# PACF

## Partial Autocorrelation Function

# Autoregressive models

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

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

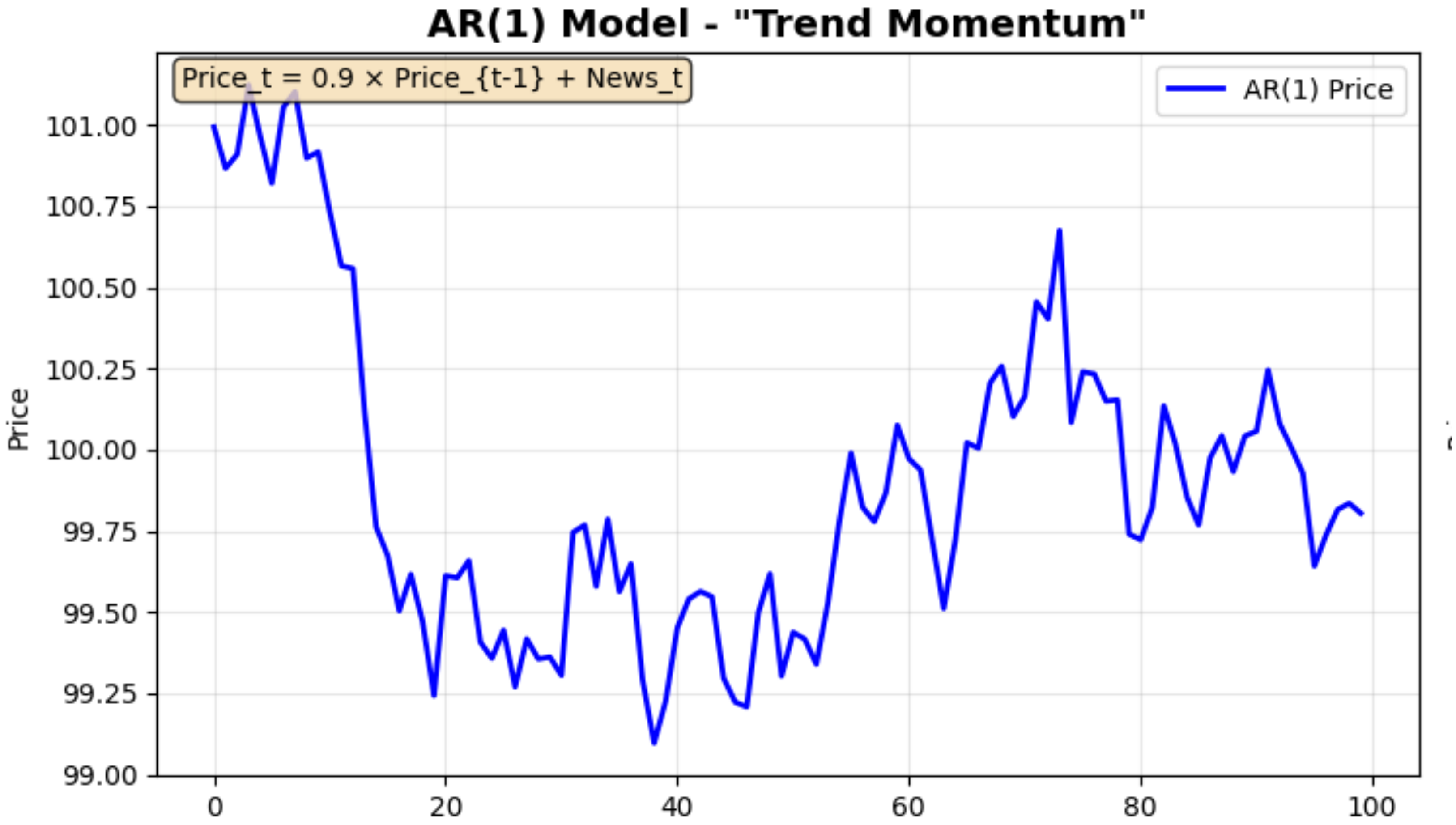
$$w_t \sim \mathrm{wn}(0, \sigma_w^2)$$

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

$$\Downarrow$$

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

# 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

# Moving average models

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

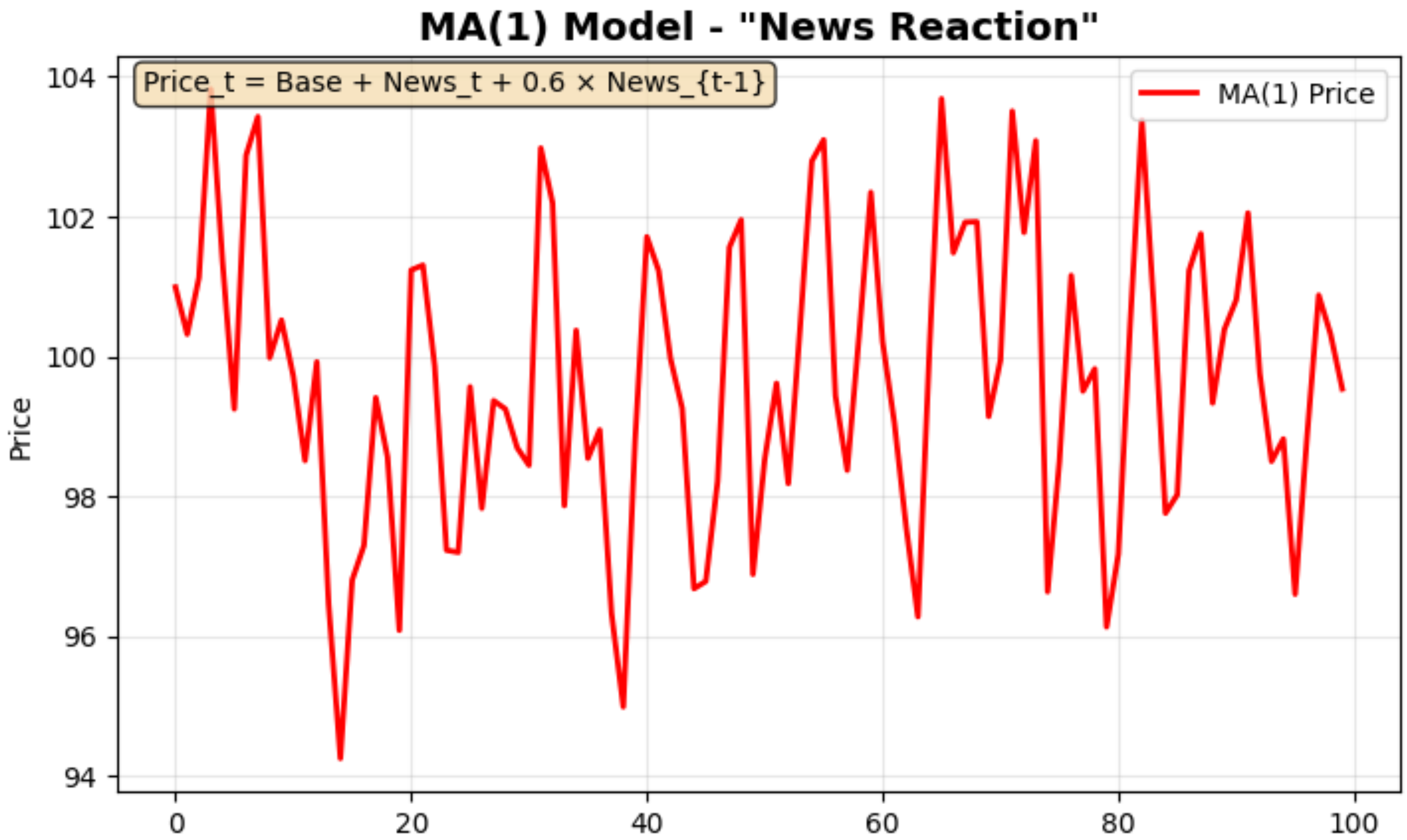
$$w_t \sim \mathrm{wn}(0, \sigma_w^2),$$

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

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

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

# 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

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."*

# 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

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

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

$$\Downarrow$$

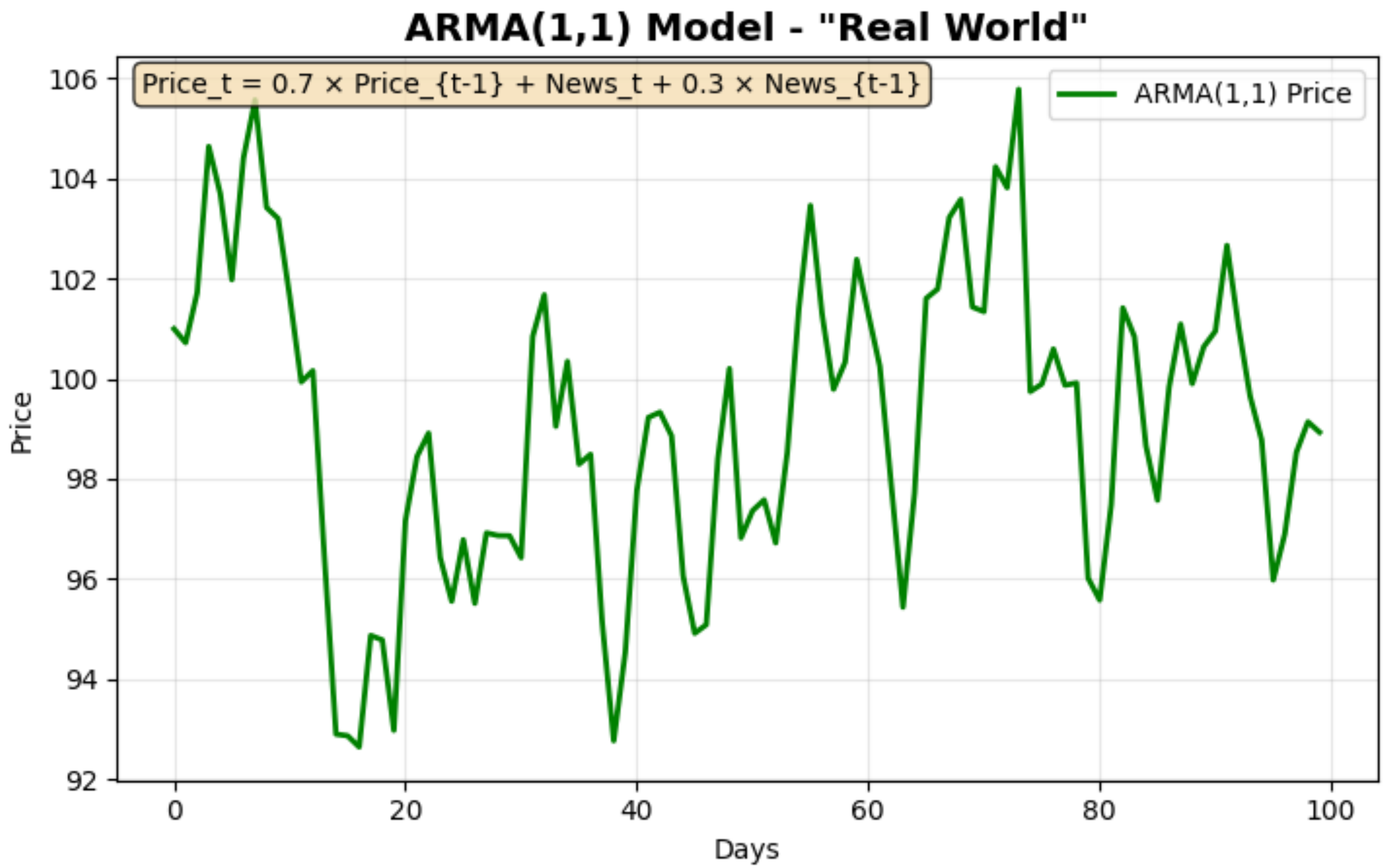
$$\phi(B)x_t = \theta(B)w_t$$

- **Special cases:**

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

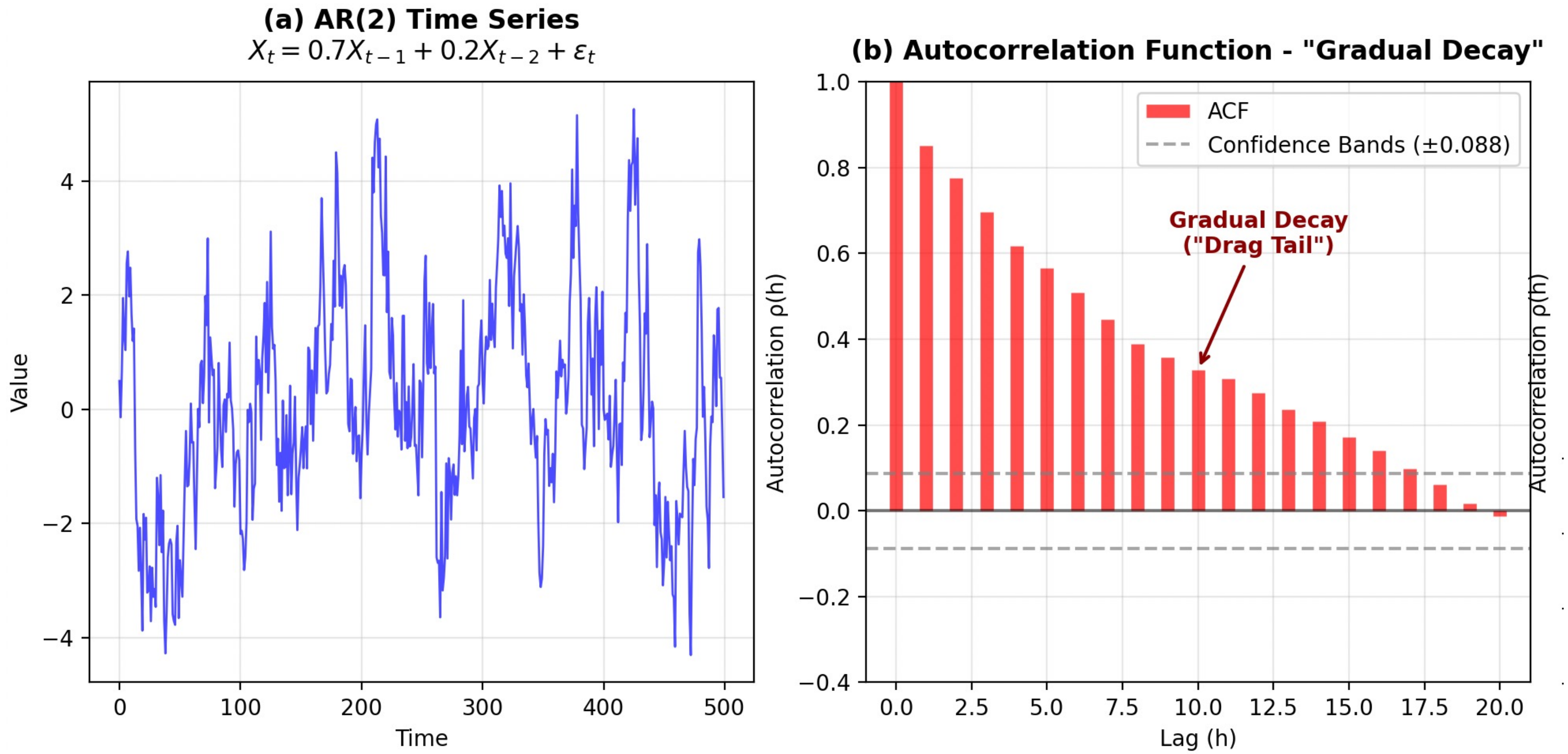
  MA(q) = ARMA(0,q) with $\phi(B) = 1$

# Examples

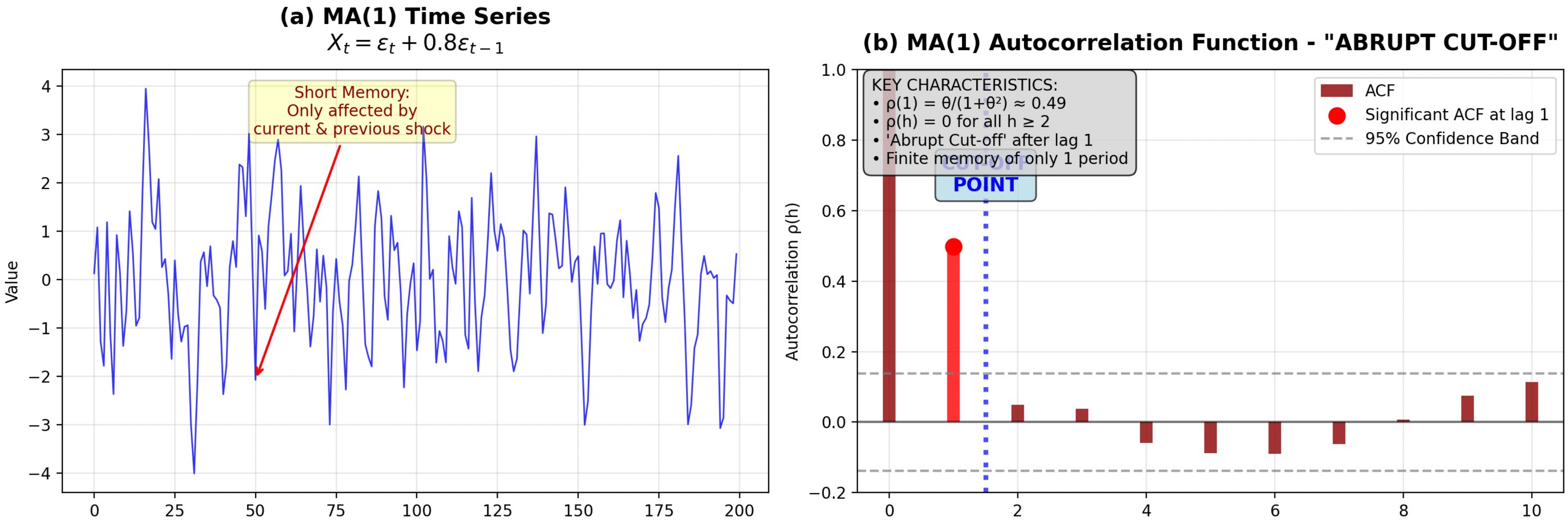

ARMA(1,1) Model - "Real World"
Price_t = 0.7 × Price_{t-1} + News_t + 0.3 × News_{t-1}
ARMA(1,1) Price
106
104
102
100
98
96
94
92
Price
0
20
40
60
80
100
Days

- **Intuition:**

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



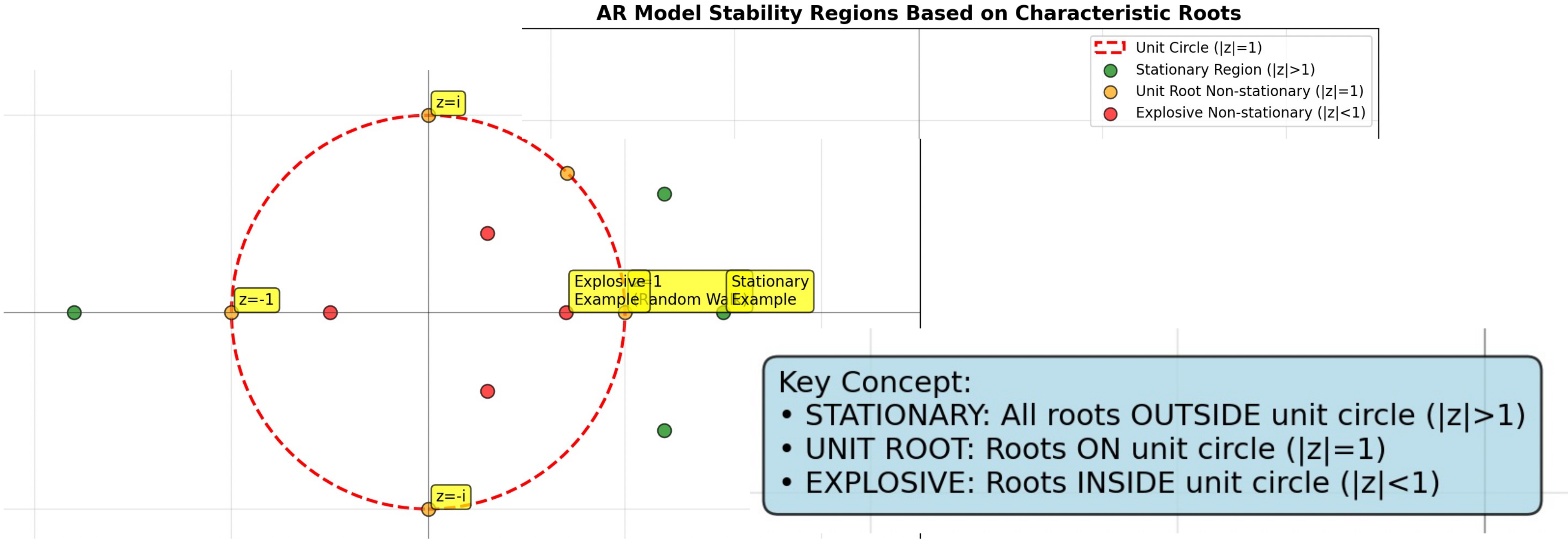
MA process: permit a few coefficients different from zero with arbitrary values



# 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



# 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:**

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

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

# 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

  $|z_1| > 1$

# Review

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

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

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

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

# Review

- **MA(q) model:**

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

$$\Downarrow$$

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

- **Mean and autocovariance function:**

$$\mu = 0, \quad \gamma(h) = \begin{cases} \sigma_w^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & |h| \le q \\ 0, & |h| > q \end{cases}$$

# Review

- **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

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

# Review

- **MA(q) model:**

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

$$\Downarrow$$

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

- **The MA(q) process is invertible if and only if**

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q = 0 \quad \Rightarrow \quad |z| > 1$$

**invertible if and only if all roots of its moving average polynomial lie outside the unit circle in the complex plane (i.e., the modulus of all roots is greater than 1).**

# Review

- **When the MA(q) is invertible:**

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

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

- **Matching coefficients to find** $\pi(B)$

# Review

- **When the MA(q) is invertible:**

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

$$\Downarrow$$

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

- **Matching coefficients to find** $\pi(B)$

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

- $B^0$: $\pi_0 = 1$
- $B^1$: $\pi_1 + \theta\pi_0 = 0 \Rightarrow \pi_1 + \theta = 0 \Rightarrow \pi_1 = -\theta$
- $B^2$: $\pi_2 + \theta\pi_1 = 0 \Rightarrow \pi_2 + \theta(-\theta) = 0 \Rightarrow \pi_2 = \theta^2$
- $B^3$: $\pi_3 + \theta\pi_2 = 0 \Rightarrow \pi_3 + \theta(\theta^2) = 0 \Rightarrow \pi_3 = -\theta^3$

# ARMA(p,q) model

- **Causality**

  The ARMA(p,q) process is causal if and only if the roots of $\phi(z)$ are outside the unit circle:

$$|z| \leq 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$$

# ARMA(p,q) model

- **Invertibility**

  The ARMA(p,q) process is invertible if and only if the roots of $\theta(z)$ are outside the unit circle:

$$|z| \leq 1 \Rightarrow \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \neq 0$$

# ARMA(p,q) model

- **Example:**

$$x_t = 0.4x_{t-1} + 0.45x_{t-2} + w_t + w_{t-1} + 0.25w_{t-2}$$

# ARMA(p,q) model

- **Example:**

$$x_t = 0.4x_{t-1} + 0.45x_{t-2} + w_t + w_{t-1} + 0.25w_{t-2}$$

$$(1 - 0.4B - 0.45B^2)x_t = (1 + B + 0.25B^2)w_t$$

$$\Downarrow$$

$$(1 - 0.9B)(1 + 0.5B)x_t = (1 + 0.5B)^2 w_t$$

$$\Downarrow$$

$$(1 - 0.9B)x_t = (1 + 0.5B)w_t$$

Causal and Invertible

# ARMA(p,q) model

- **Example:**

$$x_t = 0.4x_{t-1} + 0.45x_{t-2} + w_t + w_{t-1} + 0.25w_{t-2} \quad Bx_t = x_{t-1}$$

$$(1 - 0.4B - 0.45B^2)x_t = (1 + B + 0.25B^2)w_t$$

$$1 - 0.4B - 0.45B^2 = (1 - aB)(1 - bB)$$

$$(1 - aB)(1 - bB) = 1 - bB - aB + abB^2$$

$$= 1 - (a + b)B + abB^2$$

$$a + b = 0.4$$

$$ab = 0.45$$

$$1 - 0.4B - 0.45B^2 = 1 - (a + b)B + abB^2$$

# ARMA(p,q) model

- **Example:**

$$x_t = 0.4x_{t-1} + 0.45x_{t-2} + w_t + w_{t-1} + 0.25w_{t-2} \quad Bx_t = x_{t-1}$$

$$(1 - 0.4B - 0.45B^2)x_t = (1 + B + 0.25B^2)w_t$$

$$1 - 0.4B - 0.45B^2 = 1 - (a+b)B + abB^2$$

$$a + b = 0.4$$

$$ab = 0.45$$

- z₁ = (0.4 + 1.4) / 2 = 1.8 / 2 = 0.9
- z₂ = (0.4 − 1.4) / 2 = −1.0 / 2 = −0.5

```
z = [0.4 ± √(0.16 + 1.8)] / 2
= [0.4 ± √1.96] / 2
= [0.4 ± 1.4] / 2
```

```
(1 − 0.9B)(1 + 0.5B)
```

# ARMA(p,q) model

- **Example:**

$$x_t = 0.4x_{t-1} + 0.45x_{t-2} + w_t + w_{t-1} + 0.25w_{t-2}$$

$$(1 - 0.4B - 0.45B^2)x_t = (1 + B + 0.25B^2)w_t$$

$$\Downarrow$$

$$(1 - 0.9B)(1 + 0.5B)x_t = (1 + 0.5B)^2 w_t$$

$$\Downarrow$$

$$(1 - 0.9B)x_t = (1 + 0.5B)w_t$$

Causal and Invertible

# ARMA(p,q) model

- **Convert to MA process:**

  For a causal ARMA(p,q) model, we may write

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$$

  can use matching coefficients to find $\psi(B)$

$$\psi(B) = \frac{\theta(B)}{\phi(B)} \quad \Rightarrow \quad \phi(B)\psi(B) = \theta(B)$$

# ARMA(p,q) model

- **Convert to MA process:**

$$\phi(B)\psi(B) = \theta(B)$$

$$(1 - \phi_1 z - \phi_2 z^2 - \cdots)(\psi_0 + \psi_1 z + \psi_2 z^2 + \cdots) = 1 + \theta_1 z + \theta_2 z^2 + \cdots$$

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 - \phi_1\psi_0 &= \theta_1 \\ \psi_2 - \phi_1\psi_1 - \phi_2\psi_0 &= \theta_2 \\ &\vdots \end{aligned}$$

# ARMA(p,q) model

- **Convert to MA process:**

$$\begin{cases} \psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k} = 0, & j \geq \max(p, q+1) \\ \psi_j - \sum_{k=1}^{j} \phi_k \psi_{j-k} = \theta_j, & 0 \leq j < \max(p, q+1) \end{cases}$$

with $\phi_j = 0$ for $j > p$ and $\theta_j = 0$ for $j > q$

homogenous difference equation with initial conditions

# ARMA(p,q) model

- **Example: convert ARMA to MA:**

$$x_t = 0.9x_{t-1} + 0.5w_{t-1} + w_t$$

For the ARMA(p, q) model, we can repeatedly expand the autoregressive part $(1 - \phi_1 B - \cdots - \phi_p B^p)$ until it becomes a recursive formula in terms of the noise $w_t$, thus obtaining a purely MA process.

we want to write the equation in the form of a pure MA process:

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$$

# ARMA(p,q) model

- **Example: convert ARMA to MA:**

$$x_t = 0.9x_{t-1} + 0.5w_{t-1} + w_t$$

$$(1 - 0.9B)x_t = (1 + 0.5B)w_t$$

$$(1 - 0.9B)(\psi_0 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots) = 1 + 0.5B$$

$$(1 - 0.9B)^{-1} = 1 + 0.9B + 0.9^2 B^2 + 0.9^3 B^3 + \dots$$

$$x_t = (1 + 0.9B + 0.9^2 B^2 + \dots) \cdot (1 + 0.5B)w_t$$

# ARMA(p,q) model

- **Example: convert ARMA to MA:**

$$x_t = 0.9x_{t-1} + 0.5w_{t-1} + w_t$$

$$x_t = w_t + 0.5Bw_t + 0.9Bw_t + 0.9 \cdot 0.5B^2w_t + 0.9^2B^2w_t + \dots$$

$$x_t = w_t + 1.4Bw_t + 1.26B^2w_t + 1.134B^3w_t + \dots$$

$$\downarrow$$

$$\psi_0 = 1$$

$$\begin{aligned} \psi_1 - 0.9\psi_0 &= 0.5 \\ \psi_2 - 0.9\psi_1 &= 0 \\ \psi_3 - 0.9\psi_2 &= 0 \end{aligned} \quad \rightarrow \quad \begin{aligned} \psi_1 &= 1.4 \\ \psi_2 &= 1.26 \\ \psi_3 &= 1.134 \end{aligned}$$

# ARMA(p,q): autocovariance function

- **Approach 1:**

$$\phi(B)x_t = \theta(B)w_t \quad \Rightarrow \quad x_t = \psi(B)w_t$$

$$\Downarrow$$

$$\gamma(h) = \sigma_w^2(\psi_0\psi_h + \psi_1\psi_{h+1} + \psi_2\psi_{h+2} + \cdots)$$

# PACF: the partial autocorrelation function

- **Motivation:**

  For MA(q) models, the ACF will be zero for lags greater than q, and will not be zero at lag q.

**Example:** In an MA(1) model, you will only observe autocorrelation at lag 1, while autocorrelations at lag 2 and beyond are all zero.

However, the ACF alone does not tell us much for ARMA or AR models

Instead exhibit a gradual decay

# PACF: the partial autocorrelation function

- **Motivation:**

  For MA(q) models, the ACF will be zero for lags greater than q, and will not be zero at lag q.

  However, the ACF alone does not tell us much for ARMA or AR models

  For a causal AR(1) model: $\gamma(2) = \phi^2\gamma(0) \neq 0$ as $x_{t-2}$ is dependent on $x_t$ through $x_{t-1}$

**The Limitation of ACF**

In AR(1), cov(xt, xt-2) ≠ 0, indicates that the autocorrelation (ACF) between xt and xt-2 at lag 2 is not zero. However, this correlation is spurious, as it arises because both variables are connected through the intervening variable, xt-1.

**Therefore**

To remove the effect of $x_{t-1}$, we have $\mathrm{cov}(x_t - \phi x_{t-1}, x_{t-2} - \phi x_{t-1}) = 0$

# PACF: the partial autocorrelation function

- **Motivation:**

  For MA(q) models, the ACF will be zero for lags greater than q, and will not be zero at lag q.

  However, the ACF alone does not tell us much for ARMA or AR models

  For a causal AR(1) model: $\gamma(2) = \phi^2\gamma(0) \neq 0$ as $x_{t-2}$ is dependent on $x_t$ through $x_{t-1}$

  To remove the effect of $x_{t-1}$, we have $\operatorname{cov}(x_t - \phi x_{t-1}, x_{t-2} - \phi x_{t-1}) = 0$

The partial autocorrelation function is defined as the conditional correlation between $x_t$ and $x_{t-k}$, after controlling for the effects of the intermediate variables $x_{t-1}, x_{t-2}, \ldots, x_{t-k+1}$.

(Removing the linear dependence of both variables on the intermediate observations)

# PACF: the partial autocorrelation function

- **Notations for mean-zero stationary time series:**

$\hat{x}_{t+h}$ : regression of $x_{t+h}$ on $\{x_{t+h-1}, x_{t+h-2}, \ldots, x_{t+1}\}$ as

$$\hat{x}_{t+h} = \beta_1 x_{t+h-1} + \beta_2 x_{t+h-2} + \cdots + \beta_{h-1} x_{t+1}$$

$\hat{x}_t$ : regression of $x_t$ on $\{x_{t+h-1}, x_{t+h-2}, \ldots, x_{t+1}\}$ as

$$\hat{x}_t = \beta_1 x_{t+1} + \beta_2 x_{t+2} + \cdots + \beta_{h-1} x_{t+h-1}$$

# PACF: the partial autocorrelation function

- **Definition:**

The partial autocorrelation function (PACF) of a stationary process, $x_t$, denoted $\phi_{hh}$, for $h = 1, 2, \ldots,$ is

$$\phi_{11} = \text{corr}(x_{t+1}, x_t) = \rho(1),$$
$$\phi_{hh} = \text{corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t), \ h \geq 2$$

With a lag of 1, the partial autocorrelation is equal to the autocorrelation, as there are no intermediate values between $x_{t+1}$ and $x_t$. This directly measures the one-step dependency within the series.

the predicted value obtained from

**$\phi_1$ x_t+h-1 + $\phi_2$ x_t+h-2 + ... + $\phi$_h-1 x_t+1**

# PACF: the partial autocorrelation function

- **Example:**

$$x_t = \phi x_{t-1} + w_t, \ |\phi| < 1$$

**AR(1)**

**PACF at lag 2 ($\phi_{22}$)**

# PACF: the partial autocorrelation function

- **Example:**

$$x_t = \phi x_{t-1} + w_t, \ |\phi| < 1$$

$$\hat{x}_t = \alpha x_{t+1}$$

$$\Downarrow$$

$$\text{minimize } \mathrm{E}(x_t - \hat{x}_t)^2 = \mathrm{E}(x_t - \alpha x_{t+1})^2$$

$$= \gamma(0) - 2\alpha\gamma(1) + \alpha^2\gamma(0)$$

$$\Downarrow$$

$$\alpha = \frac{\gamma(1)}{\gamma(0)} = \phi$$

"Optimal" means that the coefficient $\alpha$ we choose should minimize the difference between the predicted value $\hat{x}_t = \alpha x_{t+1}$ and the true value $x_t$. Mathematically, we measure this difference using the Mean Square Error (MSE):

$\mathrm{MSE} = E[\,(x_t - \hat{x}_t)^2\,] = E[\,(x_t - \alpha x_{t+1})^2\,]$

Our goal is to find the value of $\alpha$ that minimizes this MSE. This problem is formally expressed as:

$\min E(\, x_t - \alpha x_{t+1}\,)^2$
with respect to $\alpha$

$$\text{minimize } \mathrm{E}(x_t - \hat{x}_t)^2 = \mathrm{E}(x_t - \alpha x_{t+1})^2$$
$$= \gamma(0) - 2\alpha\gamma(1) + \alpha^2\gamma(0)$$

$$\alpha = \frac{\gamma(1)}{\gamma(0)} = \phi$$

$$\text{minimize } \mathrm{E}(x_t - \hat{x}_t)^2 = \mathrm{E}(x_t - \alpha x_{t+1})^2$$

$$= \gamma(0) - 2\alpha\gamma(1) + \alpha^2\gamma(0)$$

$$\alpha = \frac{\gamma(1)}{\gamma(0)} = \phi$$

**Objective**: Minimize the mean square error

$$\min E(x_t - \hat{x}_t)^2 = E(x_t - \alpha x_{t+1})^2$$

**Expand the squared term**:

$$E(x_t - \alpha x_{t+1})^2 = E[x_t^2 - 2\alpha x_t x_{t+1} + \alpha^2 x_{t+1}^2]$$

**Apply the linearity property of expectation**:

$$= E[x_t^2] - 2\alpha E[x_t x_{t+1}] + \alpha^2 E[x_{t+1}^2]$$

$$\text{minimize } \mathrm{E}(x_t - \hat{x}_t)^2 = \mathrm{E}(x_t - \alpha x_{t+1})^2$$
$$= \gamma(0) - 2\alpha\gamma(1) + \alpha^2\gamma(0)$$

$$\alpha = \frac{\gamma(1)}{\gamma(0)} = \phi$$

Thus, the minimization problem becomes:

$$\min_{\alpha} \left[\gamma(0) - 2\alpha\gamma(1) + \alpha^2\gamma(0)\right]$$

**Substitute the autocovariance function** (for a zero-mean stationary series):

- $E[x_t^2] = \gamma(0)$
- $E[x_t x_{t+1}] = \gamma(1)$
- $E[x_{t+1}^2] = \gamma(0)$

**Obtain**:

$$E(x_t - \alpha x_{t+1})^2 = \gamma(0) - 2\alpha\gamma(1) + \alpha^2\gamma(0)$$

**Express in simplified notation**:

$$(0) - 2\alpha(1) + \alpha^2(0)$$

# PACF: the partial autocorrelation function

- **Example:**

$$x_t = \phi x_{t-1} + w_t, \ |\phi| < 1$$

$$\hat{x}_t = \alpha x_{t+1}$$

$$\Downarrow$$

$$\alpha = \frac{\gamma(1)}{\gamma(0)} = \phi$$

The role of this section: It serves as a lemma. It proves that in an AR(1) process, the optimal regression coefficient for predicting $x_t$ using $x_{t+1}$ is exactly the model's own parameter $\phi$. This implies that $\hat{x}_t = \phi x_{t+1}$ is the optimal predictor.

# PACF: the partial autocorrelation function

- **Example:**

$$x_t = \phi x_{t-1} + w_t,\ |\phi| < 1$$

In an AR(1) model, the optimal linear predictor of $x_t$ using $x_{t+1}$ is $\hat{x}_t = \phi x_{t+1}$.

$$\begin{aligned}\phi_{22} &= \operatorname{corr}(x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t) \\ &= \operatorname{corr}(x_{t+2} - \phi x_{t+1}, x_t - \phi x_{t+1}) \\ &= \operatorname{corr}(w_{t+2}, x_t - \phi x_{t+1}) = 0\end{aligned}$$

$$\phi_{hh} = 0 \quad h > 2$$

# PACF: the partial autocorrelation function

- **Example:**

$$x_t = \phi x_{t-1} + w_t,\ |\phi| < 1$$

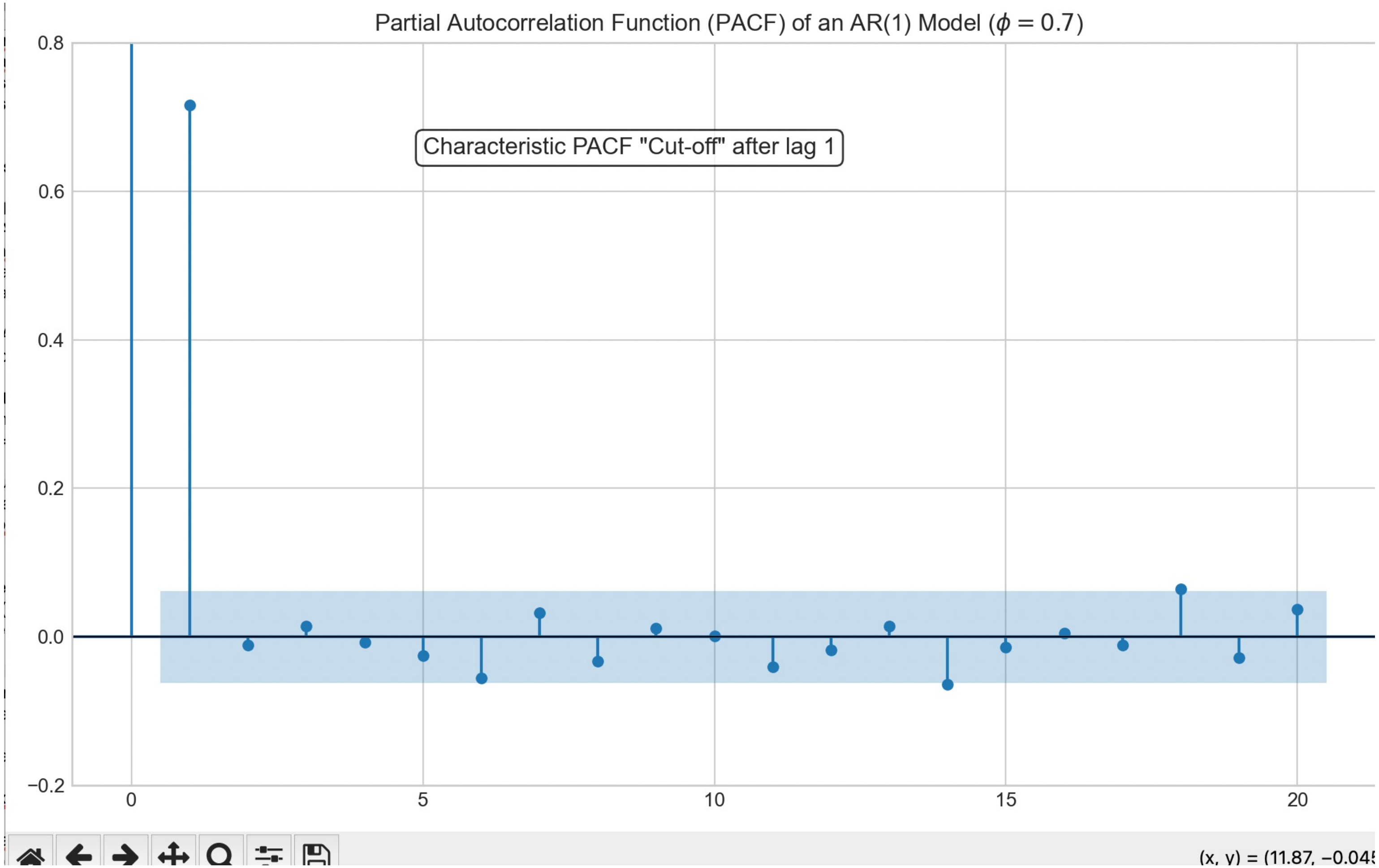
$$\begin{aligned}\phi_{22} &= \mathrm{corr}(x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t) \\ &= \mathrm{corr}(x_{t+2} - \phi x_{t+1}, x_t - \phi x_{t+1}) \\ &= \mathrm{corr}(w_{t+2}, x_t - \phi x_{t+1}) = 0\end{aligned}$$

$$\phi_{hh} = 0 \quad h > 2$$

In an AR(1) model, a core assumption is that the random shocks $w_t$ are independent and identically distributed (i.i.d.) and uncorrelated with all past values.

This implies:

- $w_{t+2}$ is uncorrelated with all random shocks at time $t+1$ and earlier ($w_{t+1}, w_t, w_{t-1}, \ldots$).
- $x_t$ and $x_{t+1}$ are both composed of random shocks up to and including time $t+1$.
- Therefore, $w_{t+2}$ is uncorrelated with any expression formed from past shocks, including $x_t - \phi x_{t+1}$.




**For an AR(1) model, all partial autocorrelation coefficients (PACF) at lag 2 and higher are equal to zero.**

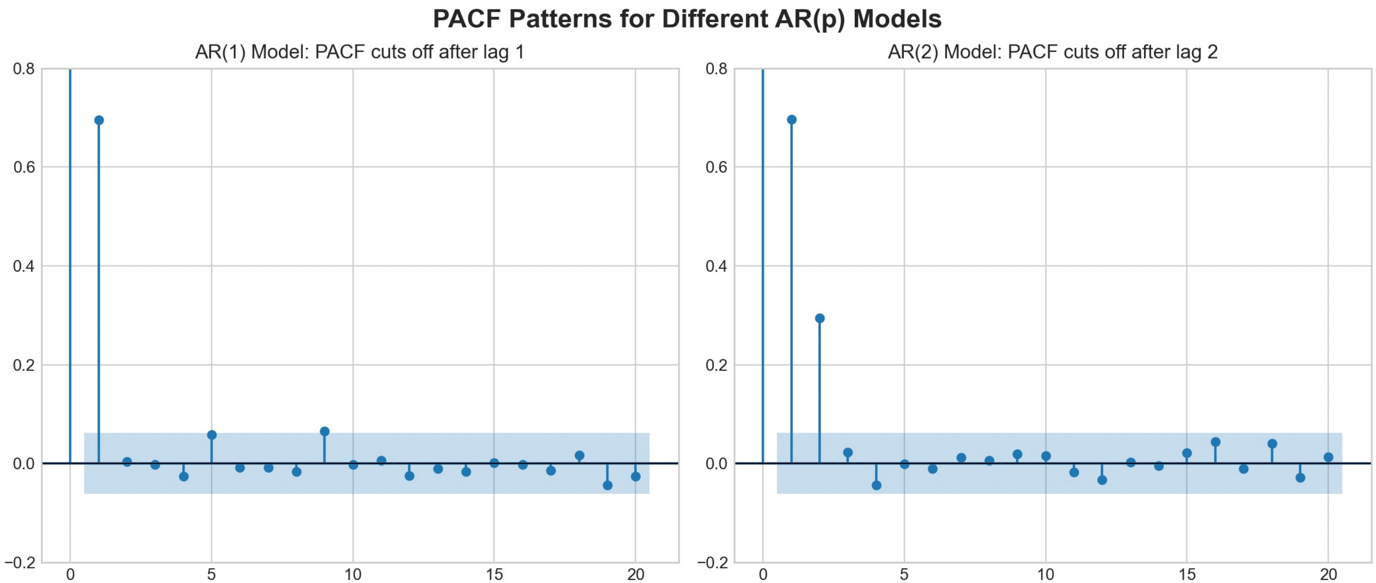
**This pattern is referred to as the PACF "cutting off" after lag 1.**

# PACF: the partial autocorrelation function

- **Example:**

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

**For an AR(p) model, the precise property of its partial autocorrelation function (PACF) is that it "cuts off" after lag p.**

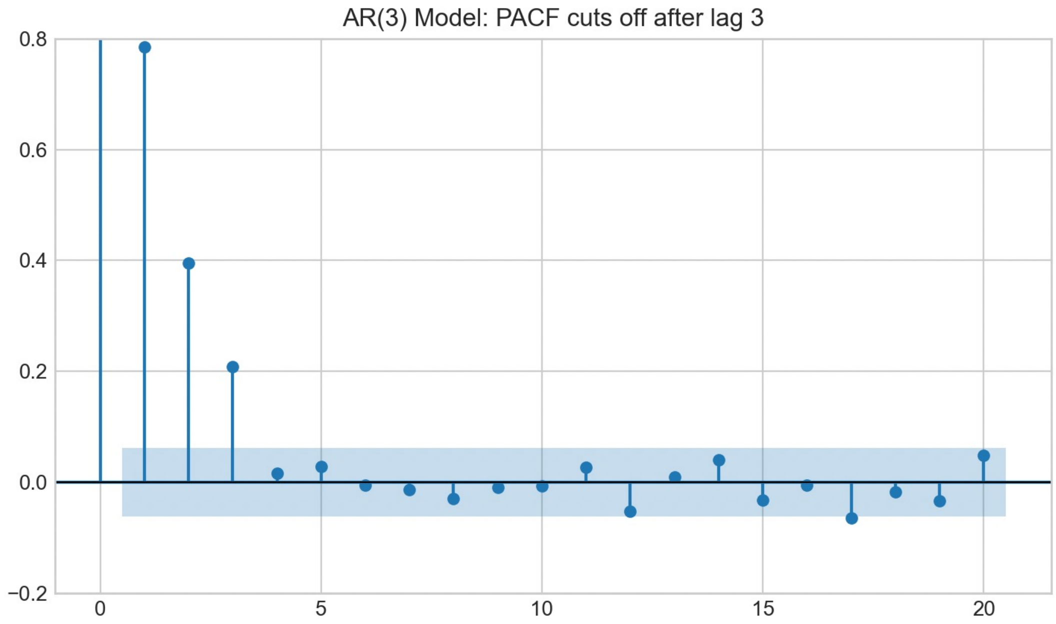


# PACF: the partial autocorrelation function

- **Example:**

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

**For an AR(p) model, the precise property of its partial autocorrelation function (PACF) is that it "cuts off" after lag p.**




PACF Cut-Off Property:

• AR(1): Significant at lag 1 only
• AR(2): Significant at lags 1-2 only
• AR(3): Significant at lags 1-3 only

The PACF 'cuts off' after lag p
for an AR(p) model, meaning:
φ_hh ≈ 0 for all h > p

This is the key identifying
characteristic of AR processes.

# PACF: the partial autocorrelation function

- **Example:**

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

When $h > p$, the regression of $x_{t+h}$ on $\{x_{t+1}, \ldots, x_{t+h-1}\}$ is

$$\hat{x}_{t+h} = \phi_1 x_{t+h-1} + \phi_2 x_{t+h-2} + \cdots + \phi_p x_{t+h-p}$$

$$x_{t+h} - \hat{x}_{t+h} = w_{t+h}$$

$$\Downarrow$$

$$\phi_{hh} = \text{corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t) = \text{corr}(w_{t+h}, x_t - \hat{x}_t) = 0$$

# PACF: the partial autocorrelation function

- **Example:**

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

# PACF: the partial autocorrelation function

- **Example:**

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

$$\phi_{11} = \rho(1) = \frac{\theta}{1+\theta^2}$$

**for an MA(1) process, when using only a single variable (**xt+1*xt*+1**) for prediction, the optimal linear prediction coefficient is precisely their correlation coefficient** ρ(1) **---**
Only a single observation xt+1

$$\hat{x}_{t+2} = \rho(1)x_{t+1} = \frac{\theta}{1+\theta^2}x_{t+1}$$

$$\hat{x}_t = \rho(1)x_{t+1} = \frac{\theta}{1+\theta^2}x_{t+1}$$

$$\mathrm{cov}(x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t) = \gamma(2) + \rho(1)^2\gamma(0) - 2\rho(1)\gamma(1) = \frac{-\theta^2}{1+\theta^2}\sigma_w^2$$

# PACF: the partial autocorrelation function

- **Example:**

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

$$\phi_{22} = \mathrm{corr}(x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t) = \frac{\mathrm{cov}(x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t)}{\sqrt{\mathrm{var}(x_{t+2} - \hat{x}_{t+2})\mathrm{var}(x_t - \hat{x}_t)}}$$

$$= \frac{-\theta^2}{1 + \theta^2 + \theta^4}$$

$$\phi_{hh} = -\frac{(-\theta)^h (1 - \theta^2)}{1 - \theta^{2(h+1)}}$$

# PACF: the partial autocorrelation function

### Deriving the General PACF Formula for MA(1)

This calculation can be generalized to any lag order $h$.

For an MA(1) model, the general formula for the partial autocorrelation function is:

$$\phi_{hh} = -\frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2(h+1)}} \quad \text{for } h \geq 1$$

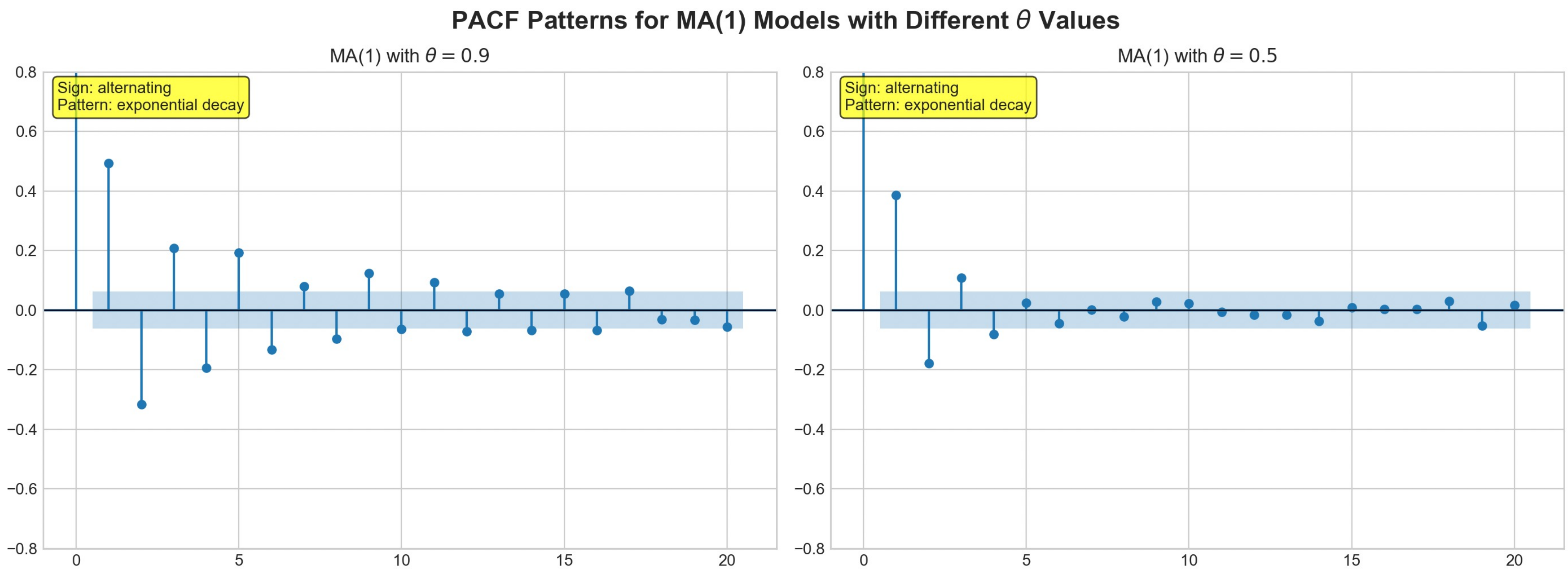
When $|\theta| < 1$ and $h$ is large, this formula approximates to:

$$\phi_{hh} \approx -(-\theta)^h(1-\theta^2)$$

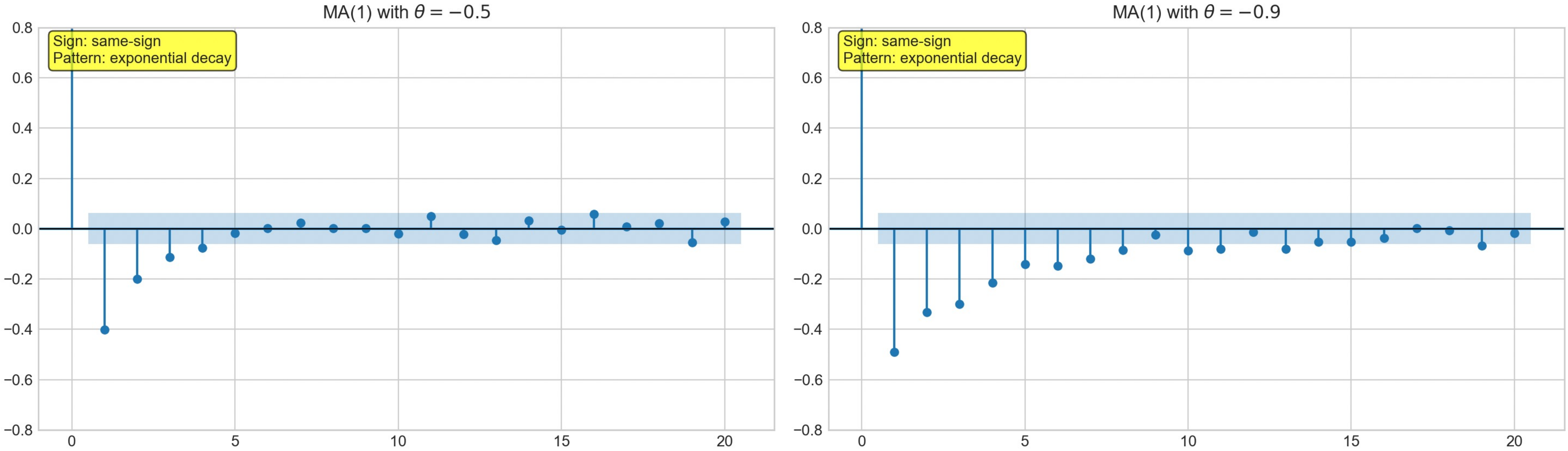
**Key Conclusions:**

- The PACF of an MA(1) model is **tail-dwelling**, meaning it decays towards zero at an exponential rate and does **not** abruptly cut off like an AR model's PACF.
- The sign of the PACF **alternates** (if $\theta > 0$, then $\phi_{11} < 0, \phi_{22} > 0, \phi_{33} < 0, \ldots$).

# PACF: the partial autocorrelation function



# PACF: the partial autocorrelation function



# PACF: the partial autocorrelation function

**Table 3.1.** Behavior of the ACF and PACF for ARMA models

| | AR($p$) | MA($q$) | ARMA($p, q$) |
|---|---|---|---|
| ACF | Tails off | Cuts off after lag $q$ | Tails off |
| PACF | Cuts off after lag $p$ | Tails off | Tails off |

*Table.3.1 in "Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer. 4th Edition.*

# Forecasting

- **Objective:**

  Predict future values of a time series, $x_{n+m}$, $m = 1, 2, \dots,$ based on the data collected to present, $x_{1:n} = \{x_1, x_2, \dots, x_n\}$

- **Mean square prediction error:**

  $$E(x_{n+m} - x_{n+m}^n)^2$$

- **Minimum mean square error predictor:**

  $$x_{n+m}^n = E(x_{n+m} | x_{1:n})$$

- **Objective:**

  Predict future values of a time series, $x_{n+m},\ m = 1, 2, \ldots,$ based on the data collected to present, $x_{1:n} = \{x_1, x_2, \ldots, x_n\}$

- **Linear predictor:**

  $$x_{n+m}^n = \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k$$

- **Best linear predictors (BLPs):**

  Linear predictors that minimize the mean square prediction error

  Can be found by solving

  $\mathrm{E}[(x_{n+m} - x_{n+m}^n)x_k] = 0,\ k = 0, 1, 2, \ldots, n,$

  where $x_0 = 1$

# Linear predictor

- **Predictors of the form:**

$$x_{n+m}^{n} = \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k$$

- **Best linear predictors (BLPs):**

  Linear predictors that minimize the mean square prediction error

  For Gaussian process, minimum mean square error predictors and best linear predictors are the same