Time-series,
Spring, 2026

Stationarity – Unit Root Tests – Differencing – AR / MA / ARMA models (Cont.)

Faculty of DS & AI
Spring semester, 2026

Trong-Nghia Nguyen

Business AI Lab

Content

- ARMA Models
- ACF & PACF for Model Selection

Content

- ARMA Models
- ACF & PACF for Model Selection

ARMA Models

ARMA (p, q) Model

- ARMA combines:
  - **Autoregressive (AR)**: dependence on past values
  - **Moving Average (MA)**: dependence on past shocks
- Designed for **stationary time series**

Mathematical form:

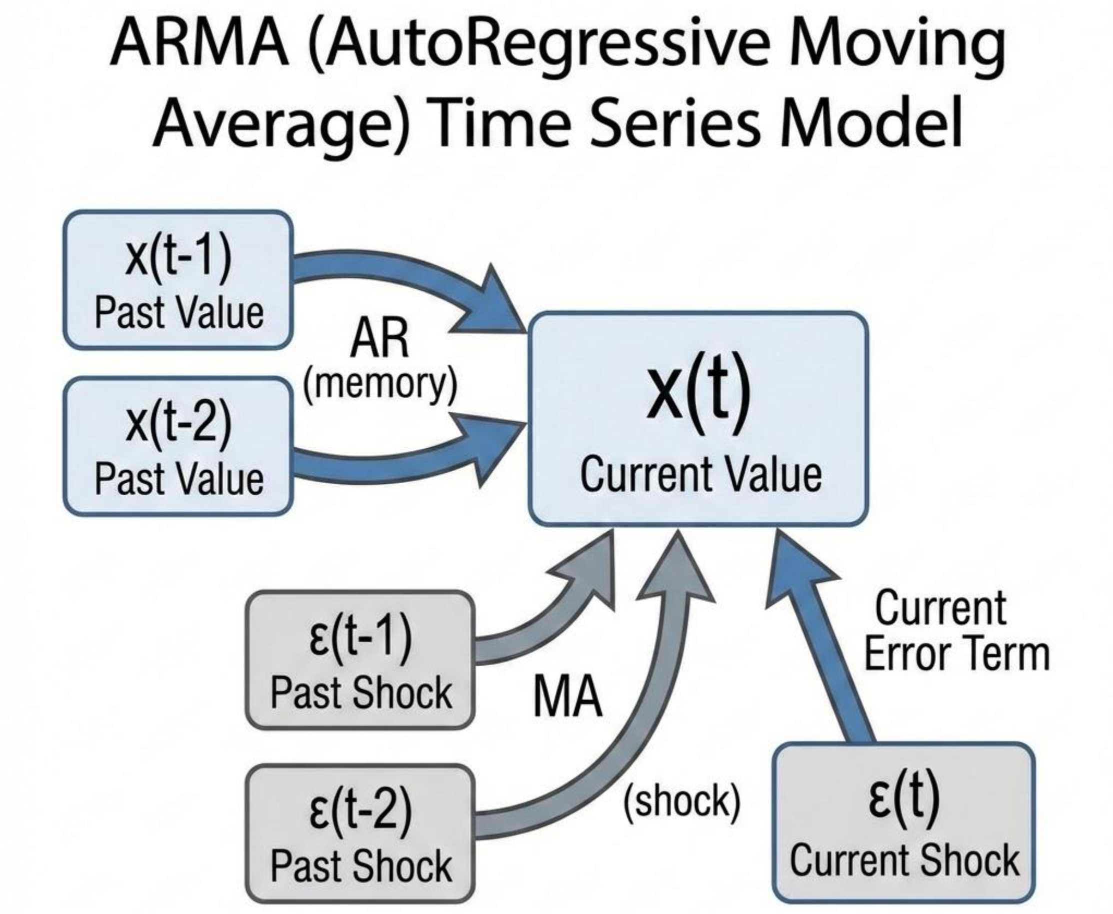
$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

Where:

- ϕ_i: AR coefficients (memory)
- θ_j: MA coefficients (shock effects)
- ε_t: white noise

Key idea:

> ARMA captures both **persistence** and **short-term disturbances** in a time series.



$$x(t) = \alpha + \varphi_1 x(t-1) + \varphi_2 x(t-2) + \theta_1 \varepsilon(t-1) + \theta_2 \varepsilon(t-2) + \varepsilon(t)$$

ARMA Models

ARMA (p, q) Model

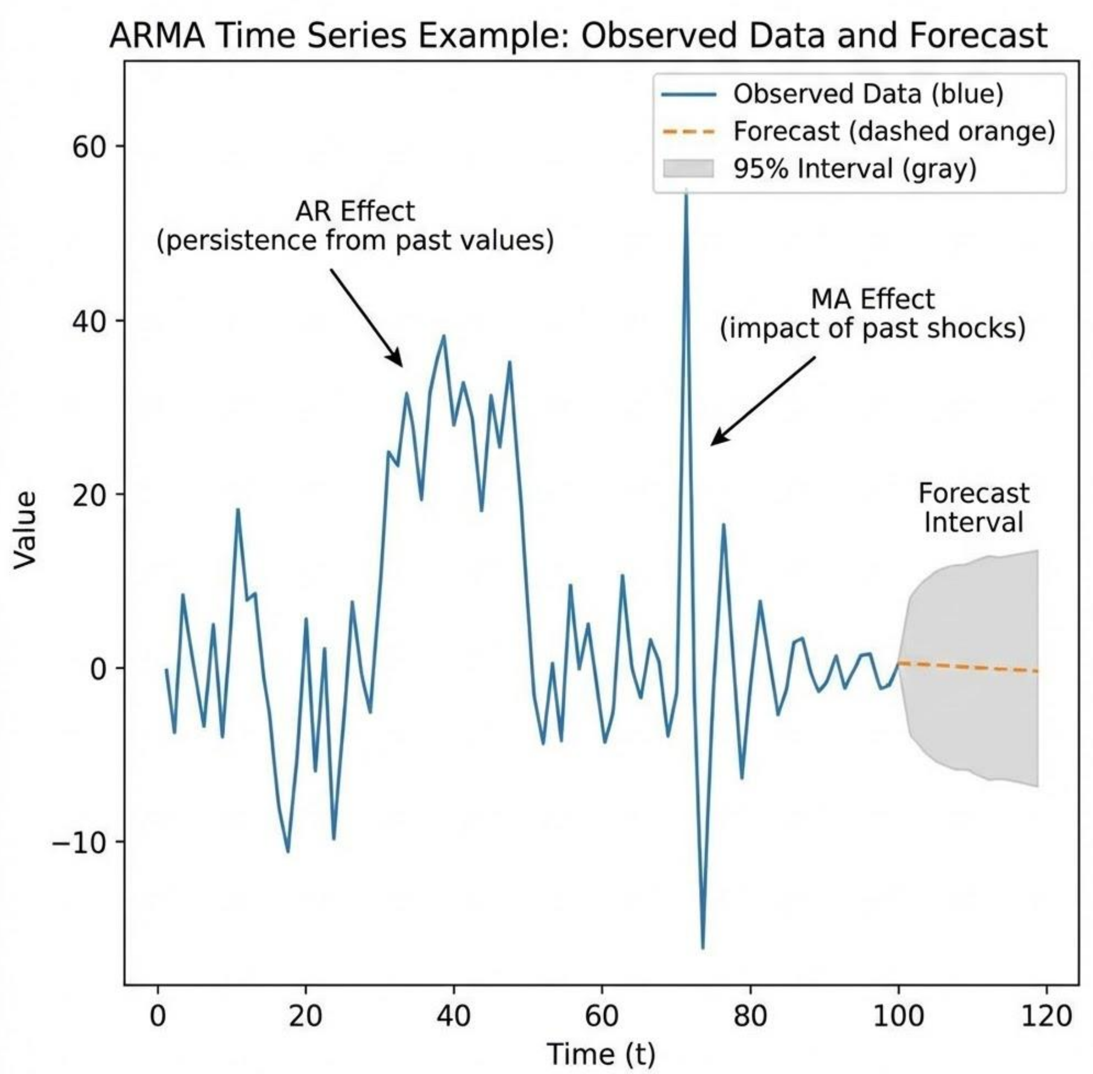
```python
from statsmodels.tsa.arima.model import ARIMA

# ARMA(p, q) = ARIMA(p, 0, q)
model = ARIMA(ts_diff, order=(1, 0, 1))
result = model.fit()

print(result.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:            SUNACTIVITY   No. Observations:                  308
Model:                 ARIMA(1, 0, 1)   Log Likelihood               -1358.256
Date:                Mon, 02 Feb 2026   AIC                           2724.512
Time:                        06:50:32   BIC                           2739.432
Sample:                             0   HQIC                          2730.478
                                - 308
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0057      2.716     -0.002      0.998      -5.328       5.317
ar.L1          0.4078      0.074      5.527      0.000       0.263       0.552
ma.L1          0.2042      0.081      2.515      0.012       0.045       0.363
sigma2       395.7234     24.175     16.369      0.000     348.342     443.105
===================================================================================
Ljung-Box (L1) (Q):                   0.23   Jarque-Bera (JB):                32.93
Prob(Q):                              0.63   Prob(JB):                         0.00
Heteroskedasticity (H):               1.78   Skew:                             0.10
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.59
===================================================================================
```

- Refer [code example 2]



ARMA Models

ARMA (p, q) Model

```
                               SARIMAX Results
==============================================================================
Dep. Variable:            SUNACTIVITY   No. Observations:                  308
Model:                 ARIMA(1, 0, 1)   Log Likelihood               -1358.256
Date:                Mon, 02 Feb 2026   AIC                           2724.512
Time:                        06:50:32   BIC                           2739.432
Sample:                             0   HQIC                          2730.478
                                - 308
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0057      2.716     -0.002      0.998      -5.328       5.317
ar.L1          0.4078      0.074      5.527      0.000       0.263       0.552
ma.L1          0.2042      0.081      2.515      0.012       0.045       0.363
sigma2       395.7234     24.175     16.369      0.000     348.342     443.105
===================================================================================
Ljung-Box (L1) (Q):                   0.23   Jarque-Bera (JB):                32.93
Prob(Q):                              0.63   Prob(JB):                         0.00
Heteroskedasticity (H):               1.78   Skew:                             0.10
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.59
===================================================================================
```

The larger (less negative) → the better the pattern

ARMA Models

ARMA (p, q) Model

```
                               SARIMAX Results
==============================================================================
Dep. Variable:            SUNACTIVITY   No. Observations:                  308
Model:                 ARIMA(1, 0, 1)   Log Likelihood               -1358.256
Date:                Mon, 02 Feb 2026   AIC                           2724.512
Time:                        06:50:32   BIC                           2739.432
Sample:                             0   HQIC                          2730.478
                                - 308
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0057      2.716     -0.002      0.998      -5.328       5.317
ar.L1          0.4078      0.074      5.527      0.000       0.263       0.552
ma.L1          0.2042      0.081      2.515      0.012       0.045       0.363
sigma2       395.7234     24.175     16.369      0.000     348.342     443.105
===================================================================================
Ljung-Box (L1) (Q):                   0.23   Jarque-Bera (JB):                32.93
Prob(Q):                              0.63   Prob(JB):                         0.00
Heteroskedasticity (H):               1.78   Skew:                             0.10
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.59
===================================================================================
```

Model selection criteria. **Penalize** multi-parameter models.

Pupose: Choose a model that fits the data well without being too complex.

ARMA Models

Model Selection Criteria

AIC – Akaike Information Criterion

$$\text{AIC} = -2\log(L) + 2k$$

- L: likelihood of the model
- k: number of parameters
- Light penalty on complexity
- Focus on **prediction performance**

BIC – Bayesian Information Criterion

$$\text{BIC} = -2\log(L) + k\log(n)$$

- n: number of observations
- Stronger penalty for complex models
- Prefers **simpler models**

HQIC – Hannan–Quinn Information Criterion

$$\text{HQIC} = -2\log(L) + 2k\log(\log(n))$$

- Penalty between AIC and BIC
- Less commonly used

Criterion	Penalty Strength	Model Preference
AIC	Light	More complex
HQIC	Medium	Balanced
BIC	Strong	Simpler

👉 **Lower value = better model**

Exercise 1: Review code example 2 and compare AR, MA, and ARMA models for these **4 criterions**.

ARMA Models

AR & MA coefficients

```
                               SARIMAX Results
==============================================================================
Dep. Variable:            SUNACTIVITY   No. Observations:                  308
Model:                 ARIMA(1, 0, 1)   Log Likelihood               -1358.256
Date:                Mon, 02 Feb 2026   AIC                           2724.512
Time:                        06:50:32   BIC                           2739.432
Sample:                             0   HQIC                          2730.478
                                - 308
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0057      2.716     -0.002      0.998      -5.328       5.317
ar.L1          0.4078      0.074      5.527      0.000       0.263       0.552
ma.L1          0.2042      0.081      2.515      0.012       0.045       0.363
sigma2       395.7234     24.175     16.369      0.000     348.342     443.105
===================================================================================
Ljung-Box (L1) (Q):                   0.23   Jarque-Bera (JB):                32.93
Prob(Q):                              0.63   Prob(JB):                         0.00
Heteroskedasticity (H):               1.78   Skew:                             0.10
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.59
===================================================================================
```

AR & MA coefficients

Testing on residuals

ARMA Models

AR & MA coefficients

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0057	2.716	-0.002	0.998	-5.328	5.317
ar.L1	0.4078	0.074	5.527	0.000	0.263	0.552
ma.L1	0.2042	0.081	2.515	0.012	0.045	0.363
sigma2	395.7234	24.175	16.369	0.000	348.342	443.105

Variance of white noise

ARMA Models

AR & MA coefficients

Statistically significant (p-value = 0.000 < 0.05)

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0057	2.716	-0.002	0.998	-5.328	5.317
ar.L1	0.4078	0.074	5.527	0.000	0.263	0.552
ma.L1	0.2042	0.081	2.515	0.012	0.045	0.363
sigma2	395.7234	24.175	16.369	0.000	348.342	443.105

The value at time t is influenced by approximately 40% from the value at t-1.

ARMA Models

AR & MA coefficients

Statistically significant (p-value = 0.012 < 0.05)

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0057	2.716	-0.002	0.998	-5.328	5.317
ar.L1	0.4078	0.074	5.527	0.000	0.263	0.552
ma.L1	0.2042	0.081	2.515	0.012	0.045	0.363
sigma2	395.7234	24.175	16.369	0.000	348.342	443.105

Errors from the past are still affecting the present.

Content

- ARMA Models
- ACF & PACF for Model Selection

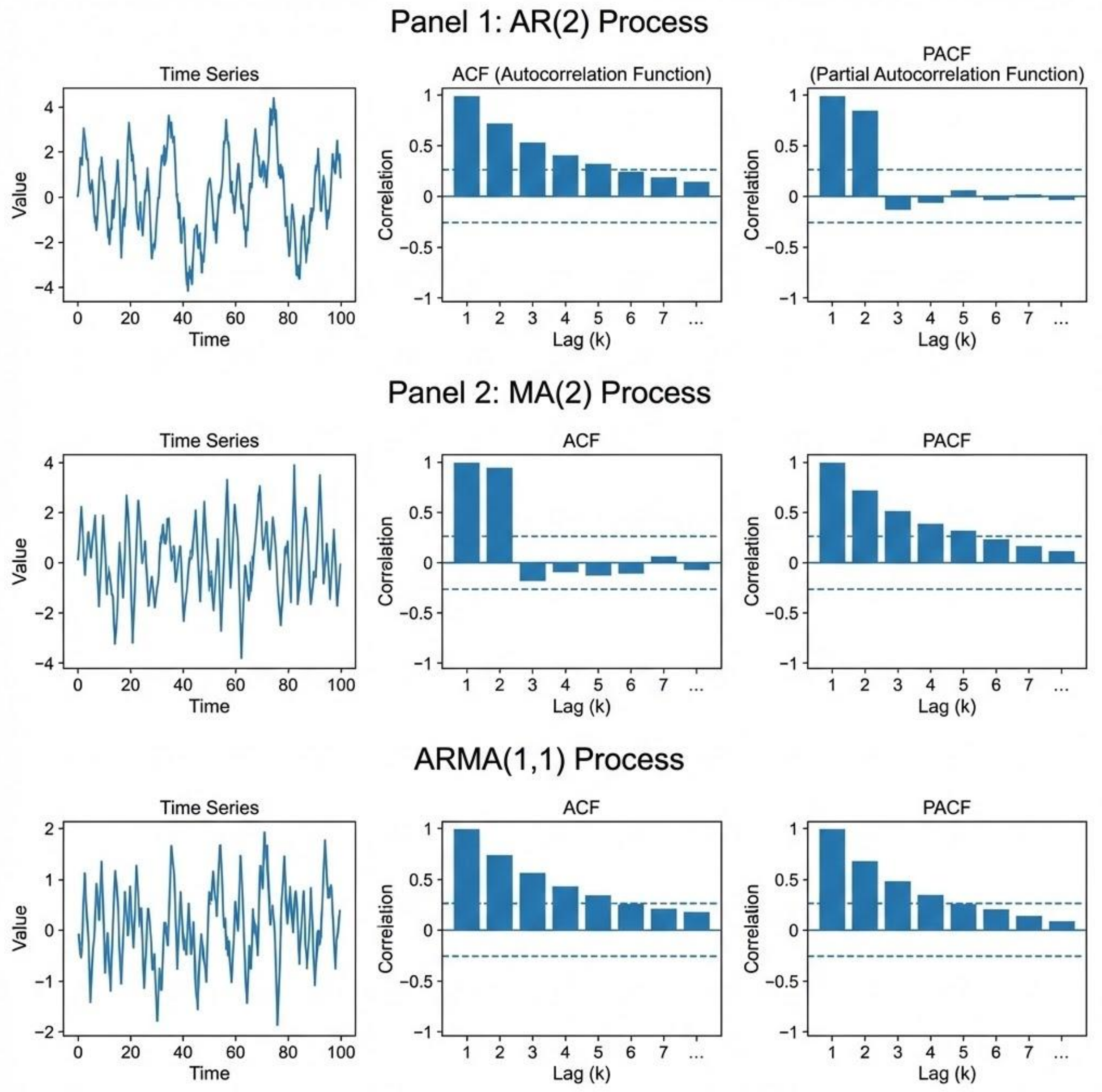
ACF & PACF for Model Selection

ACF & PACF

- Help identify **AR vs MA structure**
- Guide selection of **p** and **q**
- Based on **correlation patterns**, not optimization

Key intuition

- **ACF**: correlation between x_t and x_{t-k}
- **PACF**: *direct* correlation between x_t and x_{t-k}, removing effects of intermediate lags



ACF & PACF for Model Selection

ACF & PACF

What are p and q?

- **p**: order of **Autoregressive (AR)** part
  → number of past values (memory)
- **q**: order of **Moving Average (MA)** part
  → number of past shocks (errors)

What does "cut-off" mean?

- **Cut-off**: correlations are significant only up to a certain lag, then suddenly drop to zero
- **Decay**: correlations decrease gradually over many lags

Model identification rules

Model	ACF pattern	PACF pattern
AR(p)	Decay	Cut-off at lag **p**
MA(q)	Cut-off at lag **q**	Decay
ARMA(p, q)	Decay	Decay

Practical rule of thumb

> Use **PACF → choose p**
> Use **ACF → choose q**

ACF & PACF for Model Selection

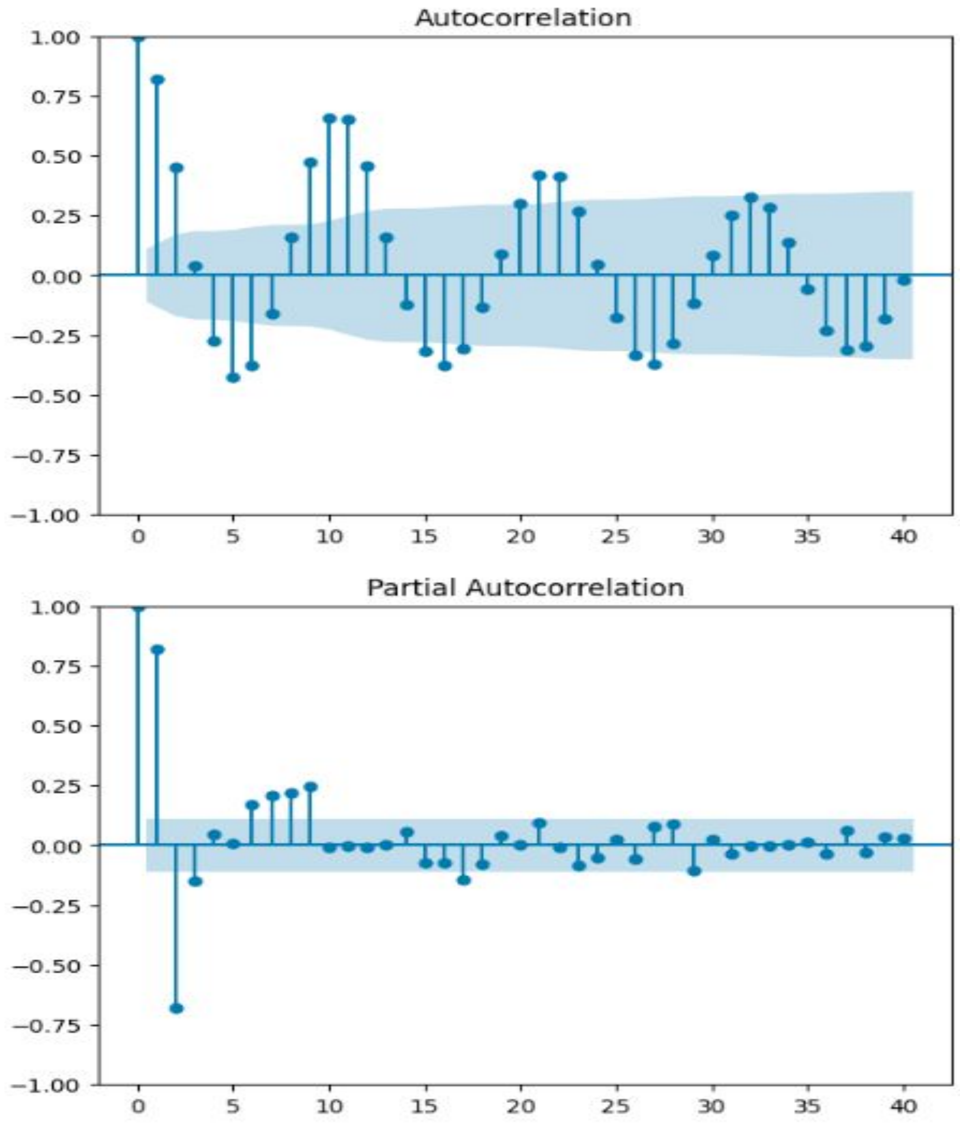
ACF & PACF

```
import matplotlib.pyplot as plt
from statsmodels.datasets import sunspots
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

# Load dataset
data = sunspots.load_pandas().data
ts = data['SUNACTIVITY']

# Plot time series
ts.plot(title="Sunspots Time Series", figsize=(6,3))
plt.show()

# ACF & PACF
plot_acf(ts, lags=40)
plot_pacf(ts, lags=40, method='ywm')
plt.show()
```



ACF & PACF for Model Selection

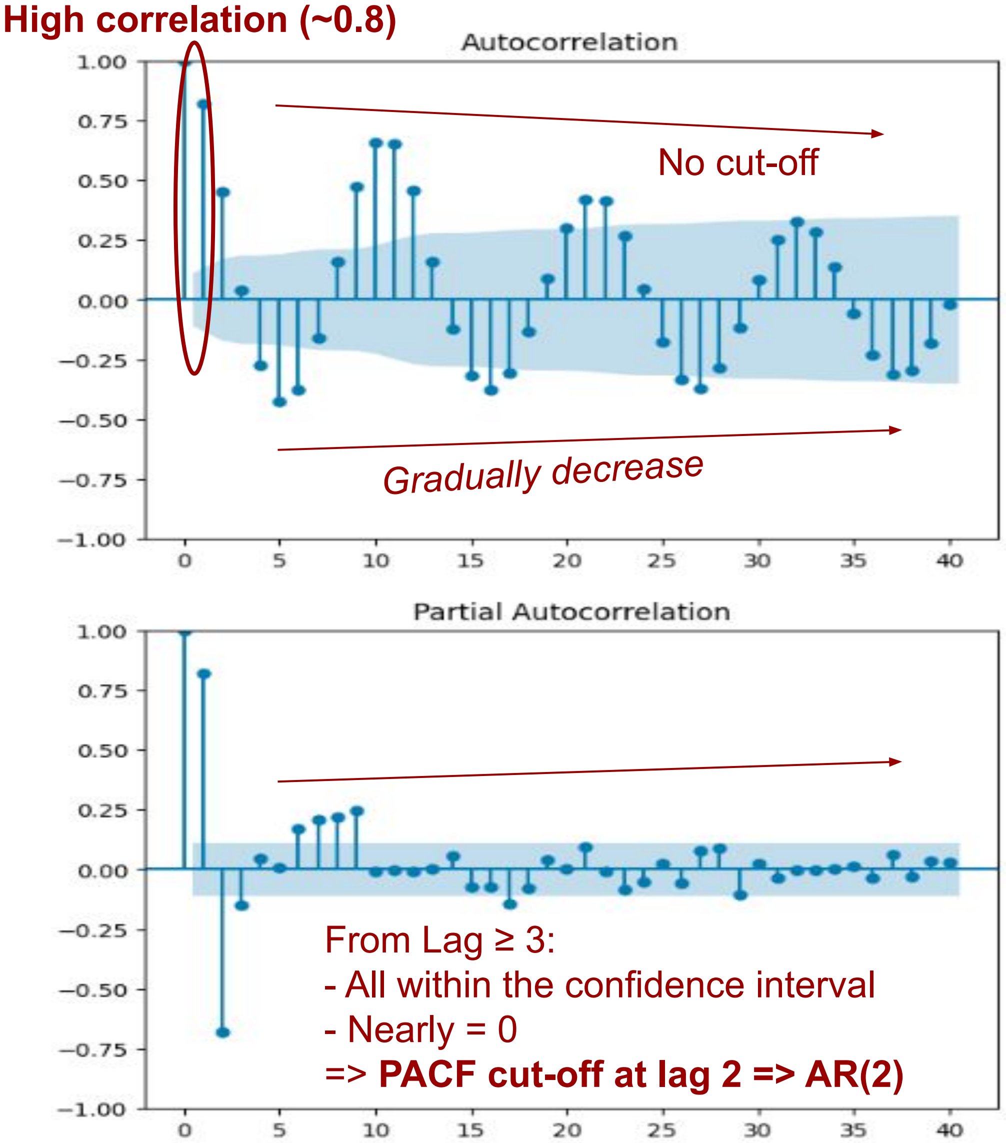
ACF & PACF

```python
import matplotlib.pyplot as plt
from statsmodels.datasets import sunspots
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

# Load dataset
data = sunspots.load_pandas().data
ts = data['SUNACTIVITY']

# Plot time series
ts.plot(title="Sunspots Time Series", figsize=(6,3))
plt.show()

# ACF & PACF
plot_acf(ts, lags=40)
plot_pacf(ts, lags=40, method='ywm')
plt.show()
```



Thank you