Autoregressive models

- **An autoregressive model of order *p* (denoted as AR(*p*)), is of the form**

  $$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

  $$w_t \sim \mathrm{wn}(0, \sigma_w^2)$$

  $$\phi_1, \phi_2, \ldots, \phi_p \text{ are constants } (\phi_p \neq 0)$$

Autoregressive models

Review

- **AR(p) model:**

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

$$\Downarrow$$

$$\phi(B)x_t = w_t, \quad \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

Moving average models

- **The moving average model with order q, or MA(q) model, is defined to be**

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$

$$w_t \sim \mathrm{wn}(0, \sigma_w^2),$$

$$\theta_1, \theta_2, \ldots, \theta_q (\theta_q \neq 0) \text{ are parameters}$$

Moving average models

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$
$$w_t \sim \mathrm{wn}(0, \sigma_w^2),$$
$$\theta_1, \theta_2, \ldots, \theta_q (\theta_q \neq 0) \text{ are parameters}$$

---

x_t the **observed value** at time point t

w_t w_{t-1} … w_{t-q} independent, random, unpredictable "events"

These are a series of **white noise** random variables.

- `w_t` represents the "shock" or "innovation" at the current time period.
- `w_{t-1}` represents the shock from the previous time period, and so on, up to `w_{t-q}` (the shock from `q` periods ago).

Moving average models

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$
$$w_t \sim \text{wn}(0, \sigma_w^2),$$
$$\theta_1, \theta_2, \ldots, \theta_q (\theta_q \neq 0) \text{ are parameters}$$

θ_1 θ_2 θ_q,

They represent **the degree of influence that past "shocks" have on the current observation x_t**.

- `θ₁` measures how much the shock from the **previous period** (`w_{t-1}`) affects the current value (`x_t`).
- `θ₂` measures the influence of the shock from **two periods ago** (`w_{t-2}`), and so on.

Moving average models

- **Moving average operator**

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$

$$\Downarrow$$

$$x_t = (\boxed{1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q})w_t$$

$$\Downarrow$$

$$x_t = \theta(B)w_t$$

Step 1: The Core Tool - The Backshift Operator (B)

The key to this derivation is the **Backshift Operator**, denoted by `B`.

- **`B` operates on a time series variable.**
- **`B` applied to the value at time `t` gives the value at time `t-1`.**

Mathematically:

$$\mathbf{B \cdot x_t = x_{t-1}}$$

$$\mathbf{B \cdot w_t = w_{t-1}}$$

By extension:

- **B^2** applies the shift twice:

  $B^2 \cdot w_t = B \cdot (B \cdot w_t) = B \cdot (w_{t-1}) = w_{t-2}$

- **B^q** applies the shift `q` times:

  $B^q \cdot w_t = w_{t-q}$

Step 2: Rewrite the MA(q) Model Using B

Now, let's rewrite each term in the original MA(q) equation using the backshift operator `B`:

- `w_t` remains as `w_t`. This can be thought of as `B⁰ · w_t` (where `B⁰ = 1`, the identity operator).
- `θ₁ w_{t-1}` becomes `θ₁ · (B · w_t)`
- `θ₂ w_{t-2}` becomes `θ₂ · (B² · w_t)`
- `...`
- `θ_q w_{t-q}` becomes `θ_q · (B^q · w_t)`

Substituting these back into the original model gives us:

$x_t = w_t + \theta_1 B w_t + \theta_2 B^2 w_t + \ldots + \theta_q B^q w_t$

Step 3: Factor Out the Common Factor `w_t`

Notice that `w_t` is a common factor in every term on the right-hand side. We can factor it out:

$$x_t = (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q)\, w_t$$

Step 4: Define the Moving Average Polynomial θ(B)

We now define a polynomial function in terms of the backshift operator B, called the **Moving Average Polynomial** and denoted as θ(B):

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$

Moving average models

- **Moving average operator**

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$

$$\Downarrow$$

$$x_t = (\boxed{1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q})w_t$$

$$\Downarrow$$

$$x_t = \theta(B) w_t$$

Difference

Autoregressive models
Moving average models

The simplest way to distinguish them:

Ask yourself one question: **What is driving the system?**

- If the answer is **"some recent external events,"** then it's more like an **MA** model.
  Example: "The market declined because bad news was released yesterday and today."
- If the answer is **"the system's own previous state,"** then it's more like an **AR** model.
  Example: "The market continued to fall today because it was already in a downward trend."

Review

- **AR(p) model:**

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

$$\Downarrow$$

$$\phi(B) x_t = w_t, \quad \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

ARMA(p,q) model

- **An ARMA(p,q) process is a stationary process that satisfies**

$$x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q},$$

$$w_t \sim \mathrm{wn}(0, \sigma_w^2)$$

$$\phi_p \neq 0,\ \theta_q \neq 0$$

$p:$ autoregressive order

$q:$ moving average order

ARMA(p,q) model

- **An ARMA(p,q) process is a stationary process that satisfies**

$$x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q},$$

Key Components:

1. **Autoregressive (AR) Part:** `x_t - φ₁x_{t-1} - ... - φ_p x_{t-p}`
   - This part represents the regression of `x_t` on its **own past values**.
   - `p` is the **autoregressive order** (number of own past lags used).
   - `φ₁, ..., φ_p` are the **autoregressive parameters**.
   - The condition `φ_p ≠ 0` ensures the AR part is truly of order `p`.

ARMA(p,q) model

- **An ARMA(p,q) process is a stationary process that satisfies**

$$x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q},$$

2. **Moving Average (MA) Part:** `w_t + θ₁ w_{t-1} + ... + θ_q w_{t-q}`
   - This part represents a weighted sum of the current and past **random shocks** (white noise terms).
   - `q` is the **moving average order** (number of past shock terms used).
   - `θ₁, ..., θ_q` are the **moving average parameters**.
   - The condition `θ_q ≠ 0` ensures the MA part is truly of order `q`.

ARMA(p,q) model

Intuition

The current value of the time series, x_t, can be explained and predicted by two components working together:

1. **The "Inertia" or "Memory" from Itself (the AR Part):**
   - This means that x_t is a **weighted linear combination of its own past values**.
   - The AR component captures the **trends and cycles that depend on its own history**.
2. **The "Random Shocks" or "Disturbances" from the Outside (the MA Part):**
   - This means that x_t is also a **weighted linear combination of current and past random shocks**.
   - This captures the effects of **sudden, temporary events**.
   - The MA component precisely captures the **short-term effects** of such external events.

ARMA(p,q) model

Why is this powerful?

The ARMA model is a versatile tool for modeling time series because it can capture a wide range of behaviors (like persistence from the AR part and abrupt changes from the MA part) with often relatively few parameters (`p + q`), making it more parsimonious and efficient than a pure AR or pure MA model might be.

The ARMA model combines both: imagine clapping your hands in a living room with some echo, but not as much as in a bathroom. The sound you hear includes both the original clap and any immediate mechanical sounds it causes (the MA part), as well as the echoes reflected from the walls (the AR part). This enables it to describe more complex and realistic phenomena in the world.

ARMA(p,q) model

- **Express in terms of the AR and MA operators:**

$$x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q},$$

$$\Downarrow$$

$$\phi(B)x_t = \theta(B)w_t$$

- **Special cases:**

  AR(p) = ARMA(p,0) with $\theta(B) = 1$

  MA(q) = ARMA(0,q) with $\phi(B) = 1$

ARMA(p,q) model

The Backshift Operator Representation

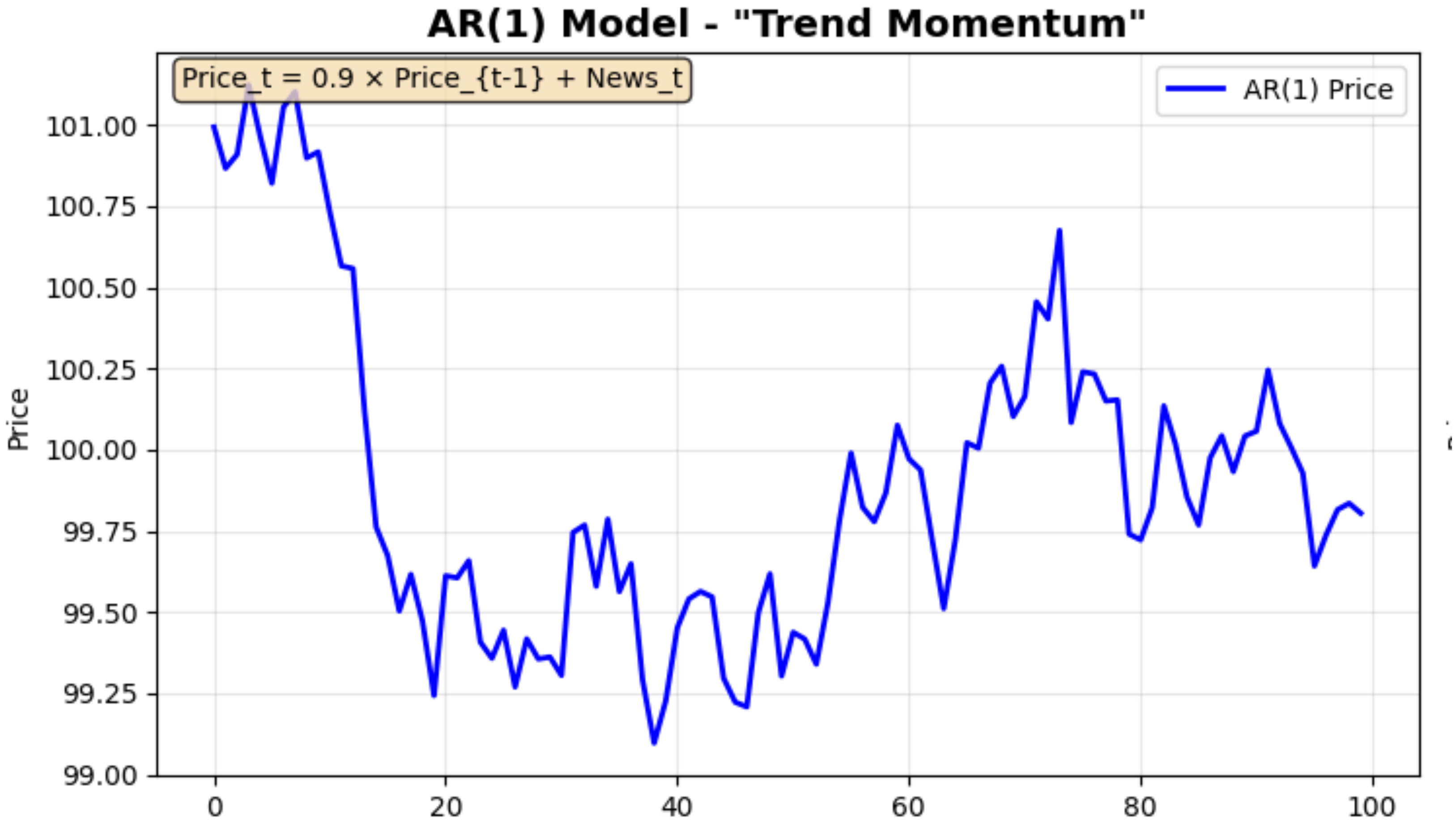
Using the **backshift operator** `B` (where `B x_t = x_{t-1}` and `B w_t = w_{t-1}`), we can write the model much more compactly.

1. **Define the AR Polynomial:**
   `**φ(B) = 1 - φ₁B - φ₂B² - ... - φ_p B^p**`
2. **Define the MA Polynomial:**
   `**θ(B) = 1 + θ₁B + θ₂B² + ... + θ_q B^q**`

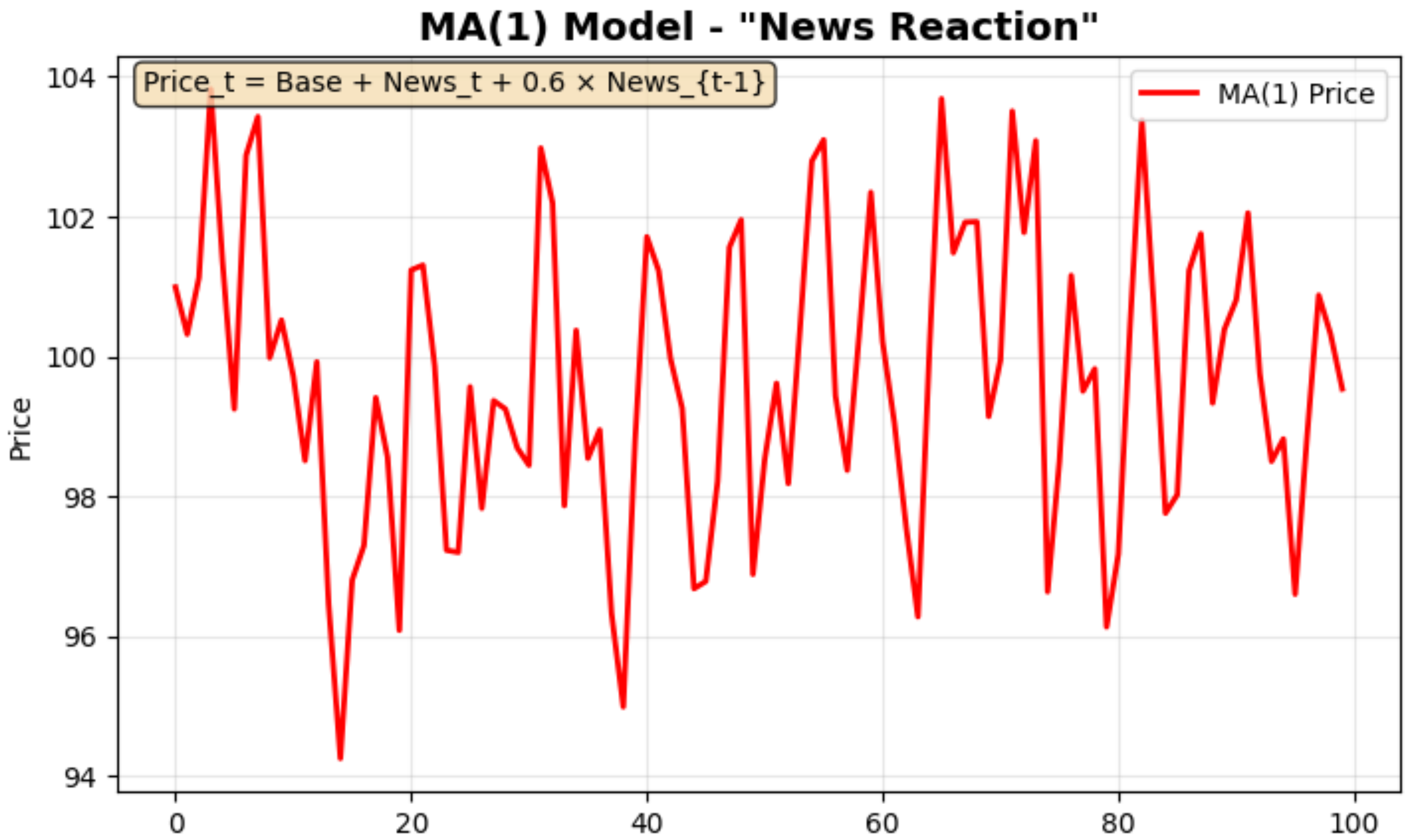
Using these polynomials, the entire ARMA(p, q) model condenses into a single, elegant equation:

`φ(B) x_t = θ(B) w_t`

Examples

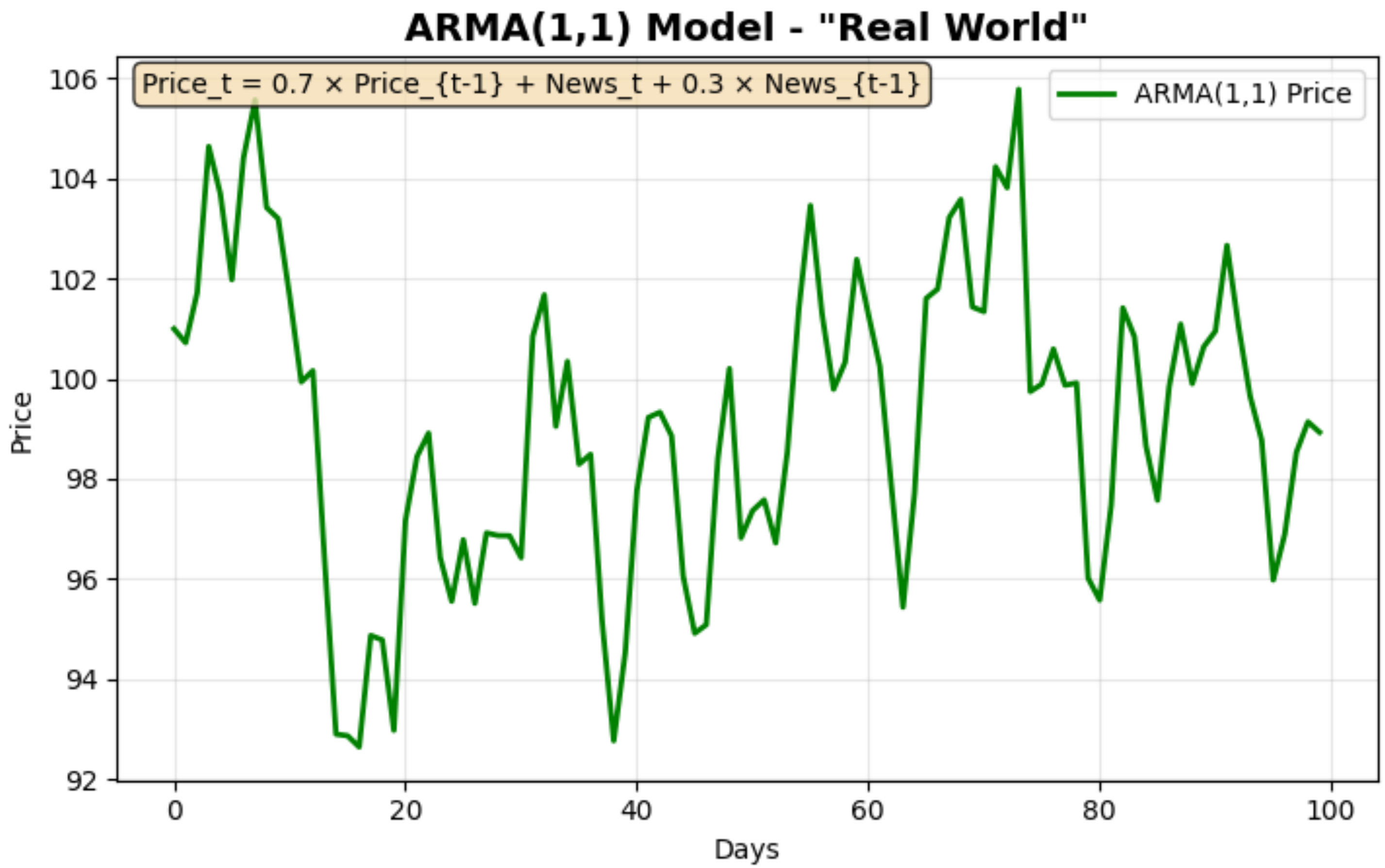

AR(1) Model - "Trend Momentum"
Price_t = 0.9 × Price_{t-1} + News_t
AR(1) Price
Price
101.00
100.75
100.50
100.25
100.00
99.75
99.50
99.25
99.00
0
20
40
60
80
100

Examples



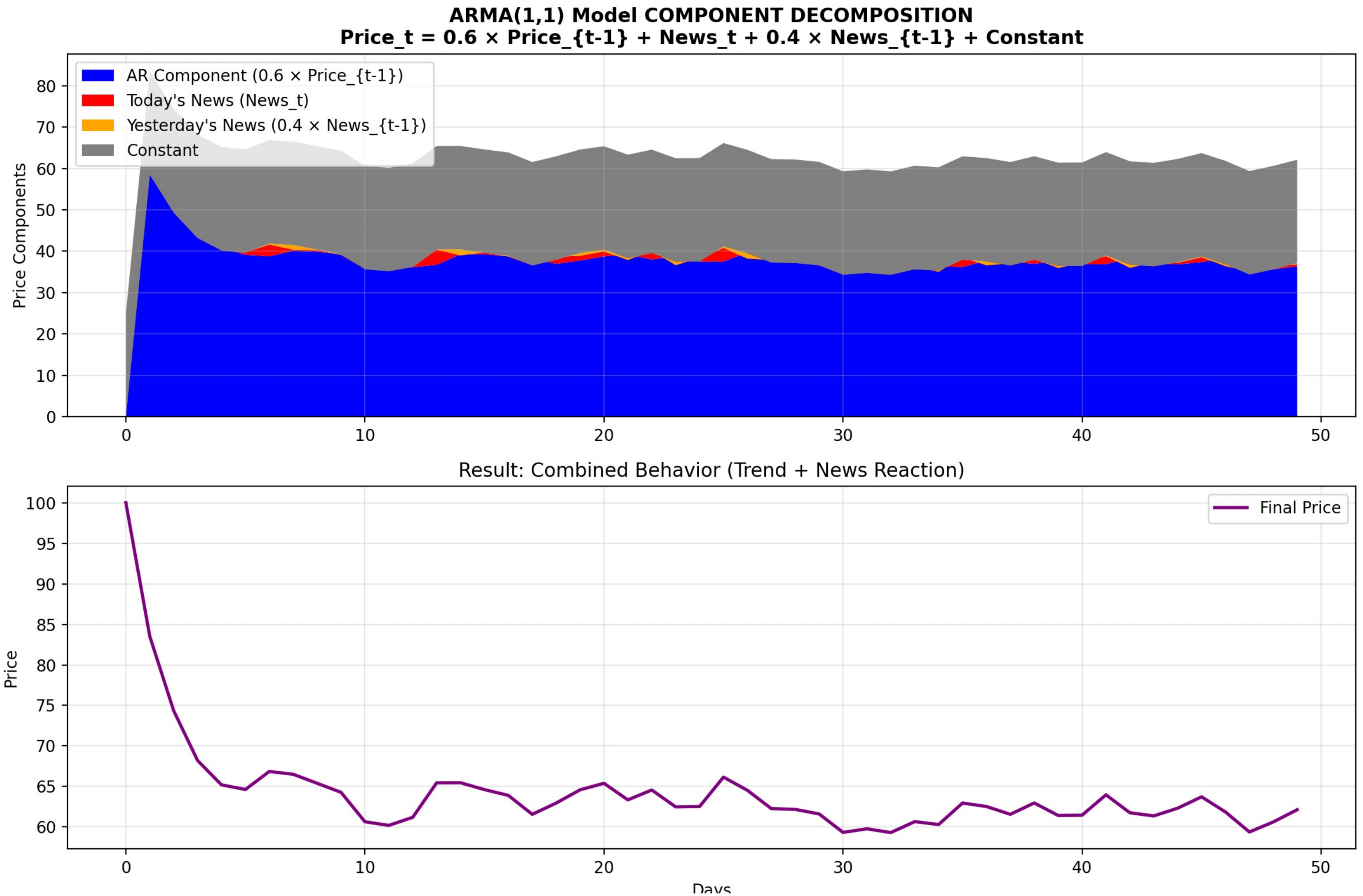
MA(1) Model - "News Reaction"
Price_t = Base + News_t + 0.6 × News_{t-1}
MA(1) Price
Price
104
102
100
98
96
94
0
20
40
60
80
100

Examples



ARMA(1,1) Model - "Real World"
Price_t = 0.7 × Price_{t-1} + News_t + 0.3 × News_{t-1}
ARMA(1,1) Price
106
104
102
100
98
96
94
92
Price
0
20
40
60
80
100
Days

Decomposition



ARMA(p,q) model

- **Can accurately approximate many stationary processes:**

  For any stationary process with autocovariance γ, and any k>0, there is an ARMA process $\{x_t\}$ for which

  Economic Data

  $$\gamma_x(h) = \gamma(h), \quad h = 0, 1, \ldots, k.$$

  ARMA model can infinitely approximate any other stationary process.

- **Intuition:**

  AR process: allow many coefficients different from zero, but with restrictions on the decay patterns

  MA process: permit a few coefficients different from zero with arbitrary values

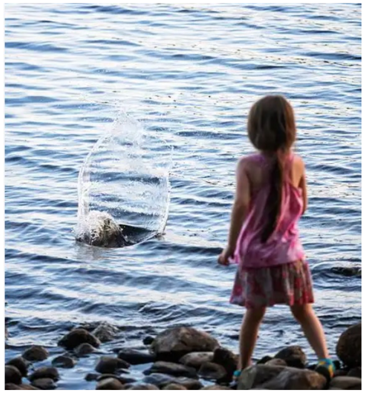
  ARMA process: combine the properties of AR and MA

ARMA(p,q) model

- **Intuition:**

  AR process: allow many coefficients different from zero, but with restrictions on the decay patterns

Today's value, x_t, can depend on values from the distant past.



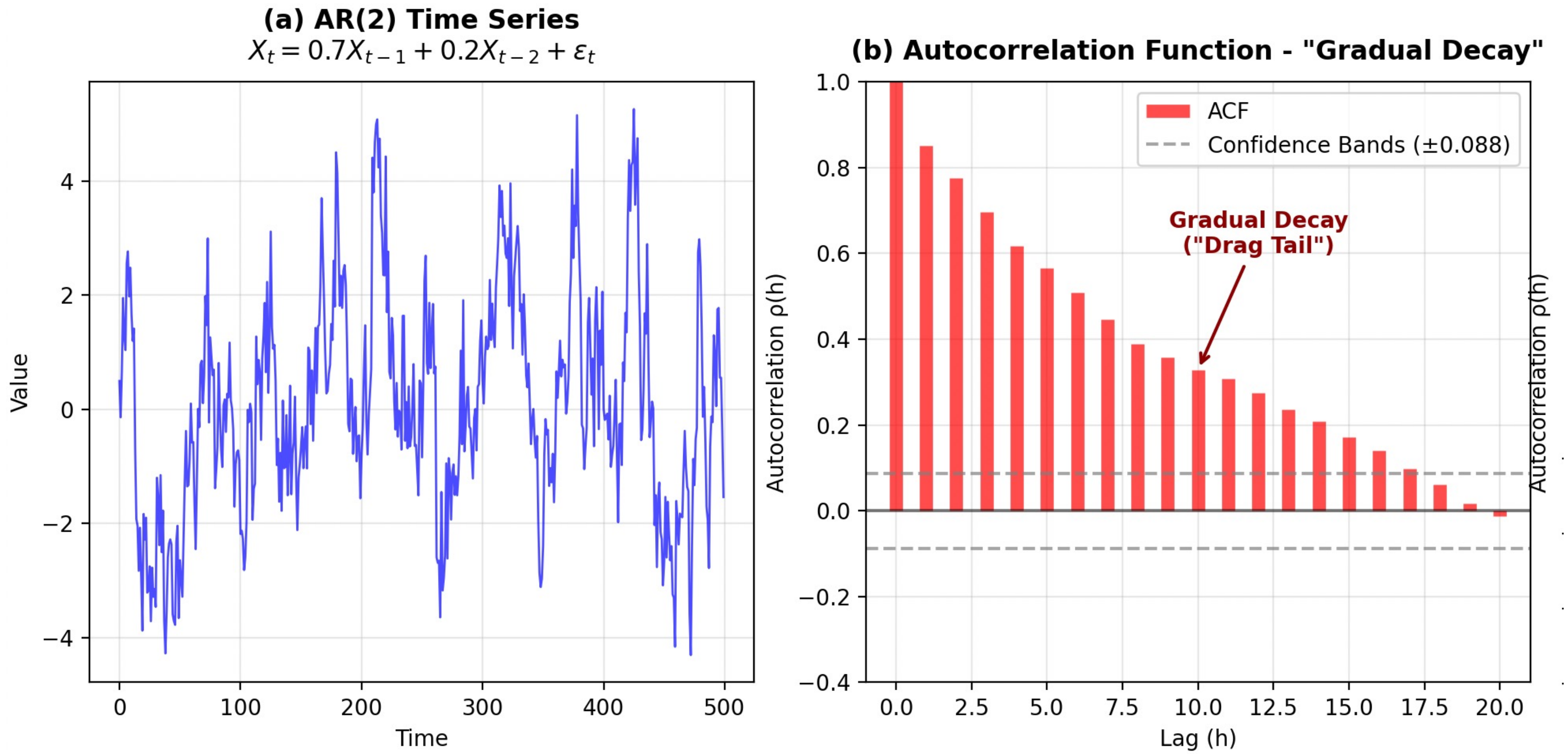
...but with restrictions on the decay patterns.

Although we can have many non-zero coefficients, these coefficients cannot take arbitrary values. To ensure the stationarity of the process, the sequence formed by these coefficients $\{\phi_1, \phi_2, \ldots, \phi_p\}$ must adhere to certain mathematical constraints, which cause their influence to exhibit a specific decay pattern (typically exponential decay or sinusoidal decay) as the lag order increases.

ARMA(p,q) model

- **Intuition:**

  AR process: allow many coefficients different from zero, but with restrictions on the decay patterns



ARMA(p,q) model

MA process: permit a few coefficients different from zero with arbitrary values

MA process: permit a few coefficients different from zero with arbitrary values

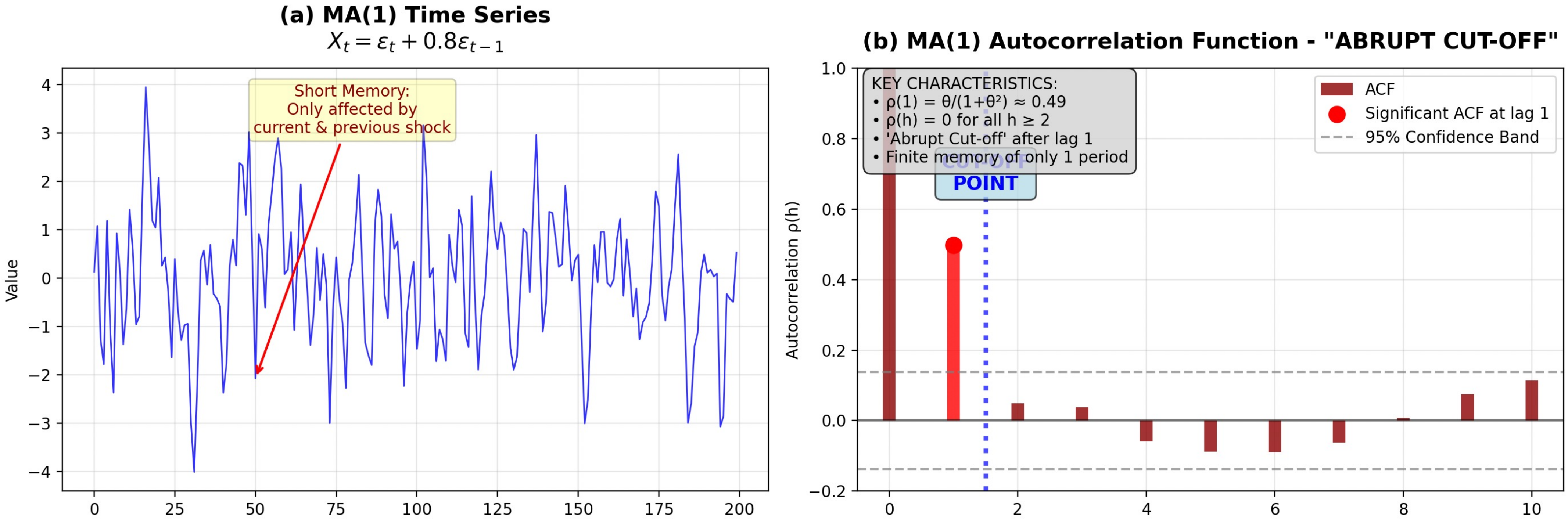
Meaning: The intuitive characteristic of the MA process is: it only allows a finite number (q) of non-zero coefficients, but the values of these coefficients can be arbitrary.

Explanation: For a pure MA(q) model, its autocorrelation function (ACF) **cuts off abruptly** (becomes zero) after lag q. This means its "memory" is limited. Its advantage lies in its flexibility—within q lags, it can generate autocorrelation patterns of any shape, as long as they truncate after lag q. However, its limitation is its short memory, making it ineffective for describing long-term dependencies.

Reason: When the lag $k > q$, x_t and x_{t-k} are composed of **completely independent** sets of shocks.
Manifestation: Cuts off abruptly (drops to zero) after lag q.

ARMA(p,q) model

MA process: permit a few coefficients different from zero with arbitrary values



ARMA(p,q) model

- **Parameter redundancy**

$$\phi(B)x_t = \theta(B)w_t$$

a common factor between the AR part and the MA part.

$$\text{with } \phi(B) = 1 - 0.5B, \quad \theta(B) = 1 - 0.5B$$

$$\Big\downarrow (1 - 0.5B)$$

$$x_t = \psi(B)w_t = \frac{\theta(B)}{\phi(B)} = w_t$$

ARMA(p,q) model

- **Parameter redundancy**

$$\phi(B)x_t = \theta(B)w_t$$

. The polynomials $\phi(z)$ and $\theta(z)$ share no common factors.

This guarantees that the resulting model is structurally unique, with identifiable parameters and of the lowest possible order.

AR Model (of order p)

- **An autoregressive model of order *p* (denoted as AR(*p*)), is of the form**

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

$$w_t \sim \mathrm{wn}(0, \sigma_w^2)$$

$$\phi_1, \phi_2, \ldots, \phi_p \text{ are constants } (\phi_p \neq 0)$$

"Order p" determines how many past data points the model looks back at when predicting today's value.

AR Model (of order p)

For an **AR(p)** model, the formula is:

$$\text{Current Value} = \varphi_1 * (\text{1st Lag Value}) + \varphi_2 * (\text{2nd Lag Value}) + ... + \varphi_p * (\text{p-th Lag Value}) + \text{Random Shock}$$

- `p = 1` **(1st Order)**:
  - Model: $\text{Today's Temperature} = \varphi_1 * \text{Yesterday's Temperature} + \text{Random Shock}$
  - Meaning: Only considers the influence of **yesterday** on today.
- `p = 2` **(2nd Order)**:
  - Model:

    $$\text{Today's Temperature} = \varphi_1 * \text{Yesterday's Temperature} + \varphi_2 * \text{Day Before Yesterday's Temperature} + \text{Random Shock}$$

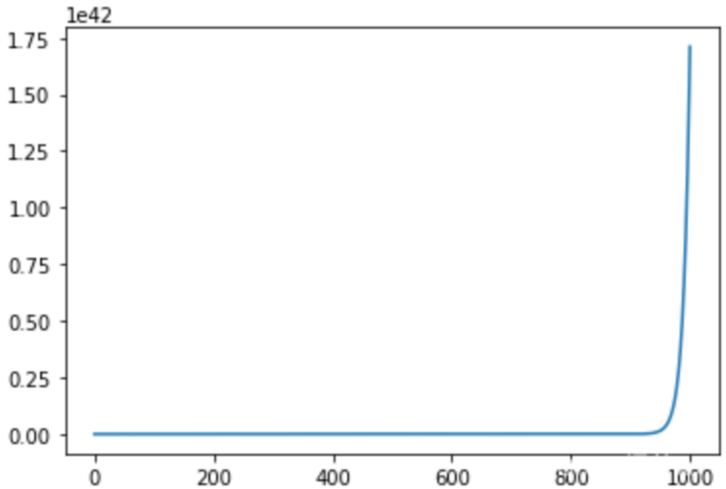
  - Meaning: Considers the influence of both **yesterday** and **the day before yesterday** on today.

AR Model (of order p)

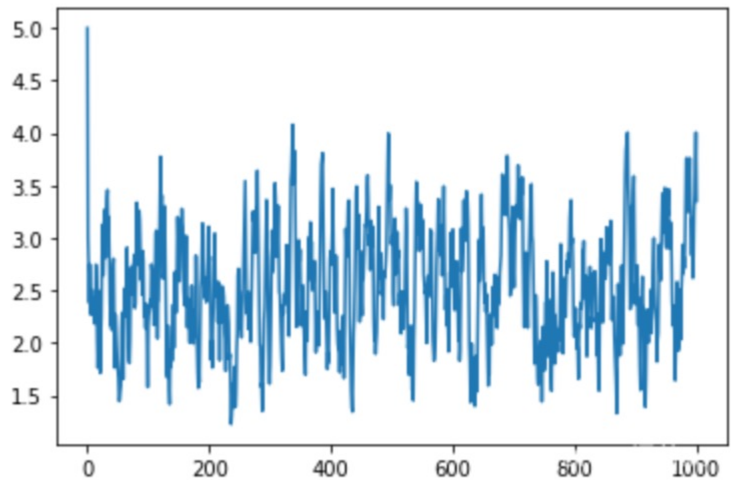
$$x_t = 1.1x_{t-1} + \epsilon_t \ldots\ldots (1)$$

$$x_t = x_{t-1} - 0.2x_{t-2} + \epsilon_t \ldots\ldots \quad (2)$$

```python
import random
import matplotlib.pyplot as plt
data_1 = [5,3]
data_2 = [5,3]
for i in range(1,1000):
    res_1 = 1.1*data_1[i]+random.random()
    res_2 = data_2[i]-0.2*data_2[i-1]+random.random()
    data_1.append(res_1)
    data_2.append(res_2)
plt.plot(data_1)
plt.plot(data_2)
```



non



Review

- **AR(p) model:**

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

$$\phi(B)x_t = w_t, \quad \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

AR polynomial

- **Stationary solution exists if and only if**

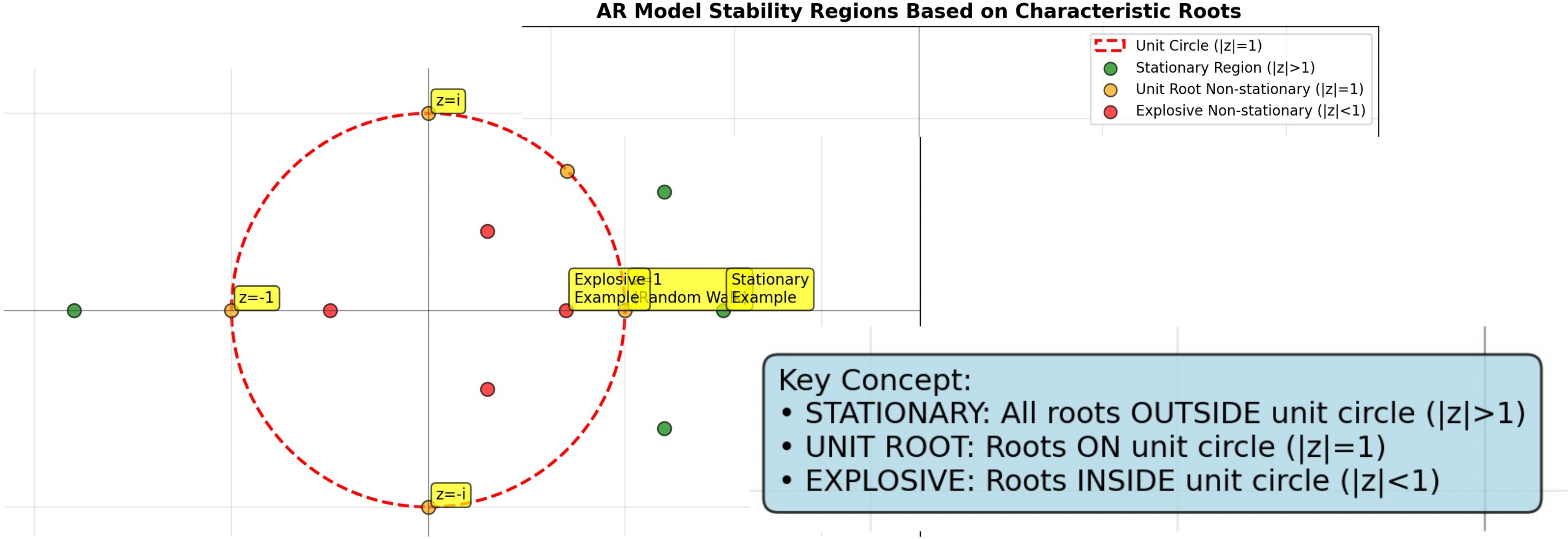
$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0$$

characteristic equation

AR Model (of order p)

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

Claim: An $AR(p)$ process is stationary if the roots of the characteristic polynomial lay *outside* the complex unit circle



AR Model (of order p)

Why is this condition related to stationarity?

Stationarity requires that the mean, variance, and covariance of the time series do not change over time. If the roots of the characteristic equation lie inside or on the unit circle (i.e., their modulus is less than or equal to 1), the model's process will exhibit explosive growth or decay, causing the variance to increase without bound, thus violating the condition for stationarity.

AR Model (of order p)

If the roots of the characteristic equation lie inside the unit circle (i.e., their modulus is less than 1), the model's shocks (white noise) will be "amplified," causing the variance of the predicted results to gradually increase, and preventing the variance from remaining finite, leading to instability in the model.

If the roots lie exactly on the unit circle (i.e., their modulus is equal to 1), the model will exhibit random walk or trending behavior, lacking stationarity.

If the roots lie outside the unit circle (i.e., their modulus is greater than 1), the model's shocks will gradually decay, ensuring that the variance of the time series remains within a finite range and does not explode, thus maintaining stationarity.

AR(1) model

- **An AR(1) process defined by** $x_t = \phi x_{t-1} + w_t$

$\phi(B)x_t = w_t$ with $\phi(B) = 1 - \phi B$

$$(1 - \phi B)x_t = w_t$$

$$(1 - \phi B) = 0 \qquad B = \frac{1}{\phi}$$

AR(1) model

- **An AR(1) process defined by**

  $\phi(B)x_t = w_t$ with $\phi(B) = 1 - \phi B$

  is causal if and only if

  $|\phi| < 1$

  or

  the root z_1 of the polynomial $\phi(z) = 1 - \phi z$ satisfies

  $|z_1| > 1$

Causal process

- **Causality** is a fundamental concept in time series analysis and signal processing.

A time series process is called **causal** if its current value depends **only** on:

- Present and **past** values of the input/shock process
- **Not** on future values

AR(p) model

- **Example: check causality**

$$x_t = 0.7x_{t-1} + 0.6x_{t-2} + w_t$$

$$x_t = -0.7x_{t-1} - 0.6x_{t-2} + w_t$$

For the given **AR(2) model**:

$$x_t = 0.7x_{t-1} + 0.6x_{t-2} + w_t$$

we want to determine whether it is **causal** or not. To do this, we analyze the **characteristic equation** of the model.

1. Characteristic Equation of AR(2)

The AR(2) model has coefficients $\phi_1 = 0.7$ and $\phi_2 = 0.6$, and can be written as:

$$(1 - 0.7B - 0.6B^2)x_t = w_t$$

where B is the lag operator.

2. Roots of the Characteristic Equation

To check causality, we solve the characteristic equation:

$$1 - 0.7B - 0.6B^2 = 0$$

This quadratic equation can be rearranged as:

$$0.6B^2 + 0.7B - 1 = 0$$

Using the quadratic formula, the roots are:

$$B = \frac{-0.7 \pm \sqrt{0.7^2 - 4 \times 0.6 \times (-1)}}{2 \times 0.6}$$

Calculate the discriminant:

$$\sqrt{0.49 + 2.4} = \sqrt{2.89} = 1.7$$

So the roots are:

$$B_1 = \frac{-0.7 + 1.7}{1.2} = \frac{1}{1.2} \approx 0.8333$$

$$B_2 = \frac{-0.7 - 1.7}{1.2} = \frac{-2.4}{1.2} = -2$$

3. Causality Condition

Here:

- $|B_1| \approx 0.8333 < 1$
- $|B_2| = 2 > 1$

Since one root lies **inside** the unit circle (less than 1 in absolute value), the process is **non-causal**.

4. Conclusion

The AR(2) model

$$x_t = 0.7x_{t-1} + 0.6x_{t-2} + w_t$$

is **non-causal** because one of the characteristic roots lies inside the unit circle, violating the causality condition.

AR(p) model

- **Example: check causality**

$$x_t = -0.7x_{t-1} - 0.6x_{t-2} + w_t$$

$$\phi(B) = 1 + 0.7B + 0.6B^2$$

$$\phi(z) = 1 + 0.7z + 0.6z^2 = 0$$

$$\Downarrow$$

$$z_{1,2} = -0.5833 \pm 1.1517\mathrm{i}$$

$$\Downarrow$$

Causal

Given:

$$x_t = -0.7x_{t-1} - 0.6x_{t-2} + w_t$$

The characteristic polynomial is:

$$\phi(B) = 1 + 0.7B + 0.6B^2$$

Step 1: Find roots of $\phi(B) = 0$:

$$1 + 0.7B + 0.6B^2 = 0$$

Rewrite as:

$$0.6B^2 + 0.7B + 1 = 0$$

Use the quadratic formula:

$$B = \frac{-0.7 \pm \sqrt{0.7^2 - 4 \times 0.6 \times 1}}{2 \times 0.6}$$

Calculate the discriminant:

$$0.7^2 - 4 \times 0.6 \times 1 = 0.49 - 2.4 = -1.91$$

Since the discriminant is **negative**, the roots are complex conjugates:

$$B = \frac{-0.7 \pm i\sqrt{1.91}}{1.2}$$

Calculate:

$$B = -0.5833 \pm 1.1517i$$

Step 2: Check the modulus (magnitude) of roots

$$|B| = \sqrt{(-0.5833)^2 + (1.1517)^2} = \sqrt{0.3402 + 1.3264} = \sqrt{1.6666} \approx 1.29$$

Step 3: Causality condition

- Since both roots have modulus $\approx 1.29 > 1$, the roots lie **outside** the unit circle.
- Therefore, the AR(2) process is **causal**.

AR(p) model

- **Example: check causality**

$$x_t = 0.7x_{t-1} + 0.6x_{t-2} + w_t$$

$$\phi(B) = 1 - 0.7B - 0.6B^2$$

$$\phi(z) = 1 - 0.7z - 0.6z^2 = 0$$

$$z_1 = -2, \quad z_2 = 0.8333$$

Not Causal

MA(1) model

- **MA(1) model:**

$$x_t = w_t + \theta w_{t-1}$$

White Noise

$$E[w_t] = 0, \quad \mathrm{Var}(w_t) = \sigma_w^2$$

$$\mathrm{Cov}(w_t, w_s) = 0 \quad t \neq s$$

- **Mean, autocovariance and autocorrelation function:**

$$\mu = 0; \quad \gamma(h) = \begin{cases} (1+\theta^2)\sigma_w^2, & h = 0 \\ \theta\sigma_w^2, & h = 1 \\ 0, & h > 1 \end{cases}$$

$$\rho(h) = \begin{cases} \frac{\theta}{1+\theta^2}, & h = 1 \\ 0, & h > 1 \end{cases}$$

MA(1) model

- **MA(1) model:**

$$x_t = w_t + \theta w_{t-1}$$

- **Mean, autocovariance and autocorrelation function:**

$$\mu = 0; \quad \gamma(h) = \begin{cases} (1+\theta^2)\sigma_w^2, & h = 0 \\ \theta\sigma_w^2, & h = 1 \\ 0, & h > 1 \end{cases}$$

$$\mu = E[x_t] = E[w_t + \theta w_{t-1}] = E[w_t] + \theta E[w_{t-1}] = 0 + \theta \cdot 0 = 0$$

MA(1) model

- **MA(1) model:**

$$x_t = w_t + \theta w_{t-1}$$

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma_w^2, & h = 0 \\ \theta\sigma_w^2, & h = 1 \\ 0, & h > 1 \end{cases}$$

$$\begin{aligned} \gamma(0) &= \mathrm{Var}(x_t) = \mathrm{Var}(w_t + \theta w_{t-1}) \\ &= \mathrm{Var}(w_t) + \theta^2 \mathrm{Var}(w_{t-1}) + 2\theta \mathrm{Cov}(w_t, w_{t-1}) \\ &= \sigma_w^2 + \theta^2\sigma_w^2 + 0 \\ &= (1+\theta^2)\sigma_w^2 \end{aligned}$$

$$\begin{aligned} \gamma(1) &= \mathrm{Cov}(x_t, x_{t+1}) \\ &= \mathrm{Cov}(w_t + \theta w_{t-1},\ w_{t+1} + \theta w_t) \\ &= \mathrm{Cov}(w_t, \theta w_t) + \mathrm{Cov}(\theta w_{t-1}, w_{t+1}) + \ldots \\ &= \theta \mathrm{Cov}(w_t, w_t) \\ &= \theta\sigma_w^2 \end{aligned}$$

$$\gamma(h) = 0 \qquad h > 1$$

The most important property of the MA(1) model is its **short memory**: its autocorrelation function **cuts off** exactly after lag 1 (becomes precisely zero).

MA(1) model

- **MA(1) model:**

$$x_t = w_t + \theta w_{t-1}$$

Step 2: Substitute into the covariance formula

$$\begin{aligned}\gamma(1) &= \mathrm{Cov}(x_t, x_{t+1}) \\ &= \mathrm{Cov}(w_t + \theta w_{t-1},\ w_{t+1} + \theta w_t)\end{aligned}$$

Step 3: The Key Step - Apply the bilinearity property of covariance

Covariance has an important property: $\mathrm{Cov}(A + B, C + D) = \mathrm{Cov}(A, C) + \mathrm{Cov}(A, D) + \mathrm{Cov}(B, C) + \mathrm{Cov}(B, D)$

Applying this property, we get:

$$\gamma(1) = \mathrm{Cov}(w_t, w_{t+1}) + \mathrm{Cov}(w_t, \theta w_t) + \mathrm{Cov}(\theta w_{t-1}, w_{t+1}) + \mathrm{Cov}(\theta w_{t-1}, \theta w_t)$$

MA(1) model

- **MA(1) model:**

$$x_t = w_t + \theta w_{t-1}$$

Step 4: Analyze each term (This is the core of understanding)

Let's examine these four covariance terms one by one:

1. $\text{Cov}(w_t, w_{t+1})$:
   - These are white noise terms at different times.
   - By the definition of white noise (serially uncorrelated), this term **equals 0**.
2. $\text{Cov}(w_t, \theta w_t)$:
   - This is the covariance of the same white noise term w_t with itself.
   - $\text{Cov}(w_t, \theta w_t) = \theta \text{Cov}(w_t, w_t) = \theta \cdot \text{Var}(w_t) = \theta \sigma_w^2$
   - This term **is NOT 0** and equals $\theta \sigma_w^2$.

MA(1) model

- **MA(1) model:**

  $x_t = w_t + \theta w_{t-1}$

3. $\mathrm{Cov}(\theta w_{t-1}, w_{t+1})$:
   - These are w_{t-1} and w_{t+1}, which are two time periods apart.
   - By the definition of white noise, this term **equals 0**.
4. $\mathrm{Cov}(\theta w_{t-1}, \theta w_t)$:
   - These are w_{t-1} and w_t, which are adjacent but still **different** white noise terms.
   - By the definition of white noise, this term **equals 0**.

Step 5: Combine the results

All the zero terms drop out, leaving only the second term:

$$\gamma(1) = 0 + \theta\sigma_w^2 + 0 + 0 = \theta\sigma_w^2$$

MA(1) model

- **MA(1) model:**

$$x_t = w_t + \theta w_{t-1}$$

$$B w_t = w_{t-1}$$

$$x_t = (1 + \theta B) w_t$$

Express w_t in terms of x_t

$$w_t = \frac{1}{(1 + \theta B)} x_t \qquad w_t = (1 + \theta B)^{-1} x_t$$

Operator inversion (applying the geometric series formula).

$$\begin{aligned}(1 + \theta B)^{-1} &= 1 - \theta B + (-\theta)^2 B^2 + (-\theta)^3 B^3 + \cdots \\ &= 1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \theta^4 B^4 - \cdots\end{aligned}$$

MA(1) model

- **Express w_t in terms of x_t?**

$$x_t = w_t + \theta w_{t-1}$$

$$x_t = (1 + \theta B) w_t$$

To ensure that this series converges (i.e., the sum is finite), we require $|\theta| < 1$.

when $|\theta| < 1$

$$\begin{aligned} w_t &= (1 + \theta B)^{-1} x_t \\ &= (1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \cdots) x_t \\ &= \sum_{j=0}^{\infty} (-\theta)^j x_{t-j} \end{aligned}$$

MA(1) model

$$
\begin{aligned}
w_t &= (1+\theta B)^{-1} x_t \\
&= (1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \cdots) x_t \\
&= \sum_{j=0}^{\infty} (-\theta)^j x_{t-j}
\end{aligned}
$$

The coefficient of the first term x_t is 1,

The coefficient of the second term x_{t-1} is $-\theta$,

The coefficient of the third term x_{t-2} is θ^2,

and so on.

This expansion shows that w_t (the white noise) is a weighted sum of the past observations x_t, where the weights decrease according to $(-\theta)^j$ as j increases.

MA(1) model

Why is this expansion important?

This expansion helps us understand how to recover the white noise w_t using past observations x_t.

It also demonstrates the causality of the MA(1) model: when $|\theta| < 1$, the white noise w_t can be expressed in terms of past observations, making the model causal.

Invertibility

- **A linear process $\{x_t\}$ is invertible (strictly, a invertible function of $\{w_t\}$) if there is a**

$$\pi(B) = \pi_0 + \pi_1 B + \pi_2 B^2 + \cdots$$

with $\sum_{j=0}^{\infty} |\pi_j| < \infty$ such that

$$w_t = \pi(B) x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}$$

MA(1) model

- **An MA(1) process defined by**

  $x_t = \theta(B)w_t$ with $\theta(B) = 1 + \theta B$

  is invertible if and only if

  $|\theta| < 1$

  or

  the root z_1 of the polynomial $\theta(z) = 1 + \theta z$ satisfies

  $|z_1| > 1$

MA(

The invertibility of an MA(q) process requires that the generation process of its white noise can be expressed in terms of past observations. If the model satisfies the invertibility condition, we can invert the process and express the current observation (x_t) as a linear combination of past white noise (w_t).

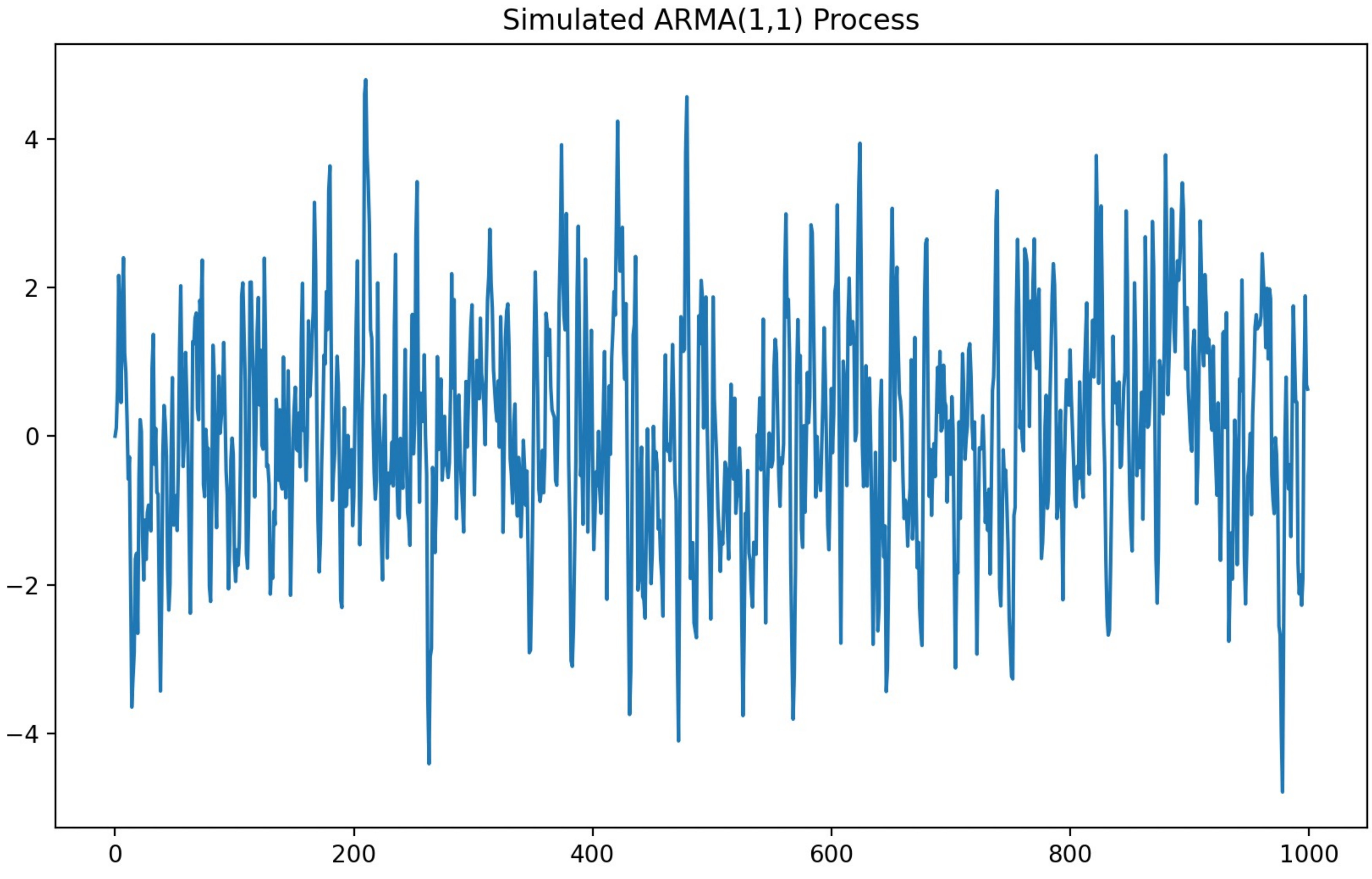
Here are some basic steps to fit these models using **Python** (for example, using the `statsmodels` library):

1. **Generate Data**:
   - Generate data using the parameters of an ARMA model. You can use `numpy` to generate white noise and then use the model formula to generate the time series.
2. **Fit the Model**:
   - Use `statsmodels`'s `AR`, `ARMA`, or `ARIMA` classes to fit the model. This library provides tools to estimate model parameters and test the fit of the model.
3. **Model Evaluation**:
   - Evaluate the model’s fit using statistics like AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion).
   - Perform residual analysis to check if the residuals are white noise, as they should be.

```python
import numpy as np
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt

# Generate ARMA(1, 1) process data
np.random.seed(42)
n = 1000  # number of data points
phi = 0.5  # AR coefficient
theta = 0.5  # MA coefficient
w = np.random.normal(size=n)  # white noise

# Generate ARMA(1, 1) data
x = np.zeros(n)
for t in range(1, n):
    x[t] = phi * x[t-1] + w[t] + theta * w[t-1]
```

```python
# Plot the generated data
plt.figure(figsize=(10, 6))
plt.plot(x)
plt.title('Simulated ARMA(1,1) Process')
plt.show()

# Fit ARMA(1, 1) model
model = sm.tsa.ARMA(x, order=(1, 1))
fitted_model = model.fit()

# Output the model's fit summary
print(fitted_model.summary())
```

$$x_t = \phi x_{t-1} + \theta w_{t-1} + w_t$$


Simulated ARMA(1,1) Process
4
2
0
−2
−4
0
200
400
600
800
1000

```
                              ARIMA Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1000
Model:                 ARIMA(1, 0, 1)   Log-Likelihood:              -1400.531
Method:                       css-mle   AIC:                          2807.063
Date:                Fri, 01 Oct 2025   BIC:                          2830.129
Sample:                      0   - 1000   HQIC:                         2816.887
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0483      0.031      1.567      0.118      -0.013       0.110
ar.L1.y        0.4689      0.048      9.741      0.000       0.374       0.564
ma.L1.y       -0.2787      0.048     -5.803      0.000      -0.373      -0.184
==============================================================================
Fitted AR(1) coefficient (phi): 0.4689
Fitted MA(1) coefficient (theta): -0.2787
```