda \in [-\pi, \pi].$$

ARMA spectra have a similar form with rational functions of $e^{-i\lambda}$.

Thank You

Questions or discussion?