ENERGY DATA SCIENCE

Statistical models for Time series

Prof. Juri Belikov

Department of Software Science
Tallinn University of Technology
juri.belikov@taltech.ee

PREVIOUSLY IN COURSE …

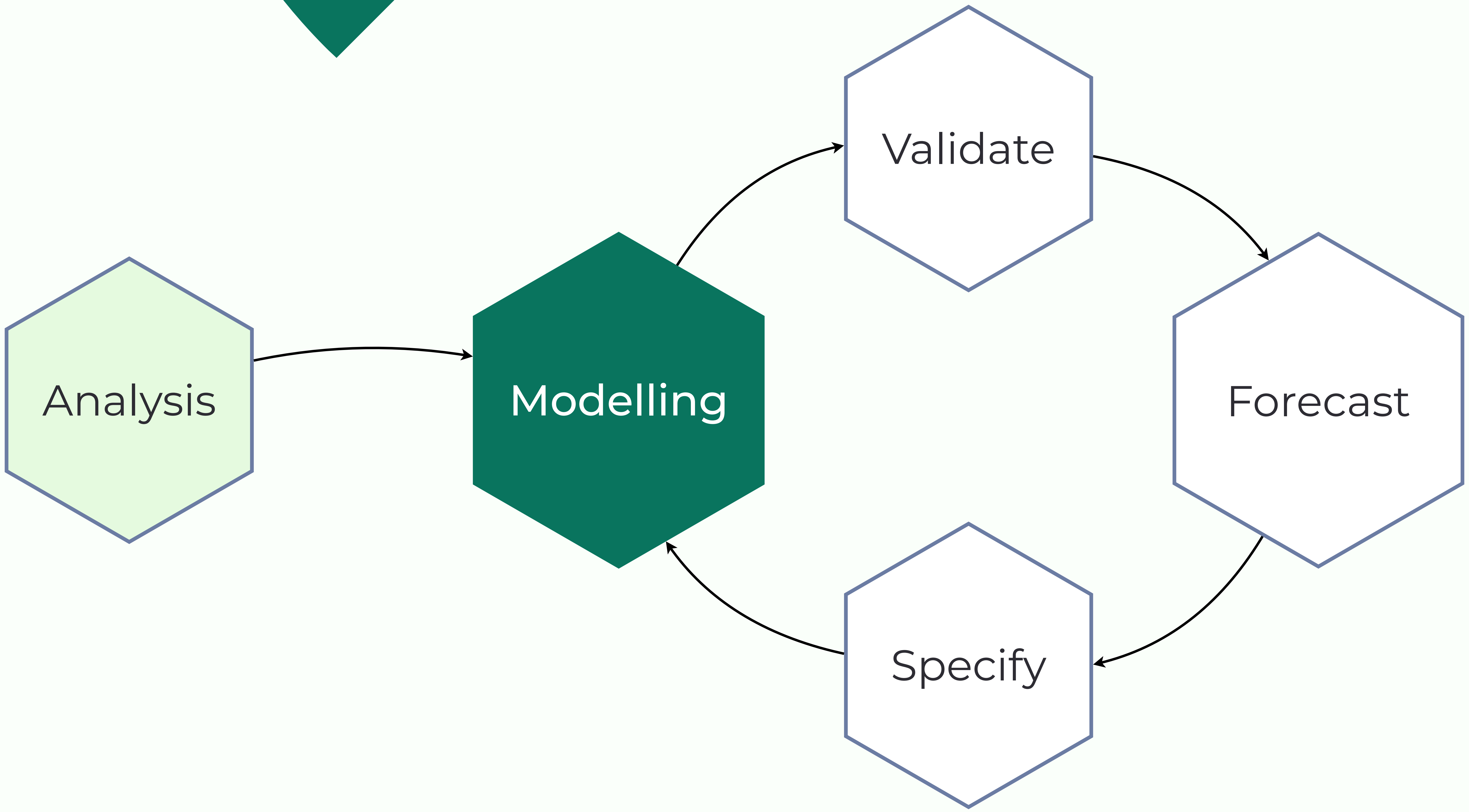
Key takeaways:

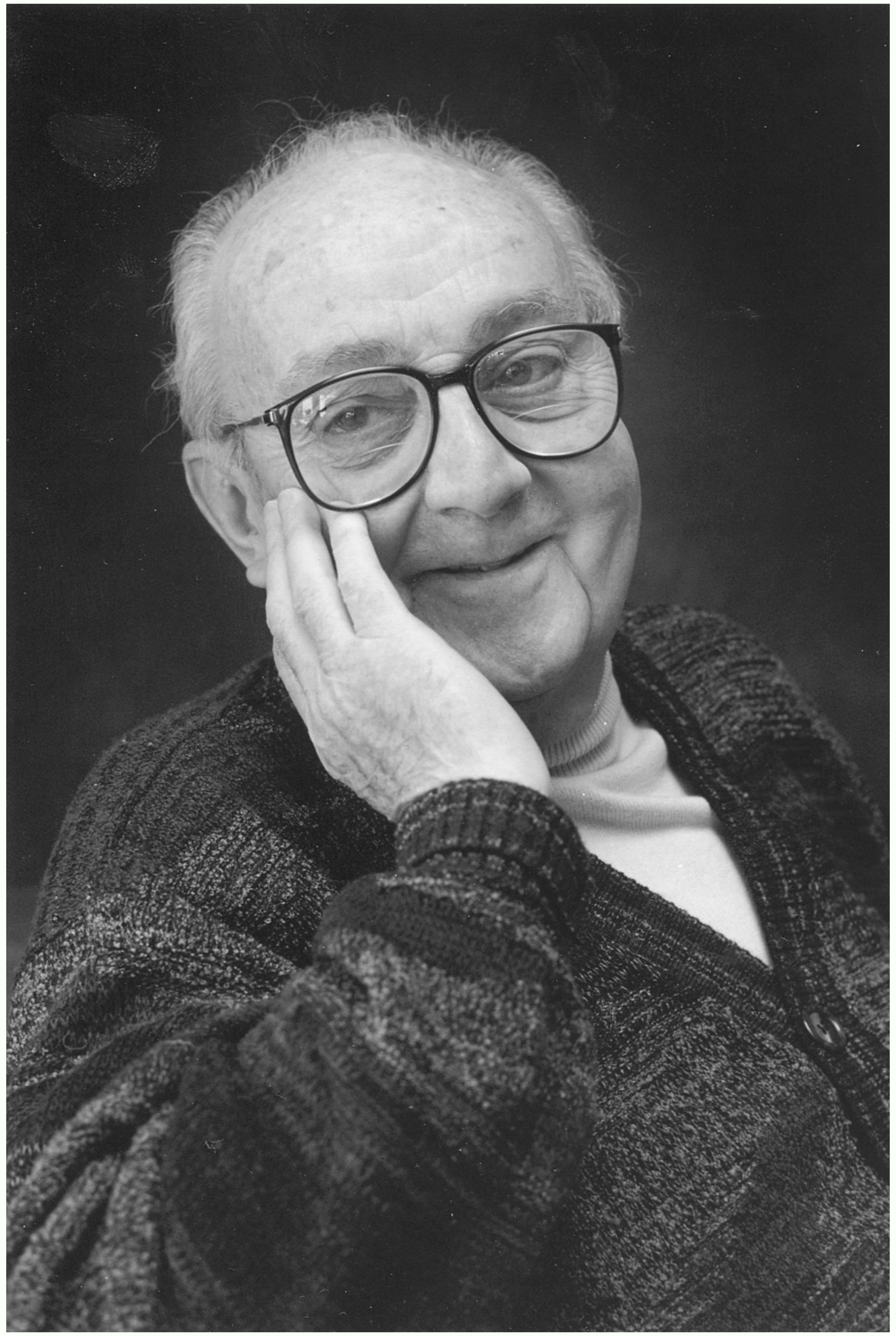
- Time series and why they are so special
- Components: trend, seasonality, cycle, remainder
- Classical decomposition

Train ML models, evaluate performance, and make predictions.

06

MODELLING





“All models are wrong, but some models are useful.”

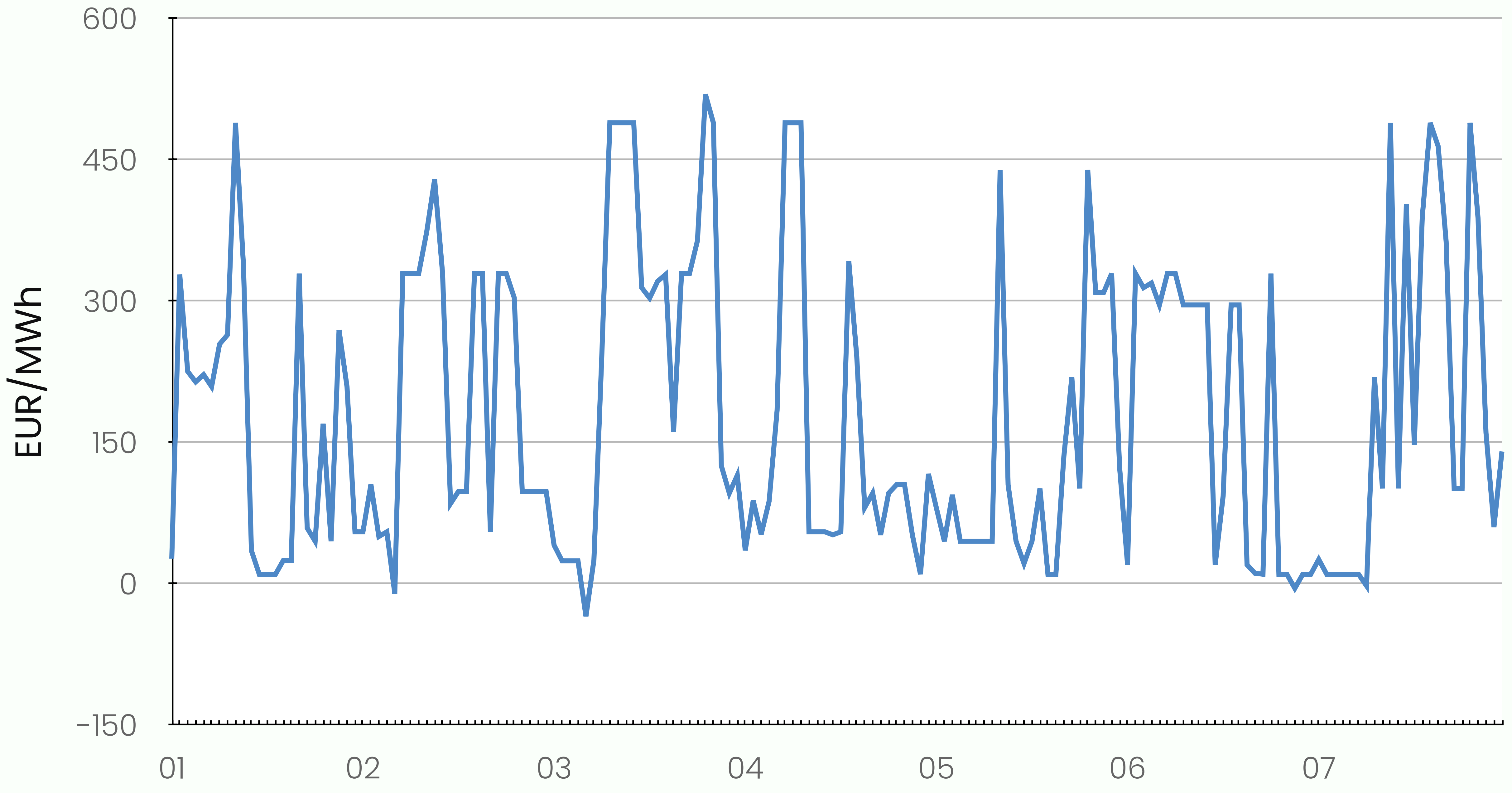
George E. P. Box (1919–2013)

Stationarity

CONSISTENCY

- ✓ Cannot have independence, but want some consistency.
- ✓ Distribution depends **only** on difference in time and **not** on location in time.

Imbalance electricity price, March 01-07, ’24, Estonia



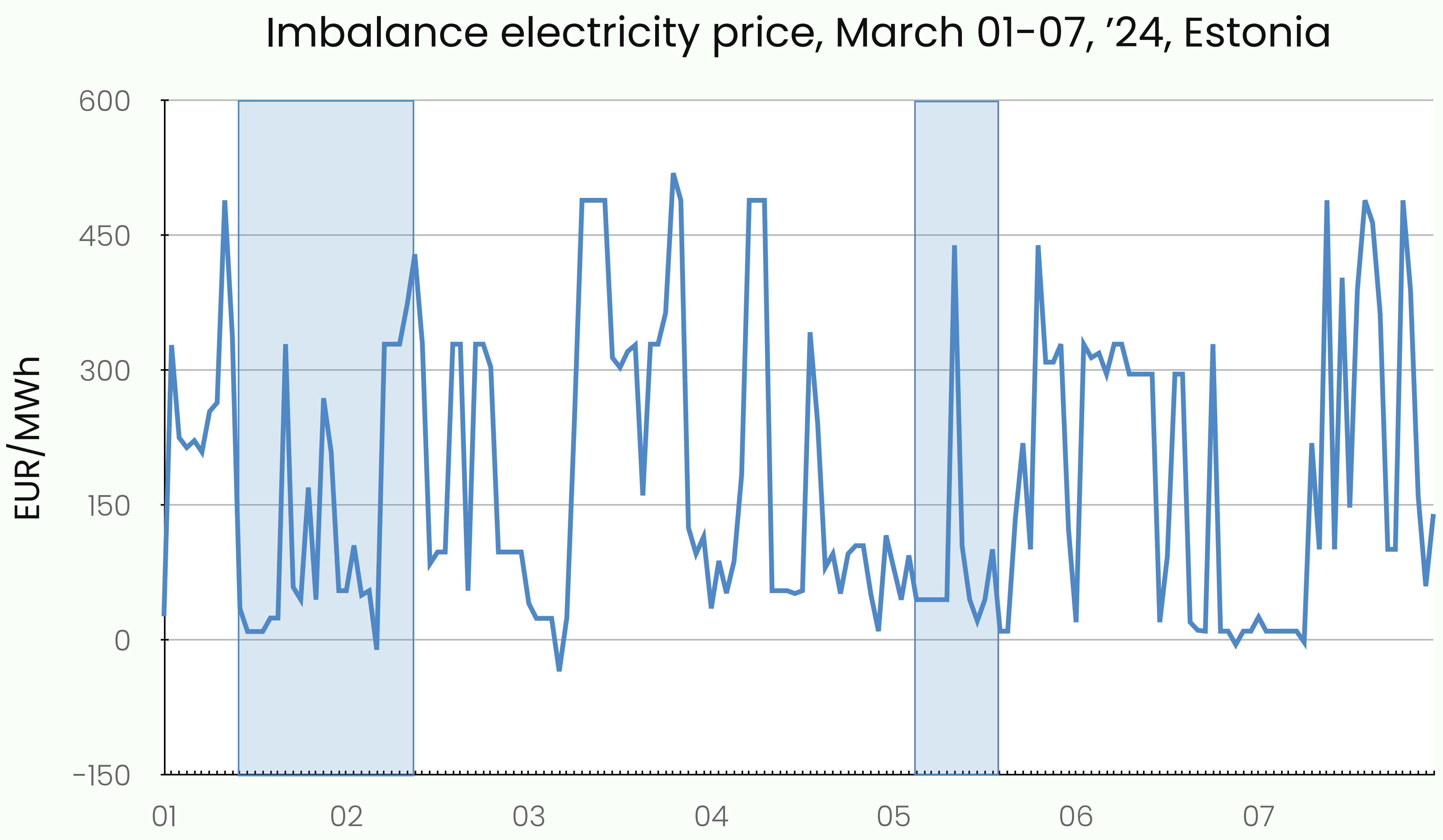
CONSISTENCY

- ✓ Cannot have independence, but want some consistency.
- ✓ Distribution depends **only** on difference in time and **not** on location in time.

Different widths

↓

Different distributions



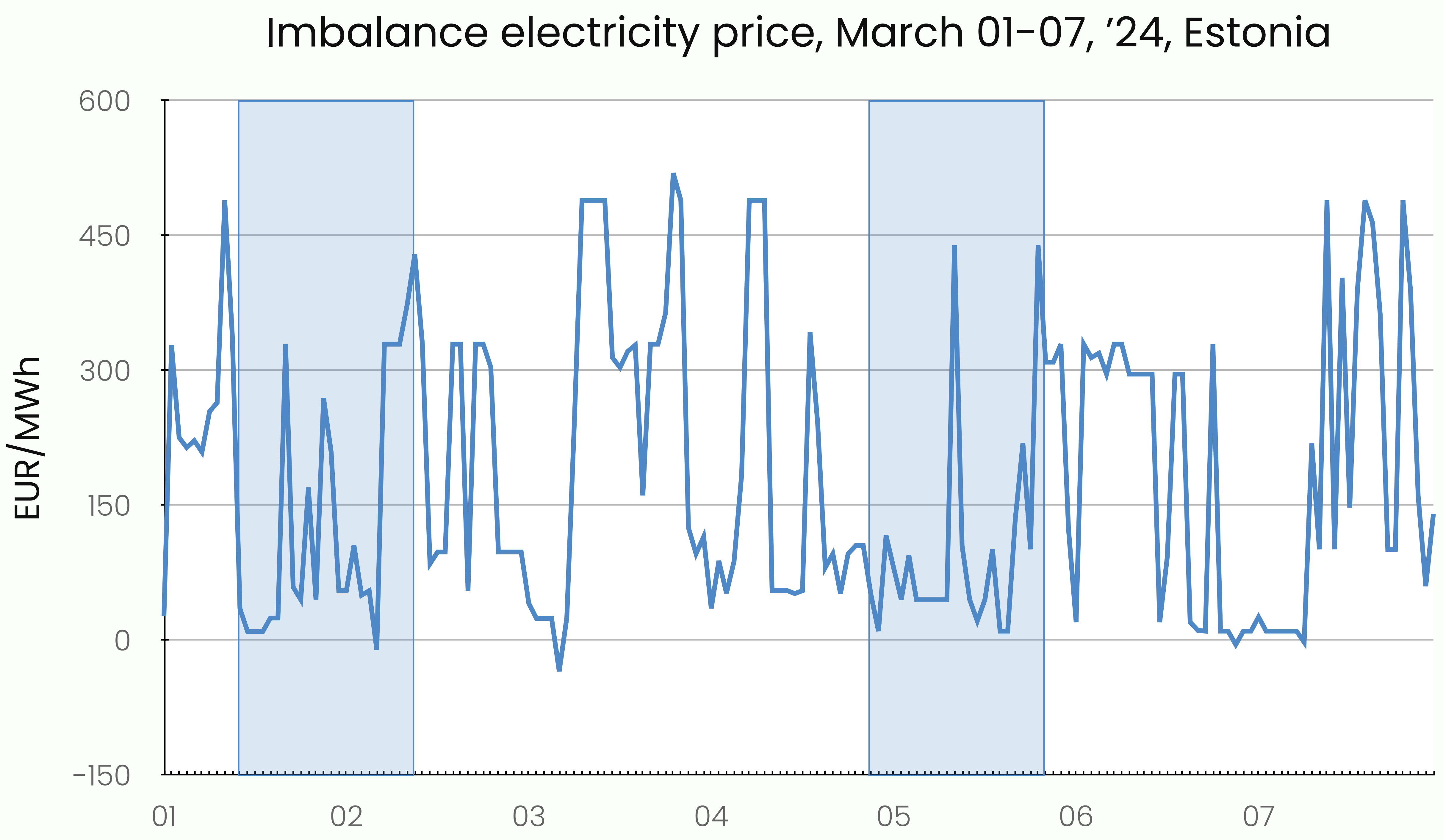
STRONG STATIONARITY

Distribution depends **only** on difference in time and **not** on location in time.

Same widths (any size)

↓

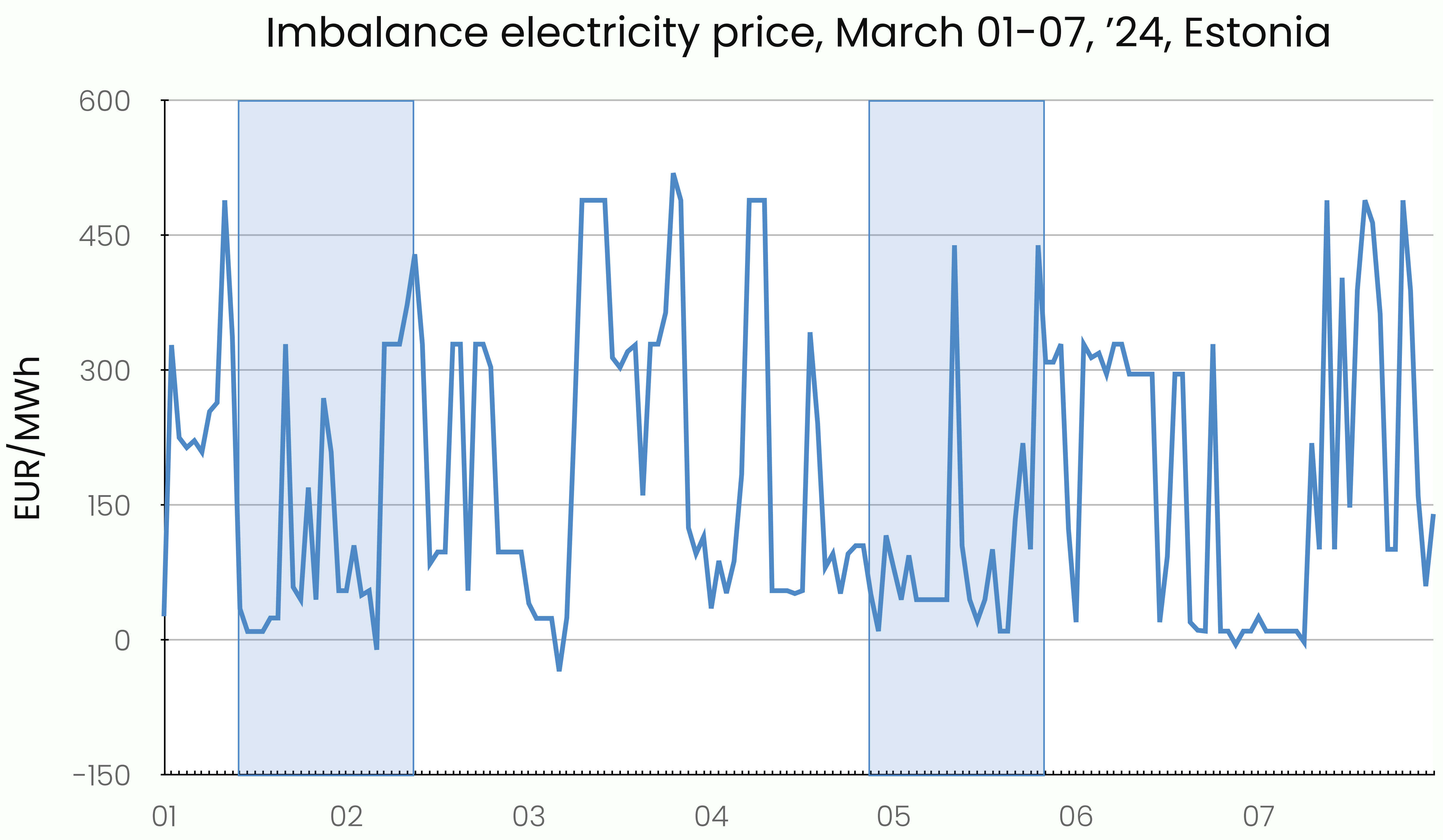
Expect same distributions



WEAK STATIONARITY

Properties depend only on difference in time, not on location in time:

- ✓ mean is constant
- ✓ variation is constant
- ✓ no seasonality

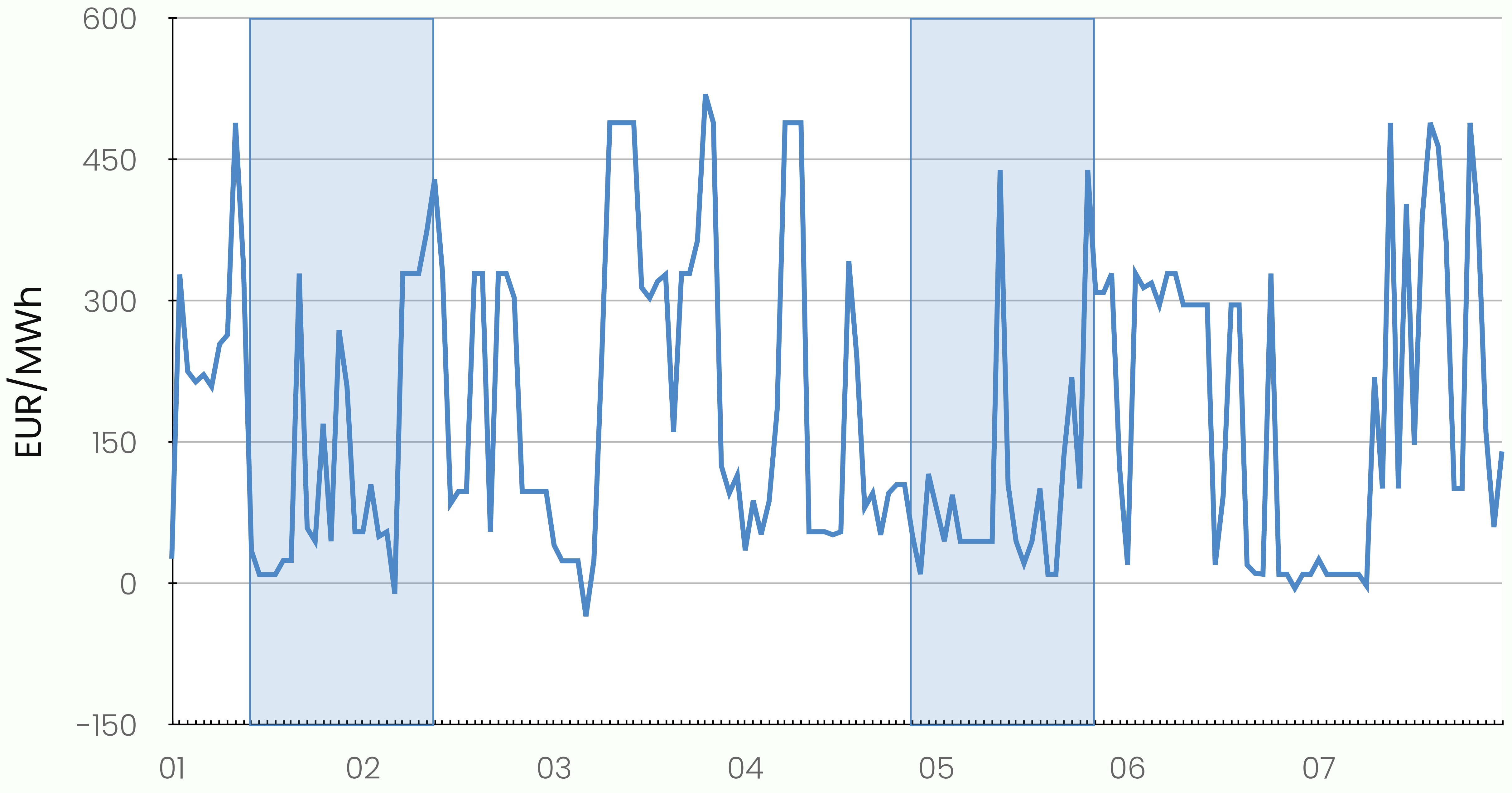


WEAK STATIONARITY

Properties depend only on difference in time, not on location in time:

- ✓mean is constant
- ✓variation is constant
- ✓no seasonality

Imbalance electricity price, March 01-07, ’24, Estonia



In most cases we only need weak stationarity.

WHY IS IT IMPORTANT?

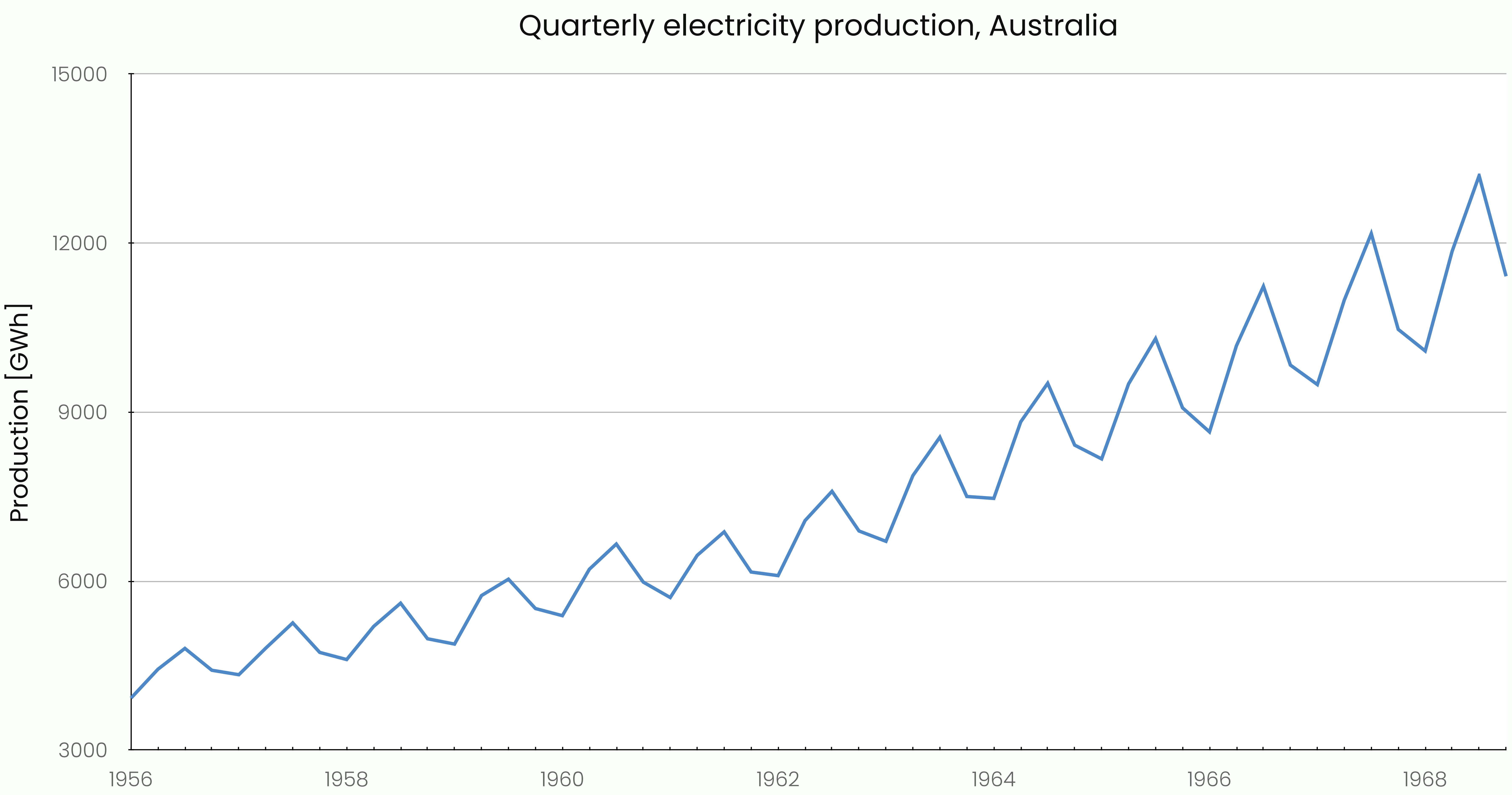
Statistical properties do not change over the time:

- Easy to analyse
- Easy to model

Easy to forecast.

Many existing statistical methods are applicable only when the data are stationary (e.g., ARMA-family models).

METHOD 1: VISUAL INSPECTION



Stationary

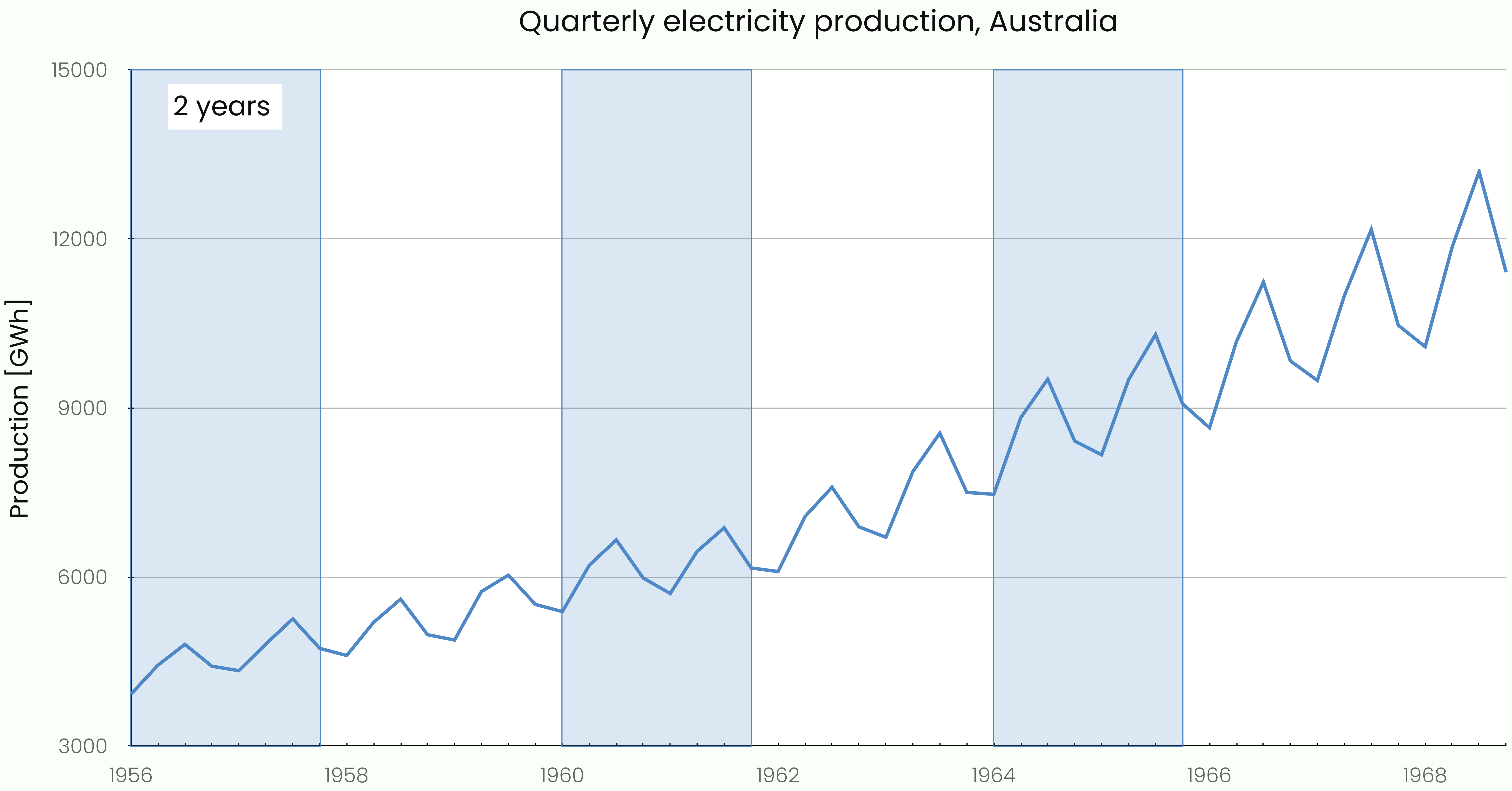
Non-stationary

Mean (trend)

Variation

Seasonal

METHOD 1: VISUAL INSPECTION



Stationary

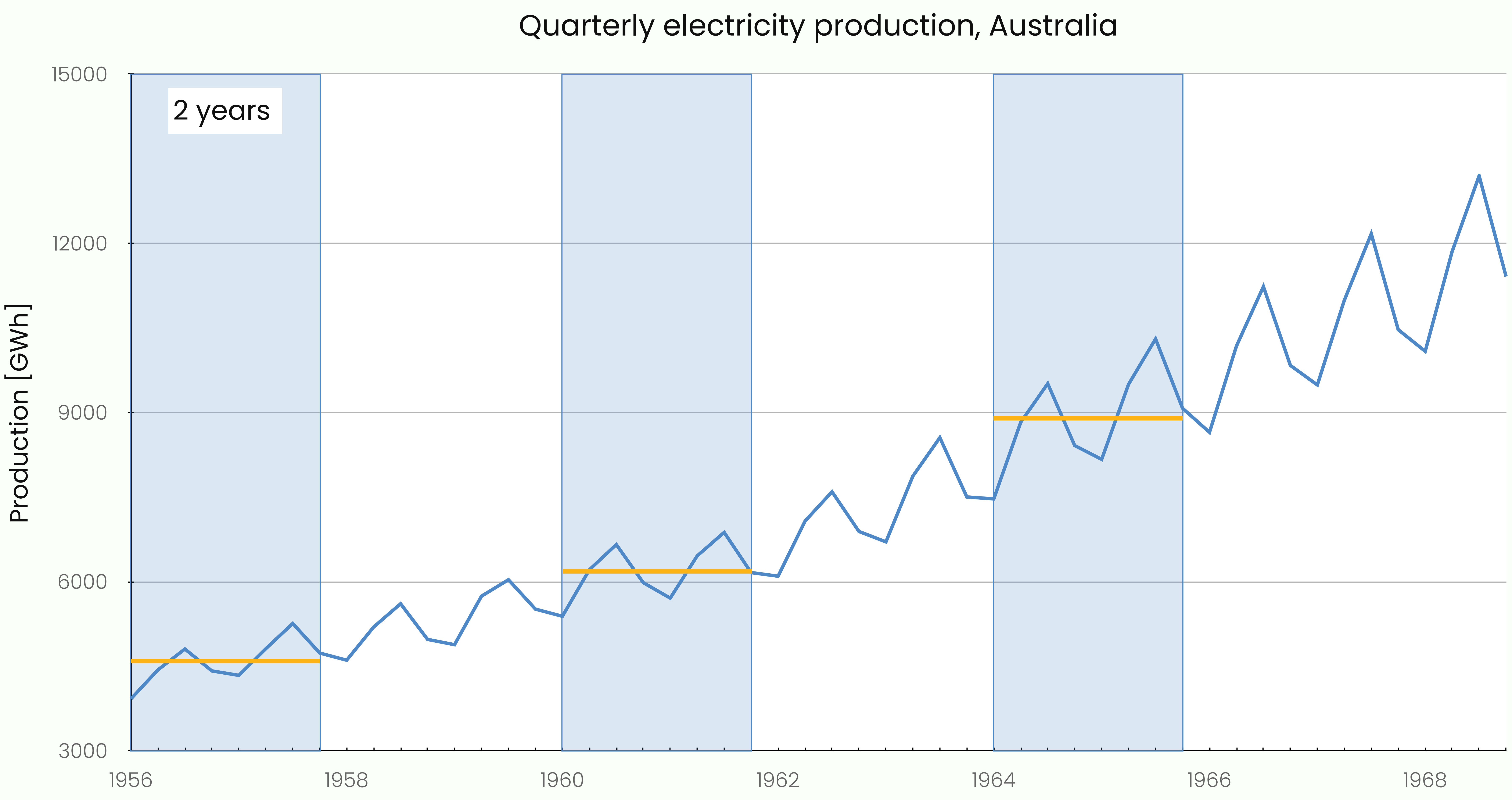
Non-stationary

Mean (trend)

Variation

Seasonal

METHOD 1: VISUAL INSPECTION



$\mu_{t,1}(y_1, \ldots, y_8) = 4591$ $\mu_{t,2} = 6181$ $\mu_{t,3} = 8910$

Stationary

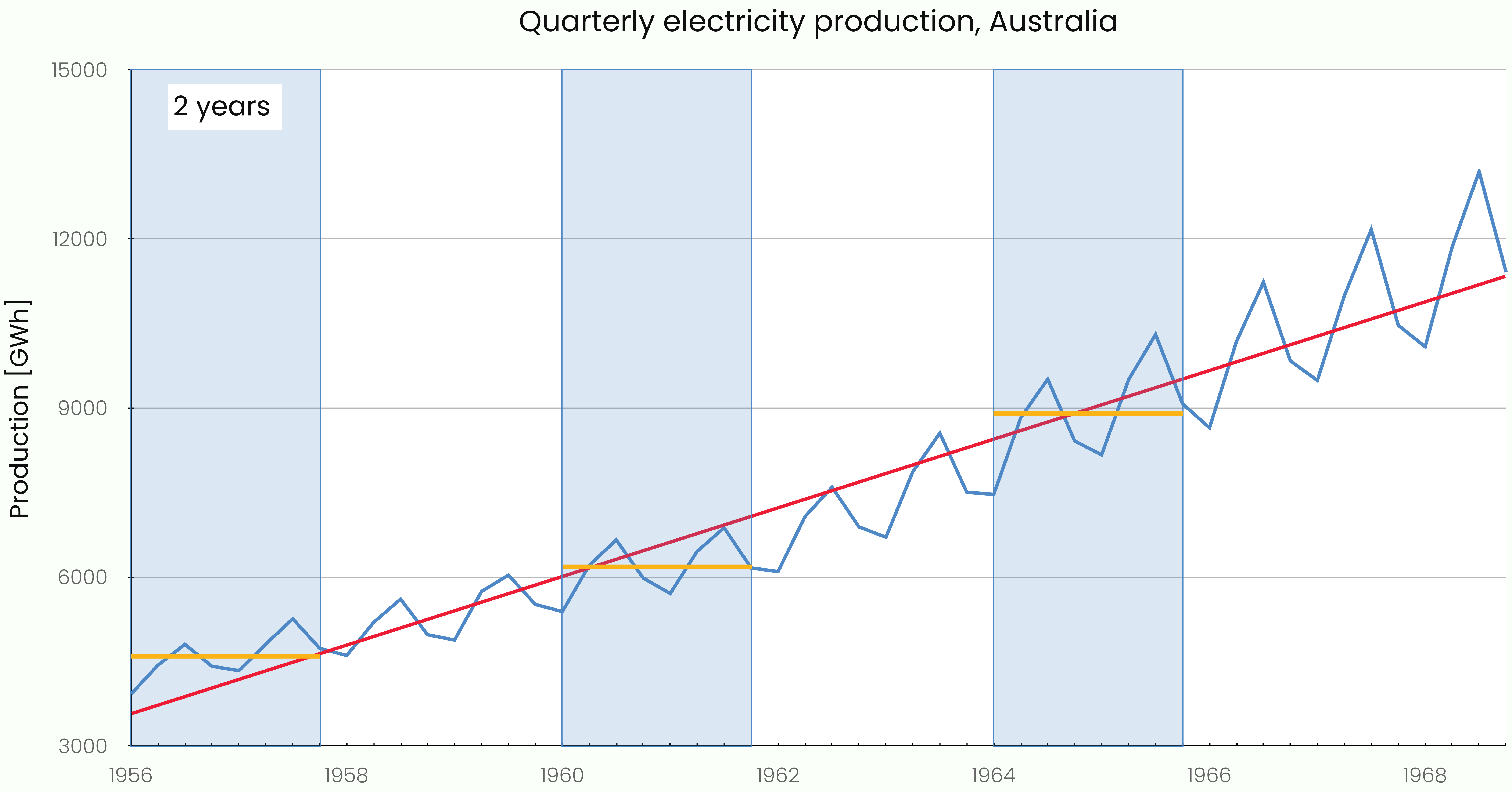
Non-stationary

Mean (trend)

Variation

Seasonal

METHOD 1: VISUAL INSPECTION



$\mu_{t,1}(y_1, \ldots, y_8) = 4591$ $\mu_{t,2} = 6181$ $\mu_{t,3} = 8910$

Stationary

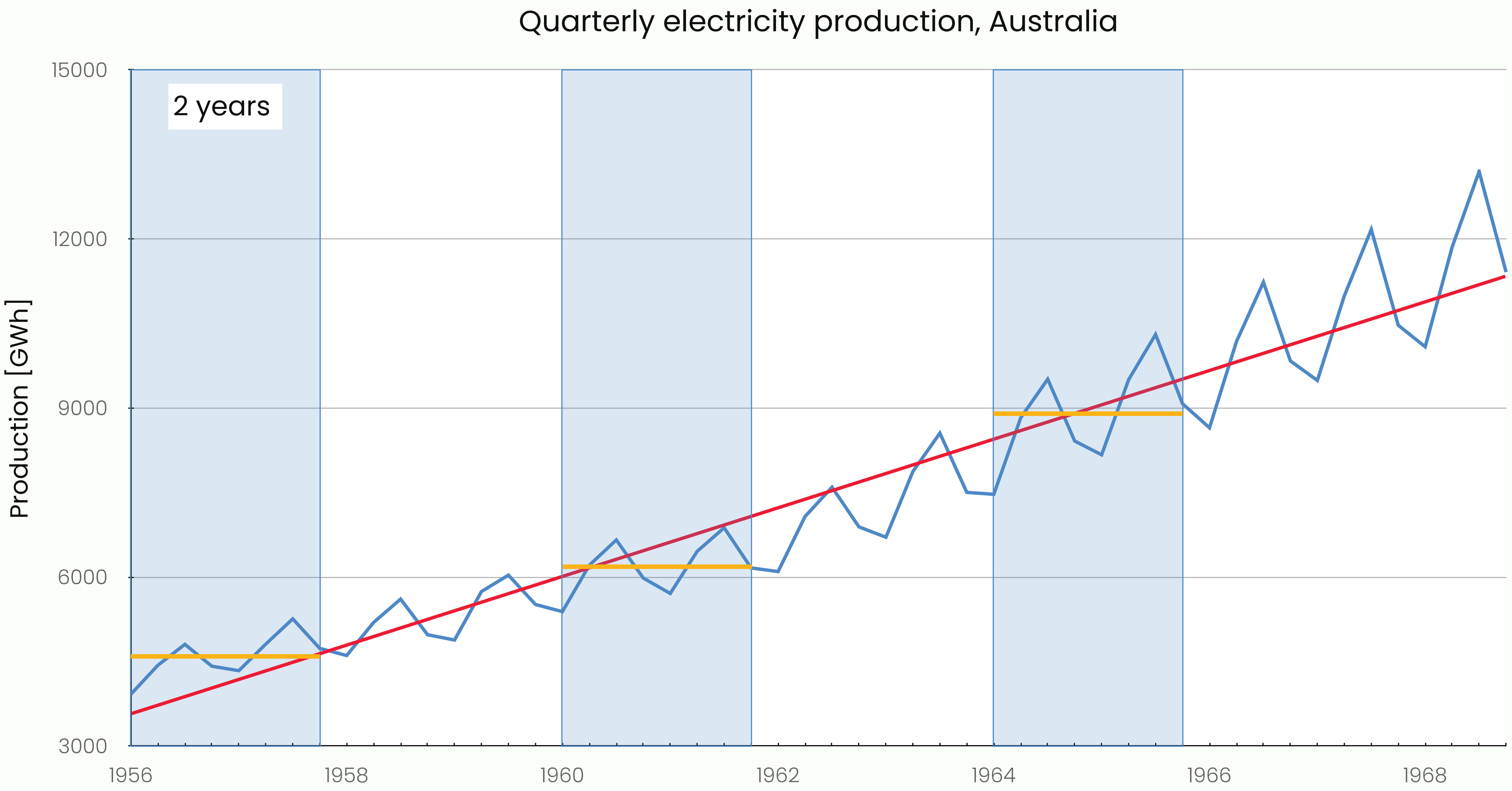
Non-stationary

Mean (trend) ✕

Variation

Seasonal

METHOD 1: VISUAL INSPECTION



$\mu_{t,1}(y_1, \ldots, y_8) = 4591$ $\mu_{t,2} = 6181$ $\mu_{t,3} = 8910$

$\sigma^2_{t,1} = 401$ $\sigma^2_{t,2} = 490$ $\sigma^2_{t,3} = 892$

Stationary

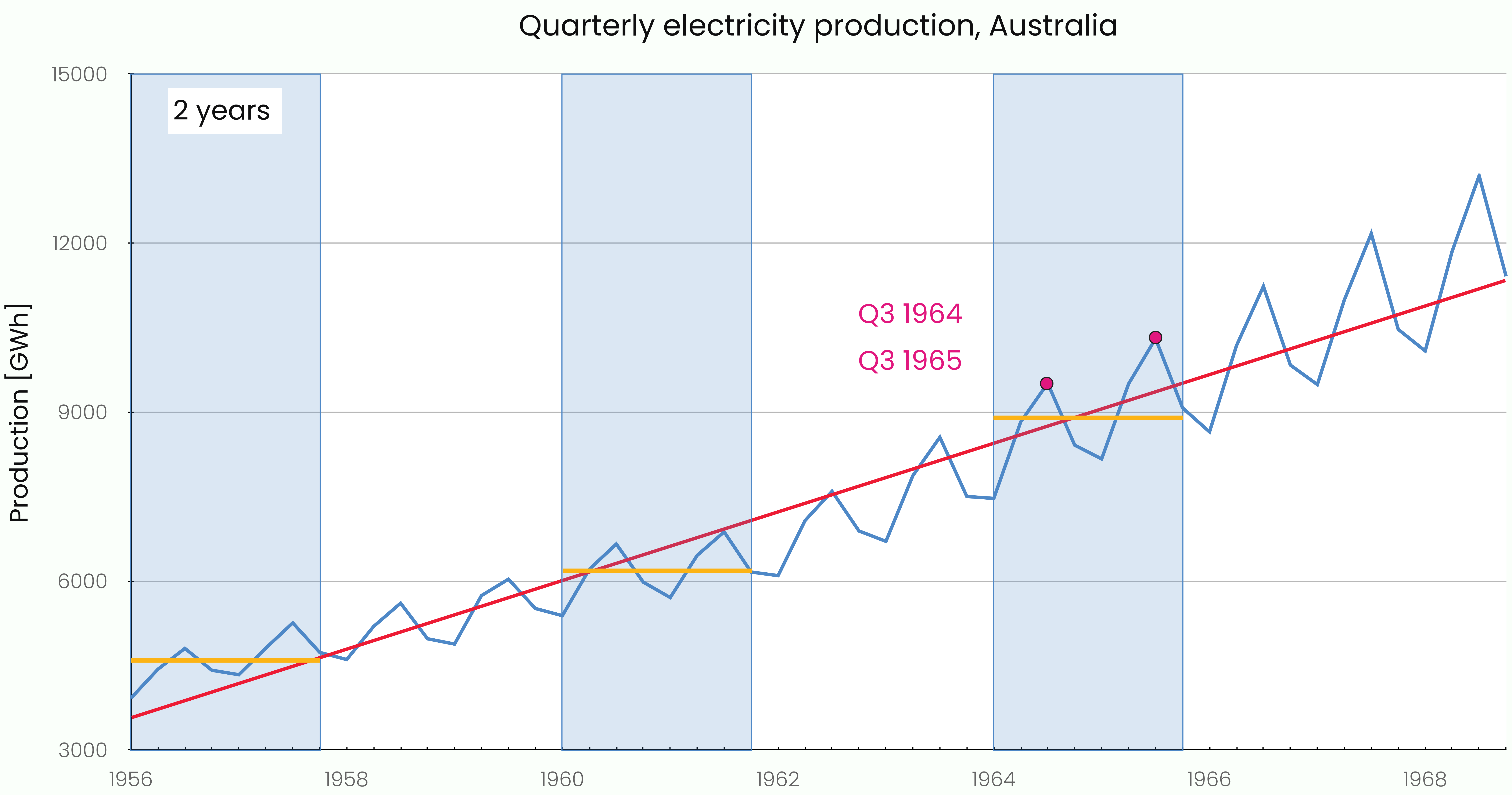
Non-stationary

Mean (trend) ✗

Variation ✗

Seasonal

METHOD 1: VISUAL INSPECTION



$\mu_{t,1}(y_1, \ldots, y_8) = 4591$ $\mu_{t,2} = 6181$ $\mu_{t,3} = 8910$

$\sigma^2_{t,1} = 401$ $\sigma^2_{t,2} = 490$ $\sigma^2_{t,3} = 892$

Stationary

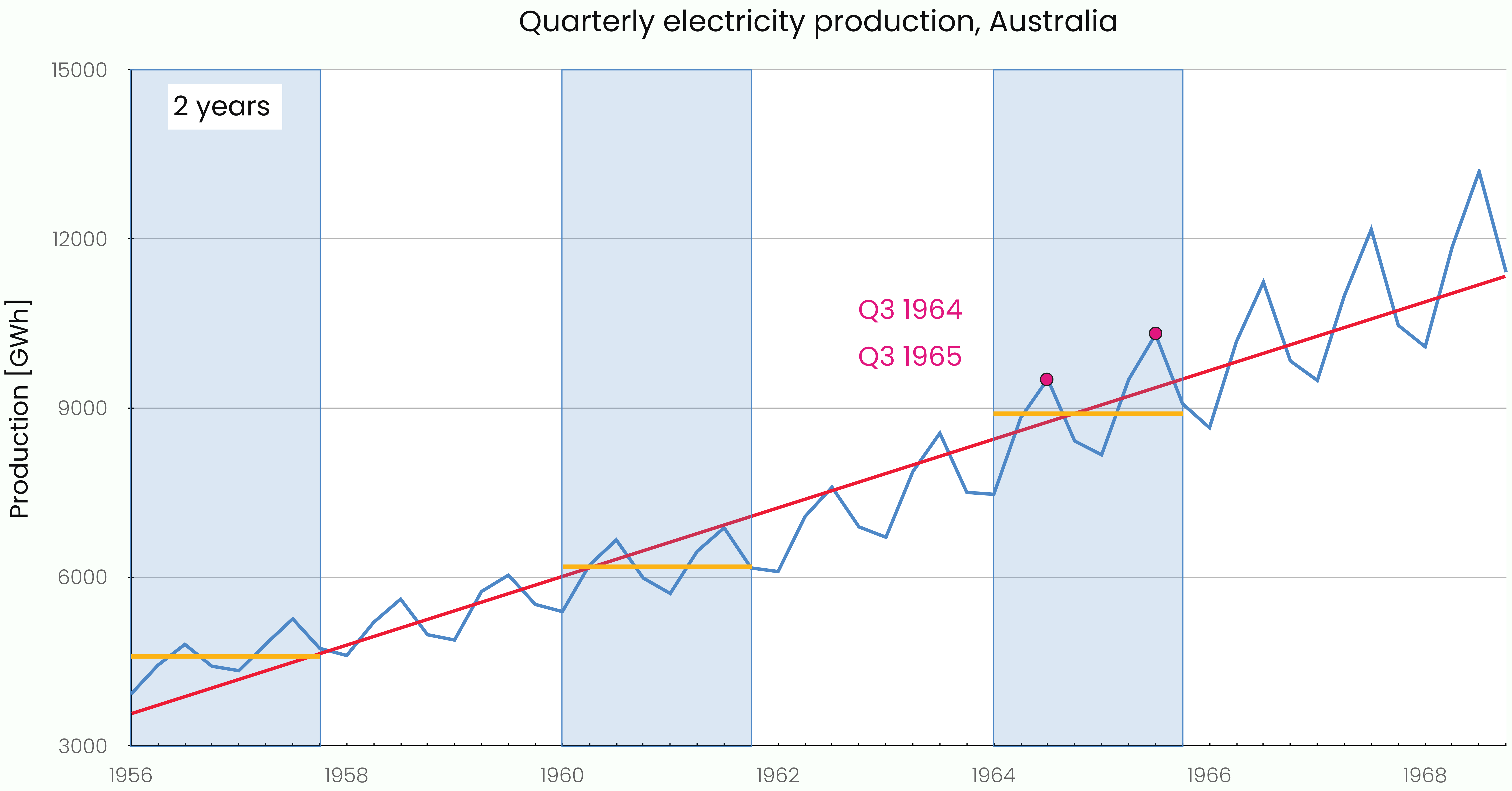
Non-stationary

Mean (trend) ✕

Variation ✕

Seasonal ✕

METHOD 1: VISUAL INSPECTION



$\mu_{t,1}(y_1, \ldots, y_8) = 4591$ $\mu_{t,2} = 6181$ $\mu_{t,3} = 8910$

$\sigma^2_{t,1} = 401$ $\sigma^2_{t,2} = 490$ $\sigma^2_{t,3} = 892$

Stationary

Non-stationary ✓

Mean (trend) ✗

Variation ✗

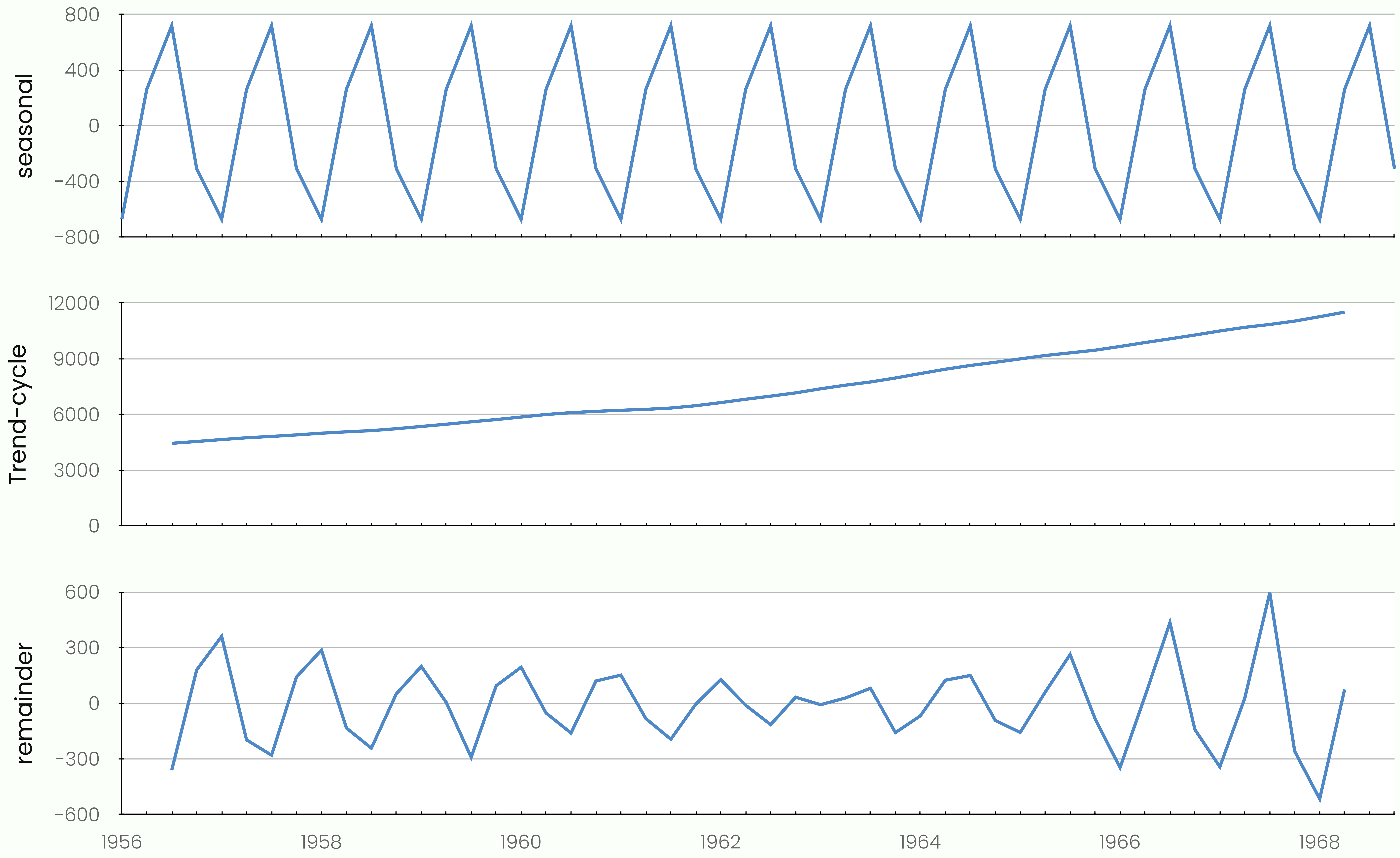
Seasonal ✗

METHOD 2: PARAMETRIC TESTS

- Dickey-Fuller test
- Augmented Dickey-Fuller (ADF) test
- Phillips-Perrone test
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test

[Accessed July 7, 2025], https://www.projectguru.in/what-is-a-stationarity-test-how-to-do-it/

METHOD 2: PARAMETRIC TESTS



METHOD 2: PARAMETRIC TESTS

- Dickey-Fuller test
- Augmented Dickey-Fuller (ADF) test
- Phillips-Perrone test
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test

METHOD 2: PARAMETRIC TESTS

- Dickey-Fuller test
- Augmented Dickey-Fuller (ADF) test
- Phillips-Perrone test
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test

Original data	Remainder (decomposed)
ADF Statistic: 2.233387	ADF Statistic: -11.701448
p-value: 0.998909	p-value: 0.0000
Critical Values:	Critical Values:
1%: -3.578	1%: -3.585
5%: -2.925	5%: -2.928
10%: -2.601	10%: -2.602

ACF & PACF

AUTOCORRELATION FUNCTION (ACF)

ACF measures the self-similarity of the signal over different delay times (*lagged values*).

ACF: INTUITION

Forecast average monthly electricity price.

ACF: INTUITION

Forecast average monthly electricity price.

y_t December

y_{t-1} November

y_{t-2} October

$\vdots$

ACF: INTUITION

Forecast average monthly electricity price.

y_t	December	y_{t-2}	y_{t-1}	y_t
y_{t-1}	November			
y_{t-2}	October			
	$\vdots$			

ACF: INTUITION

Forecast average monthly electricity price.

y_t	December
y_{t-1}	November
y_{t-2}	October
$\vdots$	

$$y_{t-2} \qquad y_{t-1} \qquad y_t$$

$$\text{corr}(y_{t-2}, y_t)$$

ACF: INTUITION

Forecast average monthly electricity price.

y_t	December
y_{t-1}	November
y_{t-2}	October
$\vdots$	

$$y_{t-2} \qquad y_{t-1} \qquad y_t$$

$$\text{corr}(y_{t-2}, y_t) \xrightarrow{\text{indirect}} y_{t-2} \rightarrow y_{t-1} \rightarrow y_t$$

ACF: INTUITION

Forecast average monthly electricity price.

y_t	December
y_{t-1}	November
y_{t-2}	October
	⋮

$$\begin{tikzcd} y_{t-2} \arrow[r, bend right] \arrow[rr, bend left] & y_{t-1} \arrow[r, bend right] & y_t \end{tikzcd}$$

$$\begin{tikzcd} & y_{t-2} \to y_t \\ \operatorname{corr}(y_{t-2}, y_t) \arrow[ur, "\text{direct}"] \arrow[dr, "\text{indirect}"'] & \\ & y_{t-2} \to y_{t-1} \to y_t \end{tikzcd}$$

ACF: INTUITION

Forecast average monthly electricity price.

y_t	December
y_{t-1}	November
y_{t-2}	October
	$\vdots$

$$
\begin{tikzcd}
y_{t-2} \arrow[r, bend right] \arrow[rr, bend left] & y_{t-1} \arrow[r, bend right] & y_t
\end{tikzcd}
$$

$$
\begin{tikzcd}
& y_{t-2} \to y_t \arrow[dr] & \\
\mathrm{corr}(y_{t-2}, y_t) \arrow[ur, "\text{direct}"] \arrow[dr, "\text{indirect}"'] & & \mathrm{ACF}(y_{t-2}) \\
& y_{t-2} \to y_{t-1} \to y_t \arrow[ur] &
\end{tikzcd}
$$

MATHEMATICALLY SPEAKING

Let r_i denote the autocorrelation coefficient.

The value of r_i can be calculated as:

$$r_i = \frac{\sum_{t=i+1}^{T} (y_t - \bar{y})(y_{t-i} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2},$$

where T is the length of the time series, and $\bar{y}$ is the average of a time series signal.

AUTOCORRELATION PLOT

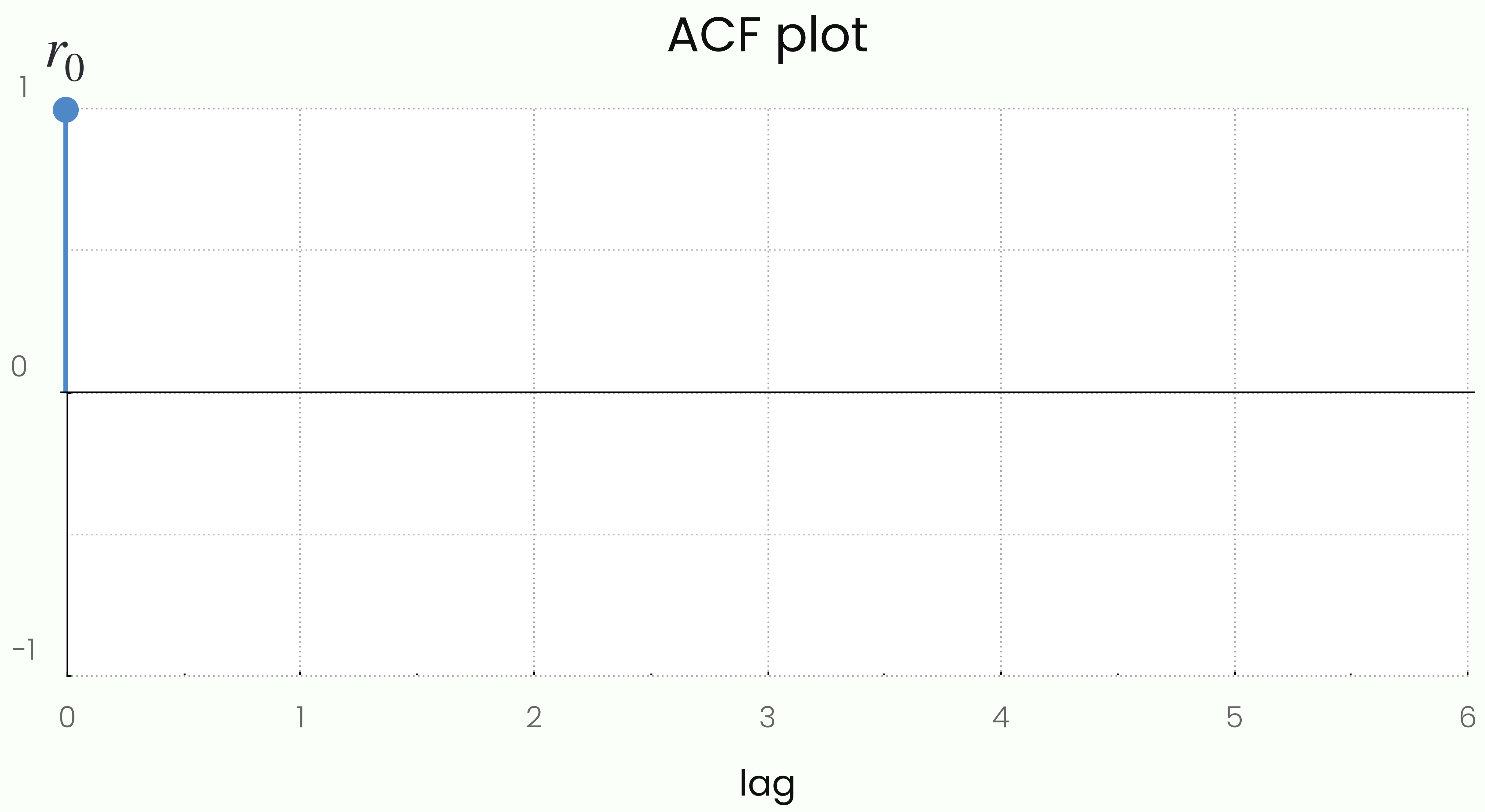
Autocorrelation coefficient r_i measures the relationship between y_t and y_{t-i}.

AUTOCORRELATION PLOT

Autocorrelation coefficient r_i measures the relationship between y_t and y_{t-i}.

For example,

- r_0 measures the relationship between y_t and itself;

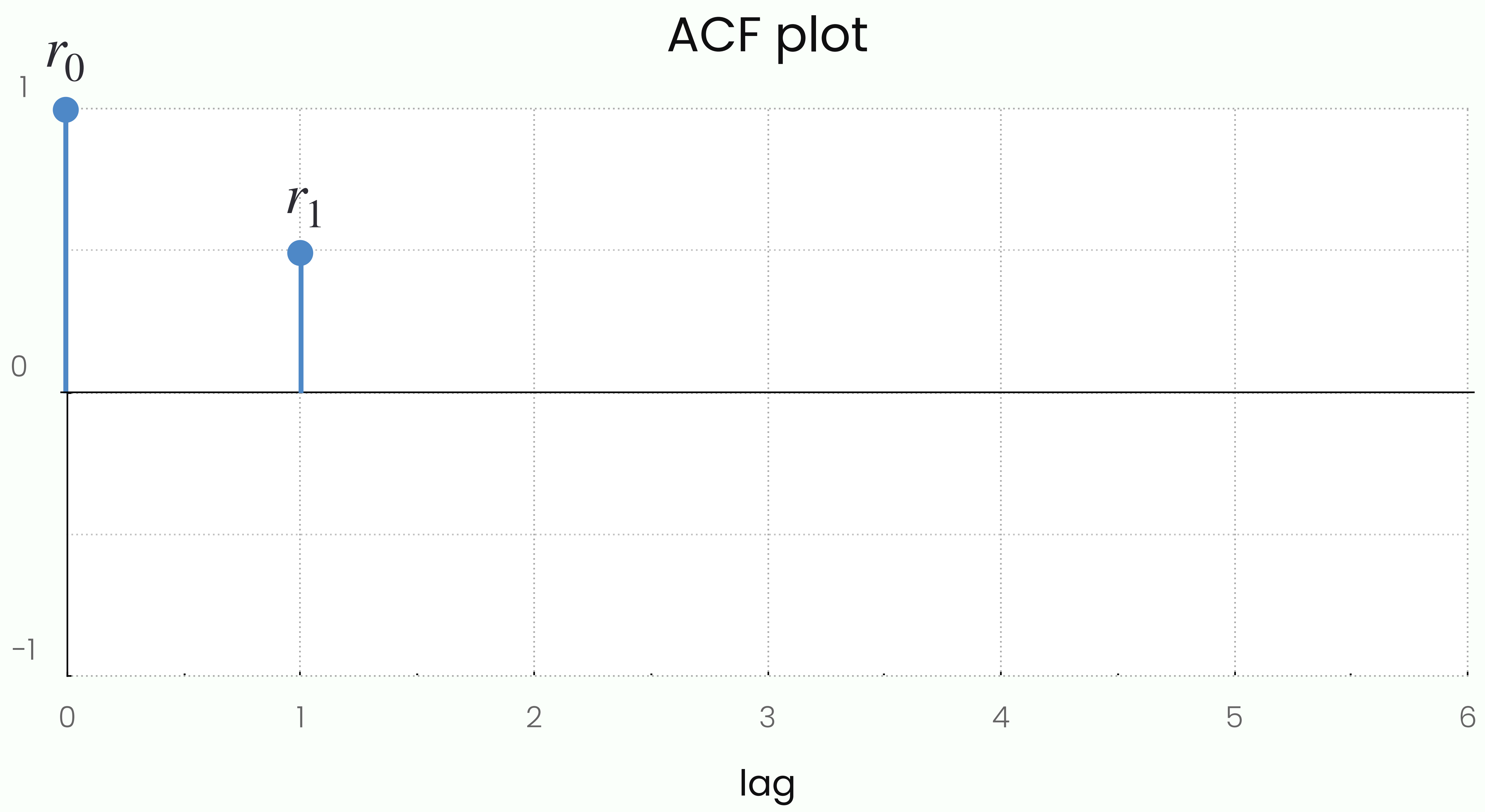


AUTOCORRELATION PLOT

Autocorrelation coefficient r_i measures the relationship between y_t and y_{t-i}.

For example,

- r_0 measures the relationship between y_t and itself;
- r_1 between y_t and y_{t-1};

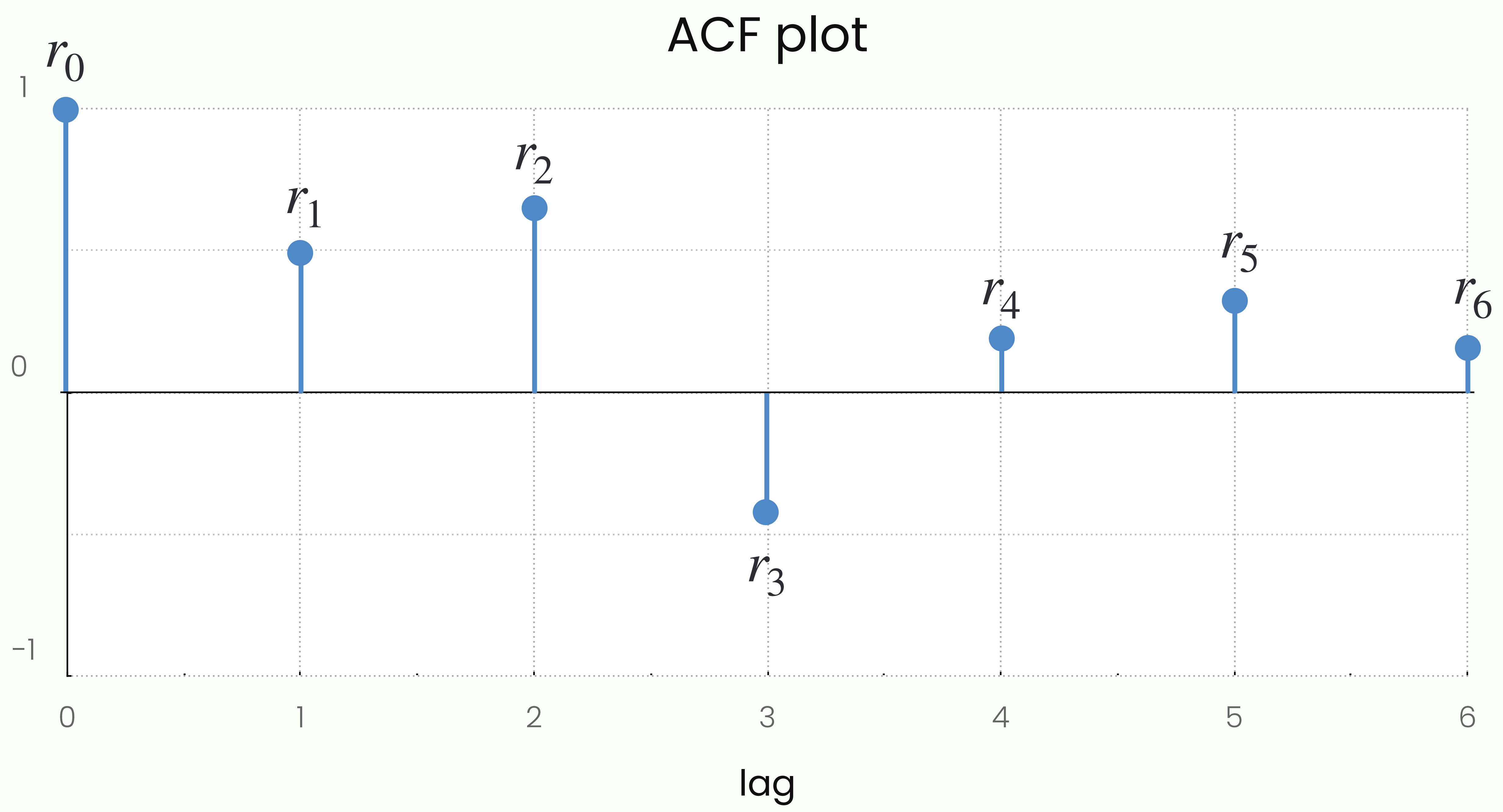


AUTOCORRELATION PLOT

Autocorrelation coefficient r_i measures the relationship between y_t and y_{t-i}.

For example,

- r_0 measures the relationship between y_t and itself;
- r_1 between y_t and y_{t-1};
- r_2 between y_t and y_{t-2};
- ...

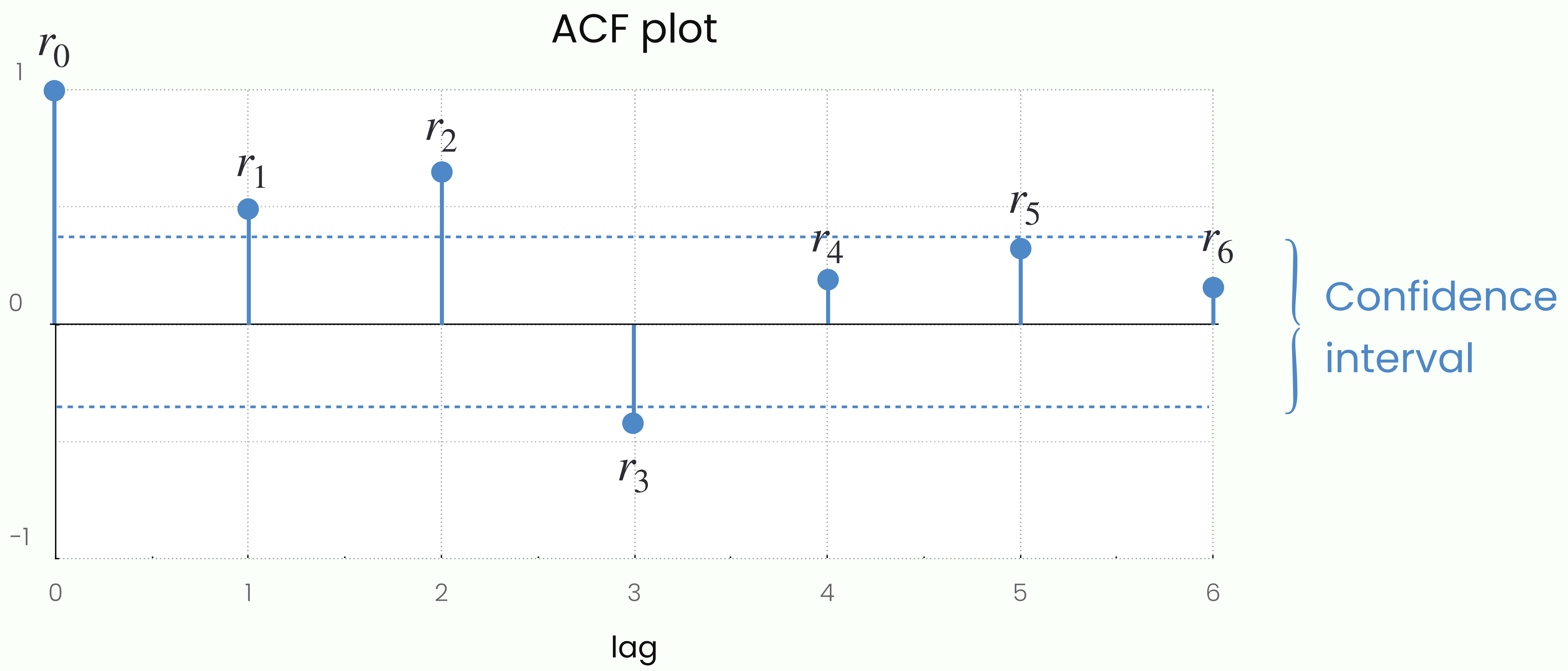


AUTOCORRELATION PLOT

Autocorrelation coefficient r_i measures the relationship between y_t and y_{t-i}.

For example,

- r_0 measures the relationship between y_t and itself;
- r_1 between y_t and y_{t-1};
- r_2 between y_t and y_{t-2};
- ...



INTERPRETATION

Positive ACF at lag i indicates a positive correlation between the current observation (y_t) and the observation at lag i (y_{t-i}).

Negative ACF at lag i indicates a negative correlation between the current observation (y_t) and the observation at lag i (y_{t-i}).

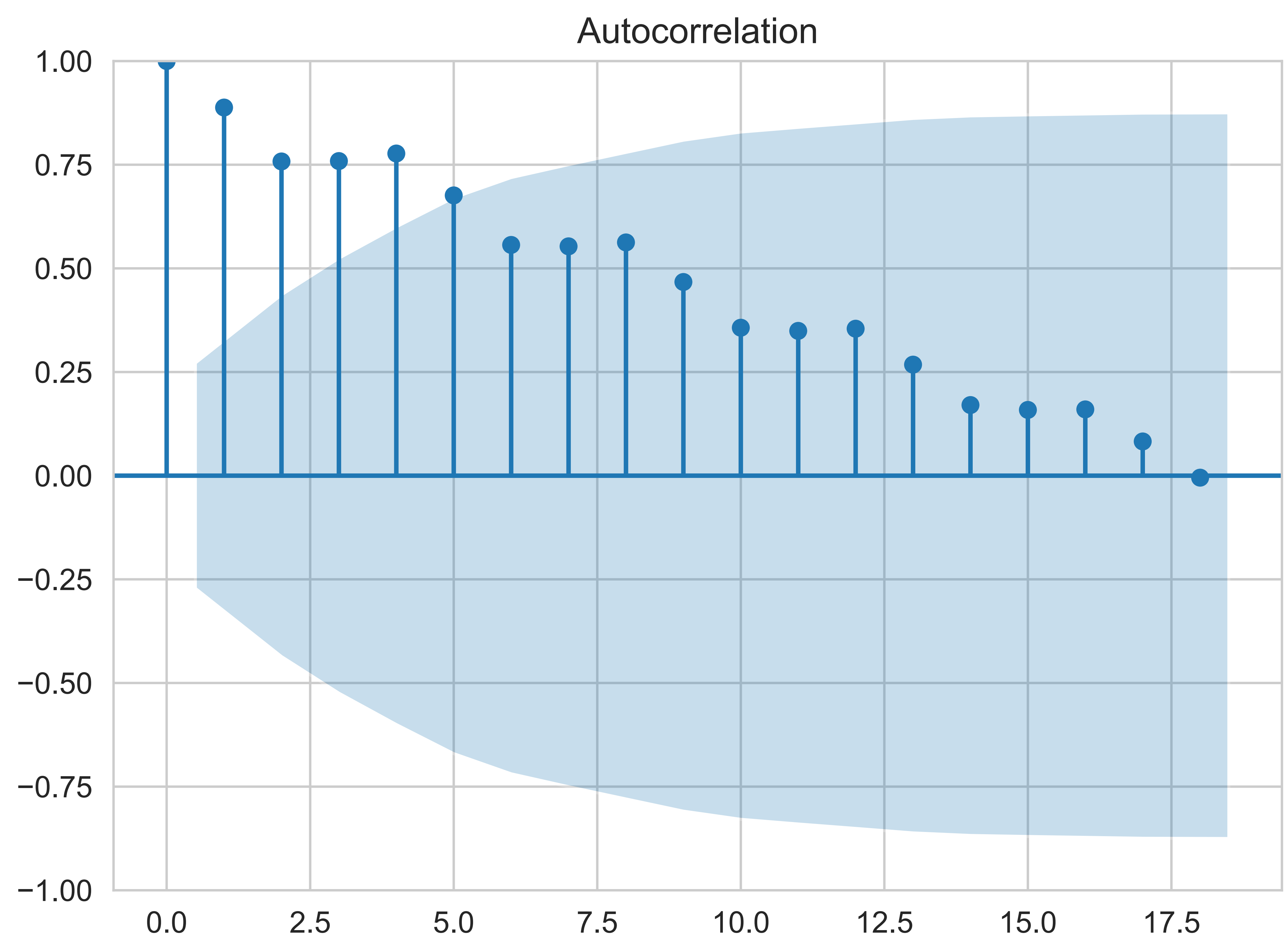
Decay in ACF as lag increases often signifies the presence of a trend or seasonality in the time series.

Significant ACF values at certain lags may suggest potential patterns or relationships in the time series.

ACF: EXAMPLE

Recall original Australian quarterly electricity production data.

lag (1Q)	0	1	2	3	4	5
AC coef (r_i)	1	0.8879	0.7579	0.7588	0.777	0.676



When data have a trend, the autocorrelations for small lags tend to be large and positive because observations nearby in time are also nearby in value.

That's why ACF for the non-stationary data does not make sense.

ACF: DATA TRANSFORMATION

Seasonal difference (m is the number of seasons):

$$y'_t = y_t - y_{t-m} = (1 - B^m)y_t$$

Difference:

$$y''_t = y'_t - y'_{t-1} = (1 - B)y_t$$

ACF: DATA TRANSFORMATION

Seasonal difference (m is the number of seasons):

$$y'_t = y_t - y_{t-m} = (1 - B^m)y_t$$

Difference:

$$y''_t = y'_t - y'_{t-1} = (1 - B)y_t$$

The final transformation (for $m = 4$) is then given as:

$$\begin{aligned}(1 - B)\left(1 - B^4\right) y_t &= \left(1 - B - B^4 + B^5\right) y_t \\ &= y_t - y_{t-1} - y_{t-4} + y_{t-5}\end{aligned}$$

ACF: DATA TRANSFORMATION (2)

When both seasonal and first differences are applied:

- Order does not matter—result will be the same.
- Try seasonal first as there might no need for further first difference.

ACF: DATA TRANSFORMATION (2)

When both seasonal and first differences are applied:

- Order does not matter—result will be the same.
- Try seasonal first as there might no need for further first difference.

NB! Differences must be interpretable.

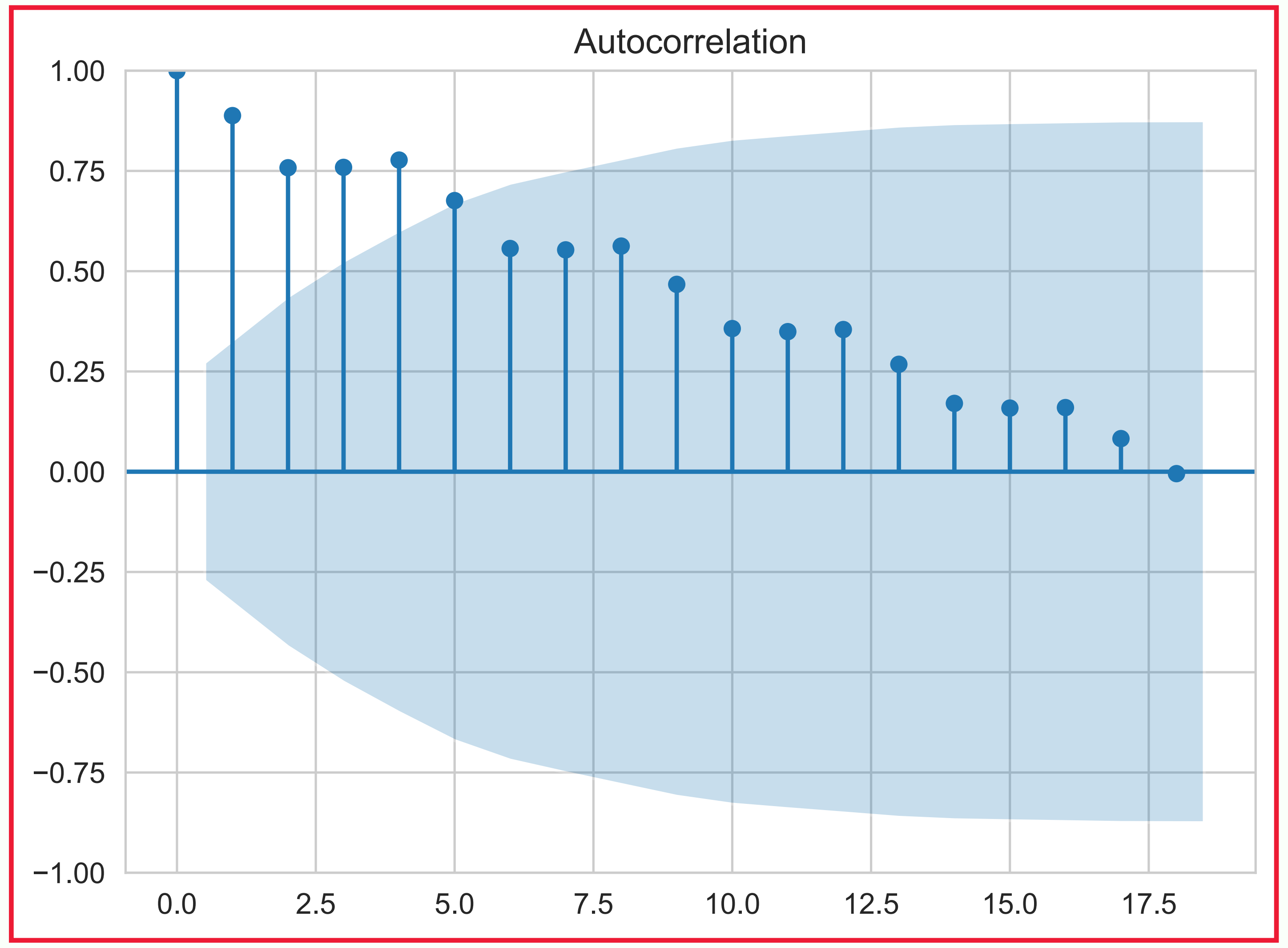
✓ First differences are the change between close observations.

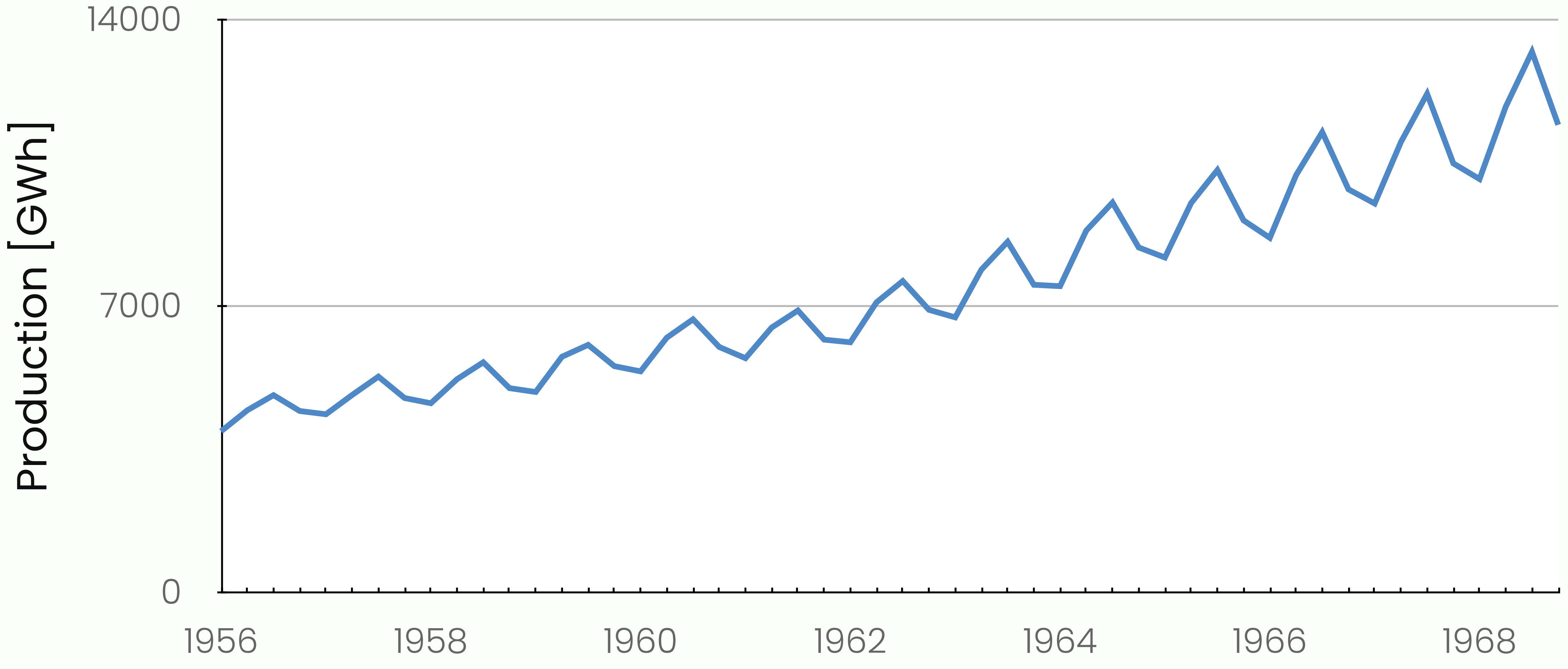
✓ Seasonal differences are the change between corresponding observations.

✓ But, for example, $(1 - B^3)$ difference (lag3) for quarterly data is not interpretable.

ACF: EXAMPLE

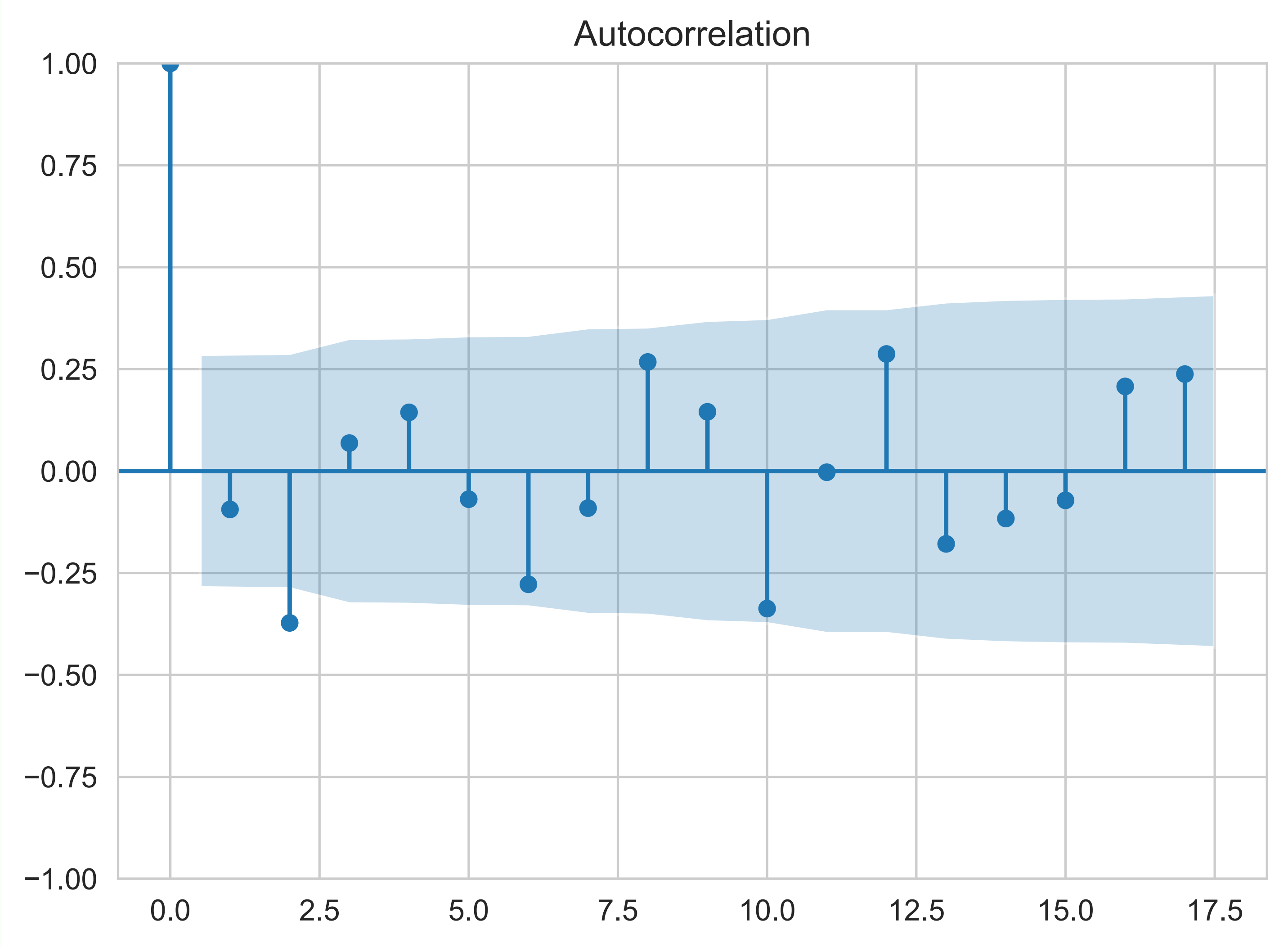
ORIGINAL

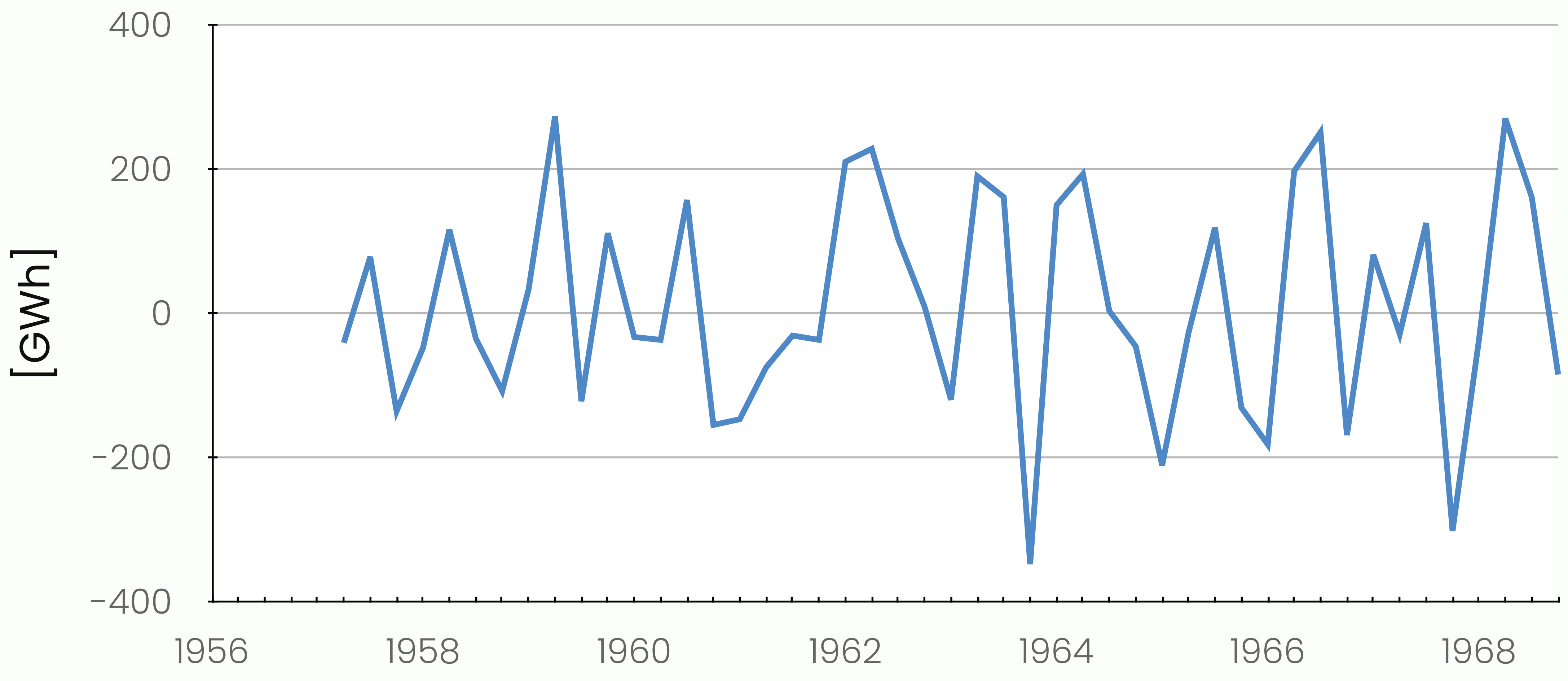




TRANSFORMED

SEAS + 1st DIFF





PARTIAL ACF (PACF)

Partial ACF measures the correlation between a variable and its lagged values, excluding the linear influence of the intermediate lags.

PACF: INTUITION

Forecast average monthly electricity price.

y_t	December
y_{t-1}	November
y_{t-2}	October
	$\vdots$

$$y_{t-2} \qquad y_{t-1} \qquad y_t$$

PACF: INTUITION

Forecast average monthly electricity price.

y_t	December
y_{t-1}	November
y_{t-2}	October
$\vdots$	

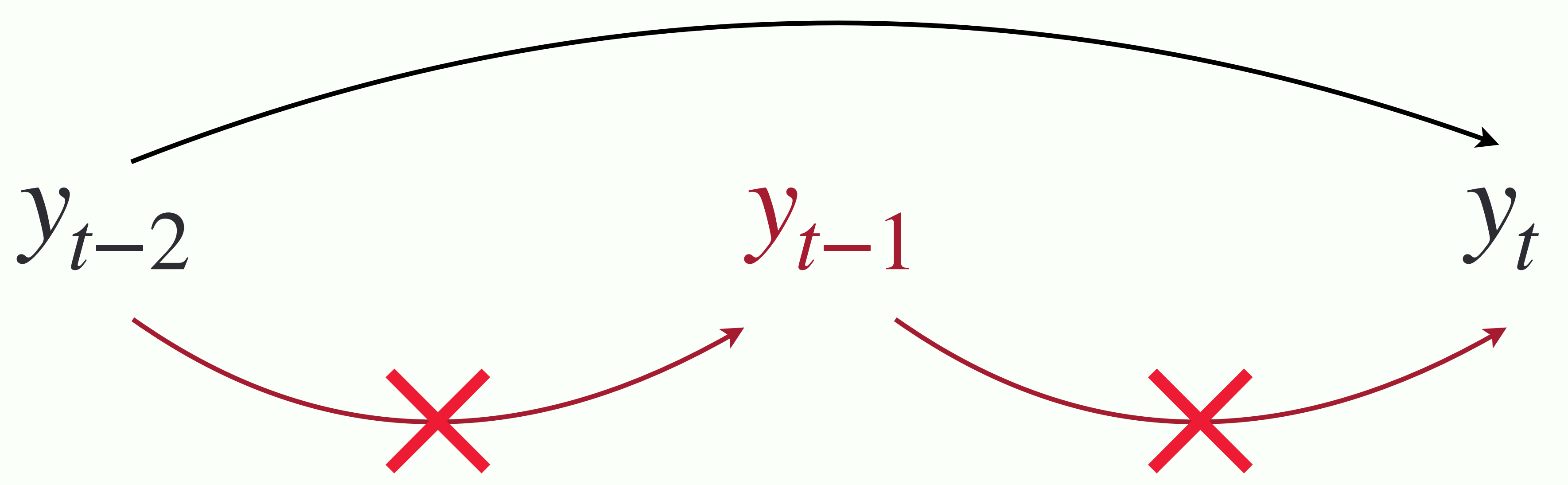
$$
\begin{tikzcd}
y_{t-2} \arrow[rr, bend left] \arrow[r, bend right] & y_{t-1} \arrow[r, bend right] & y_t
\end{tikzcd}
$$

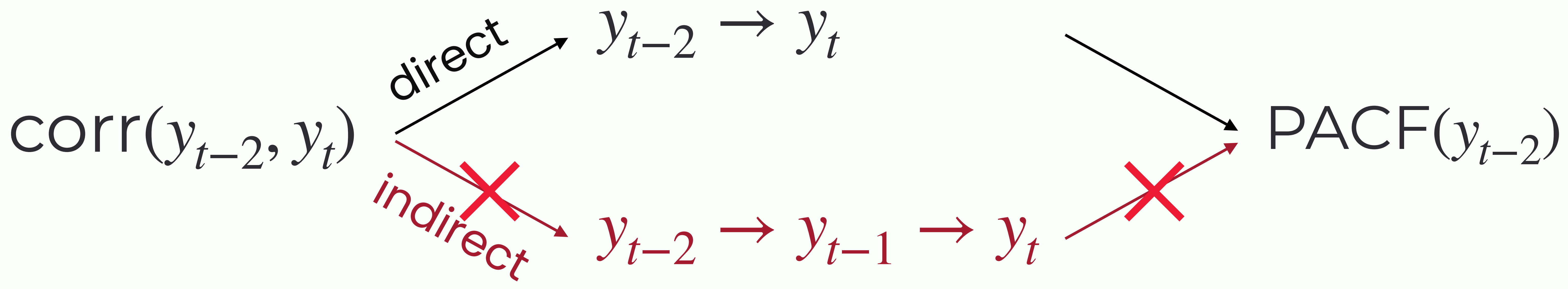
$$
\begin{tikzcd}
 & y_{t-2} \to y_t \arrow[dr] & \\
\mathrm{corr}(y_{t-2}, y_t) \arrow[ur, "\text{direct}"] \arrow[dr, "\text{indirect}"'] & & \mathrm{PACF}(y_{t-2}) \\
 & y_{t-2} \to y_{t-1} \to y_t \arrow[ur] &
\end{tikzcd}
$$

PACF: INTUITION

Forecast average monthly electricity price.

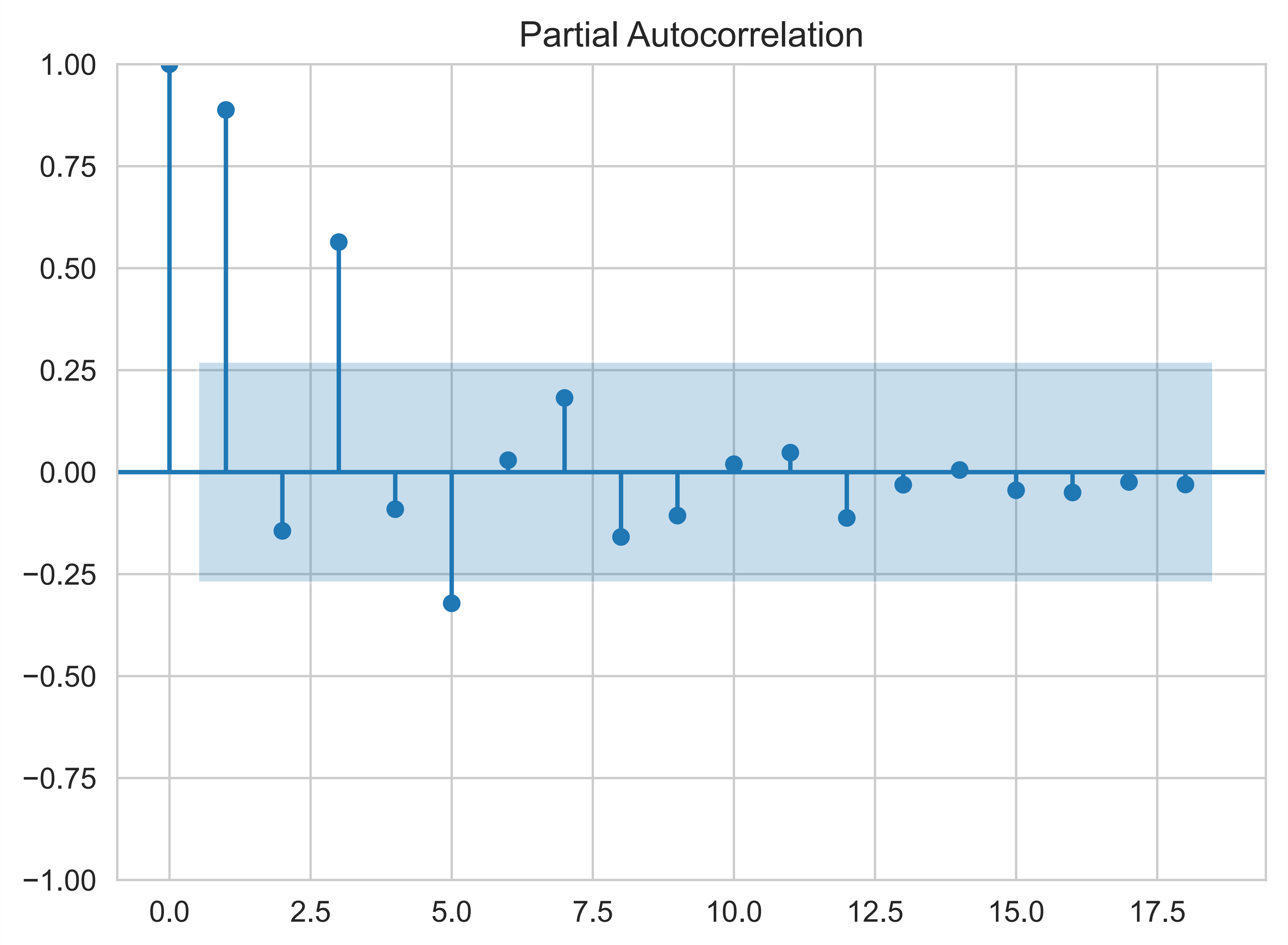
y_t	December
y_{t-1}	November
y_{t-2}	October
$\vdots$	





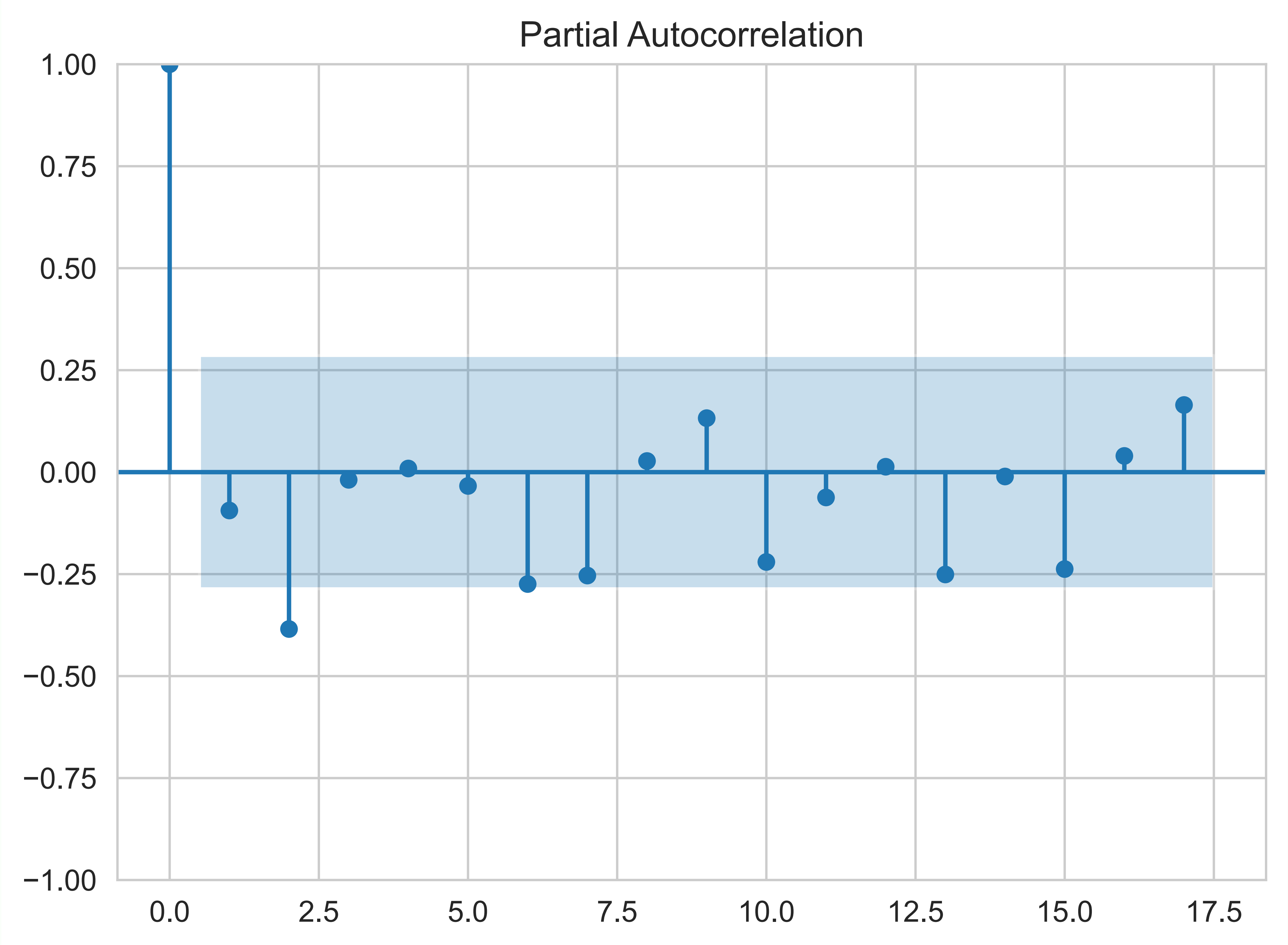
EXAMPLE

ORIGINAL



TRANSFORMED

SEAS + 1st DIFF

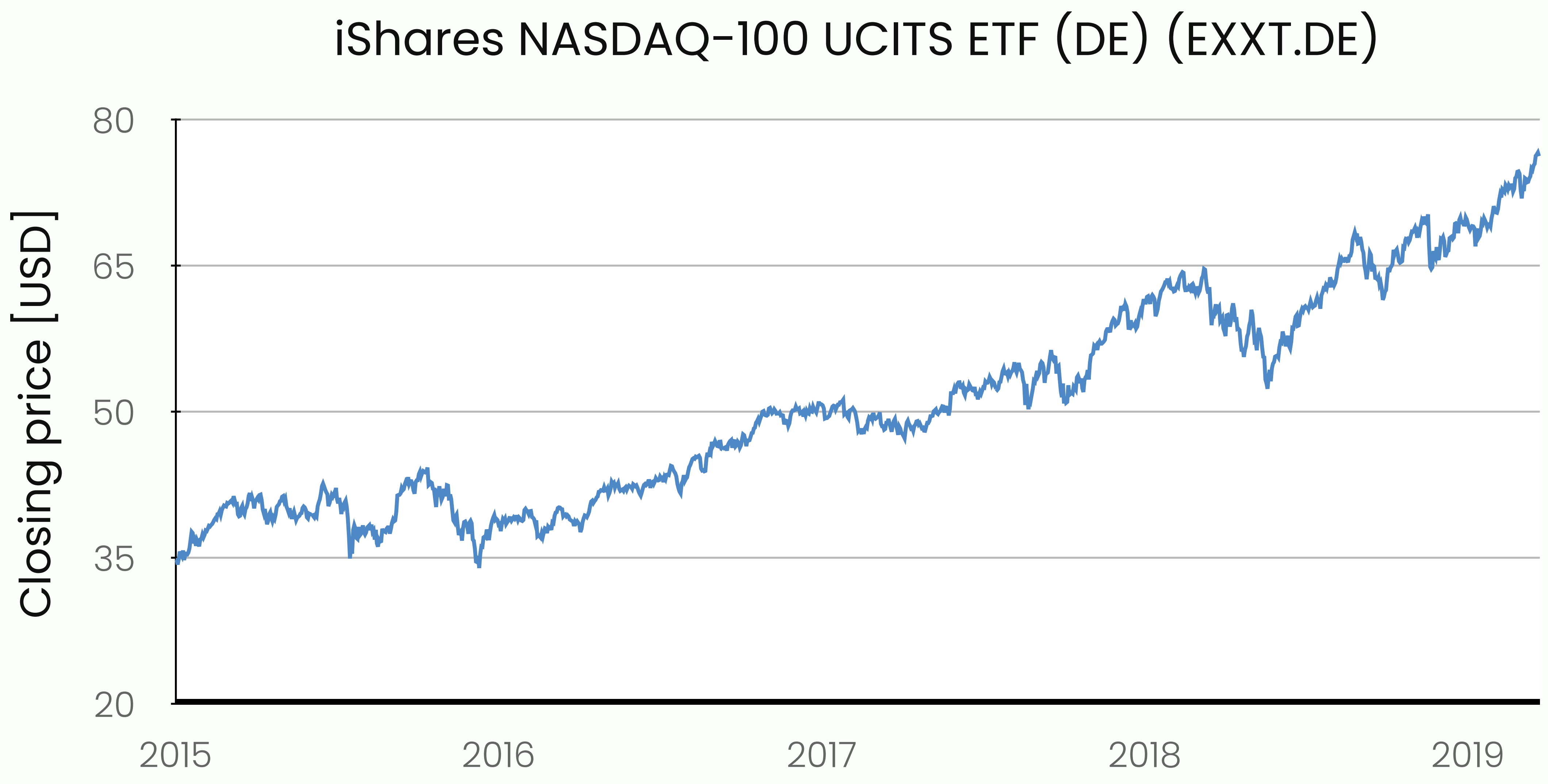


ACF VS PACF

Autocorrelation (ACF) and Partial Autocorrelation (PACF) are both measures used in time series analysis to understand the relationships between observations at different time points.

ACF	PACF
Shows order of a moving average (MA) process	Shows order of an autoregressive (AR) process
Represents the overall correlation structure of the time series	Highlights the direct relationships between observations at specific lags
Measures the linear relationship between an observation and its previous observations at different lags	Measures the direct linear relationship between an observation and its previous observations at a specific lag, excluding the contributions from intermediate lags

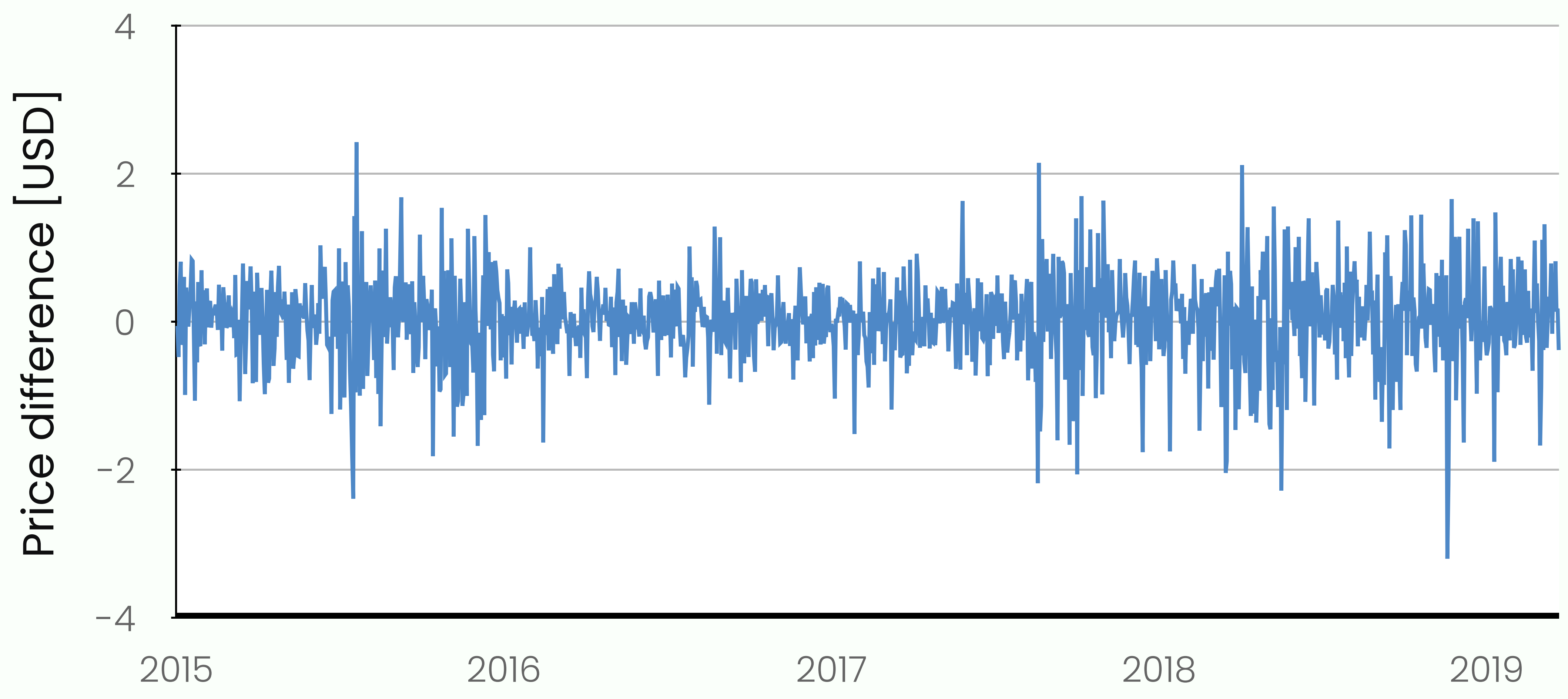
ACF & PACF AT STOCK MARKET



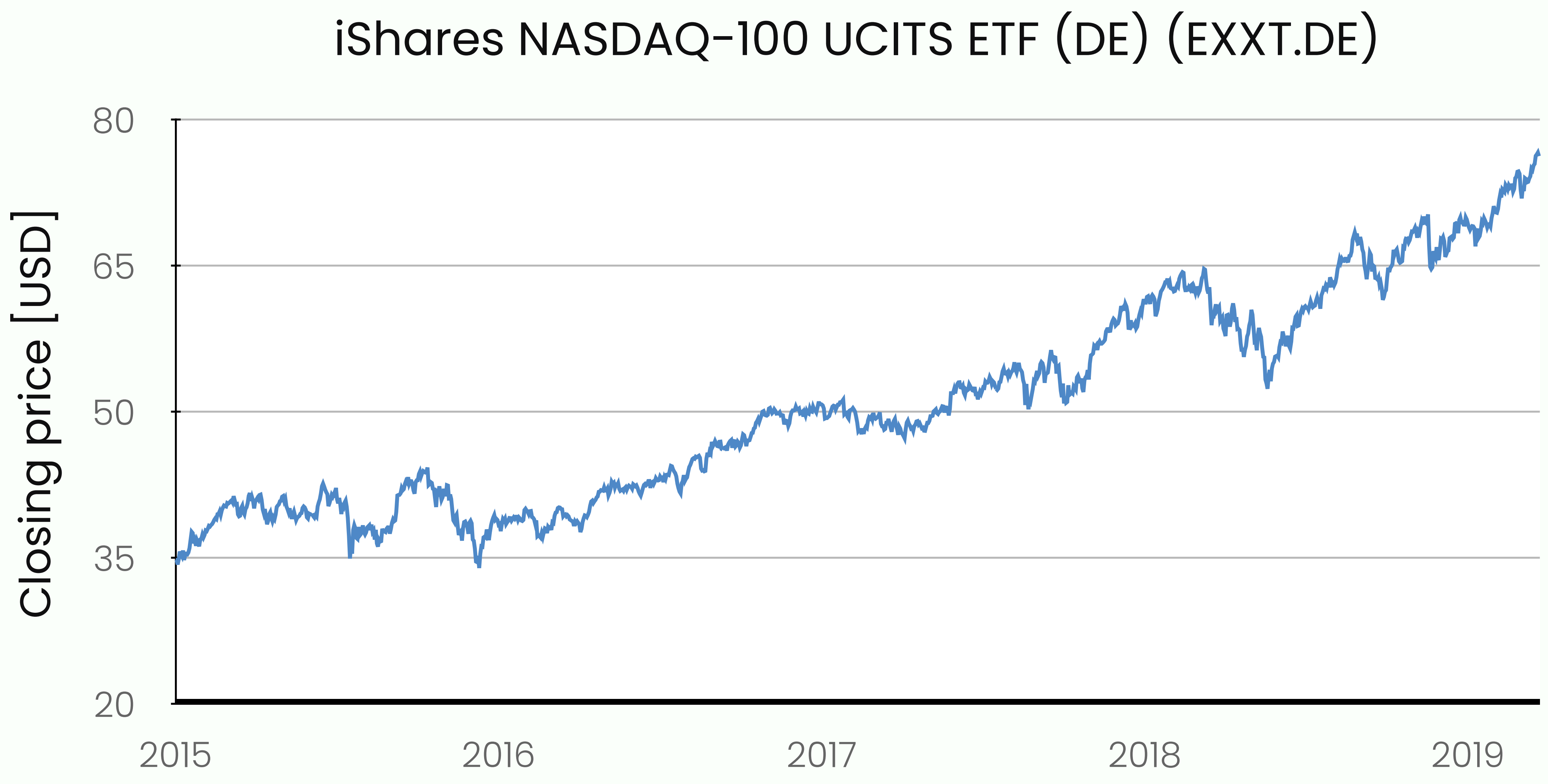
ACF & PACF AT STOCK MARKET

iShares NASDAQ-100 UCITS ETF (DE) (EXXT.DE)

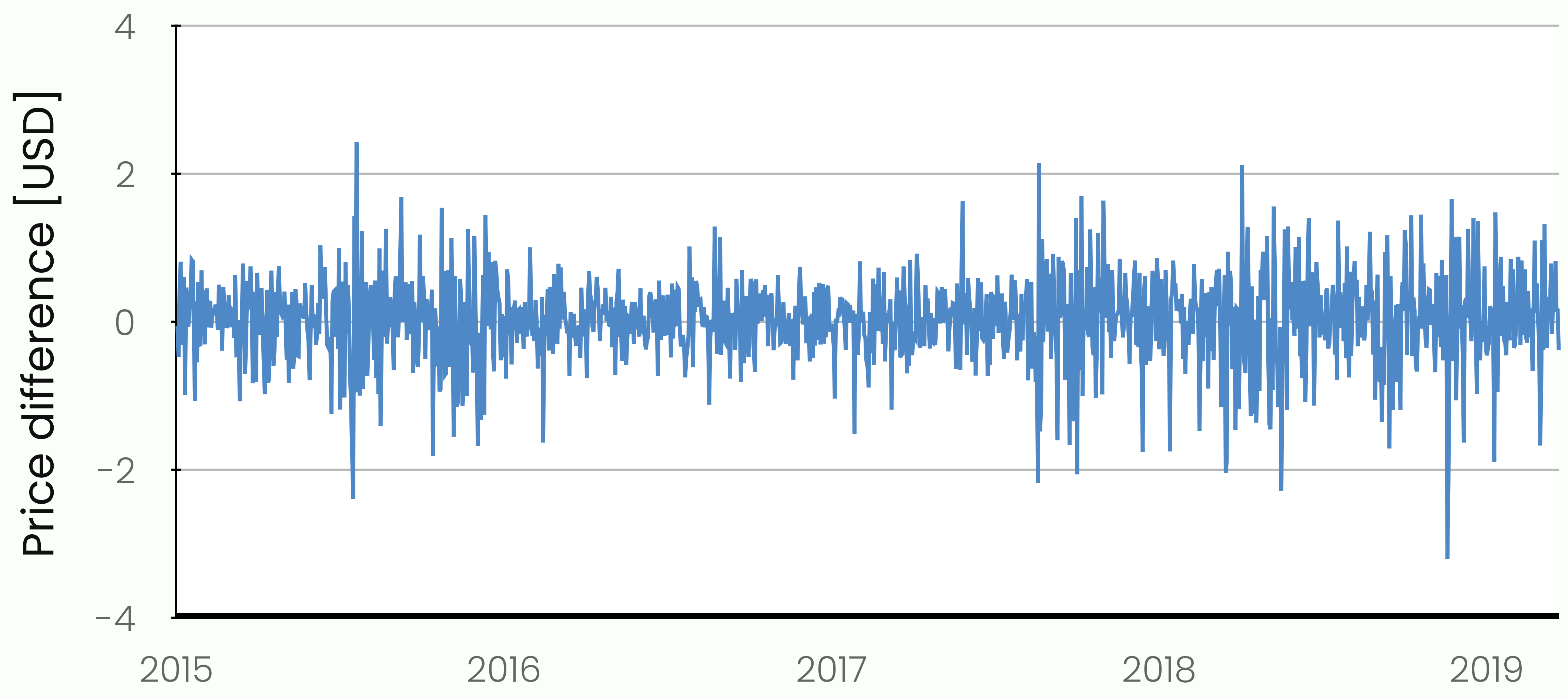
$$(1 - B)y_t$$



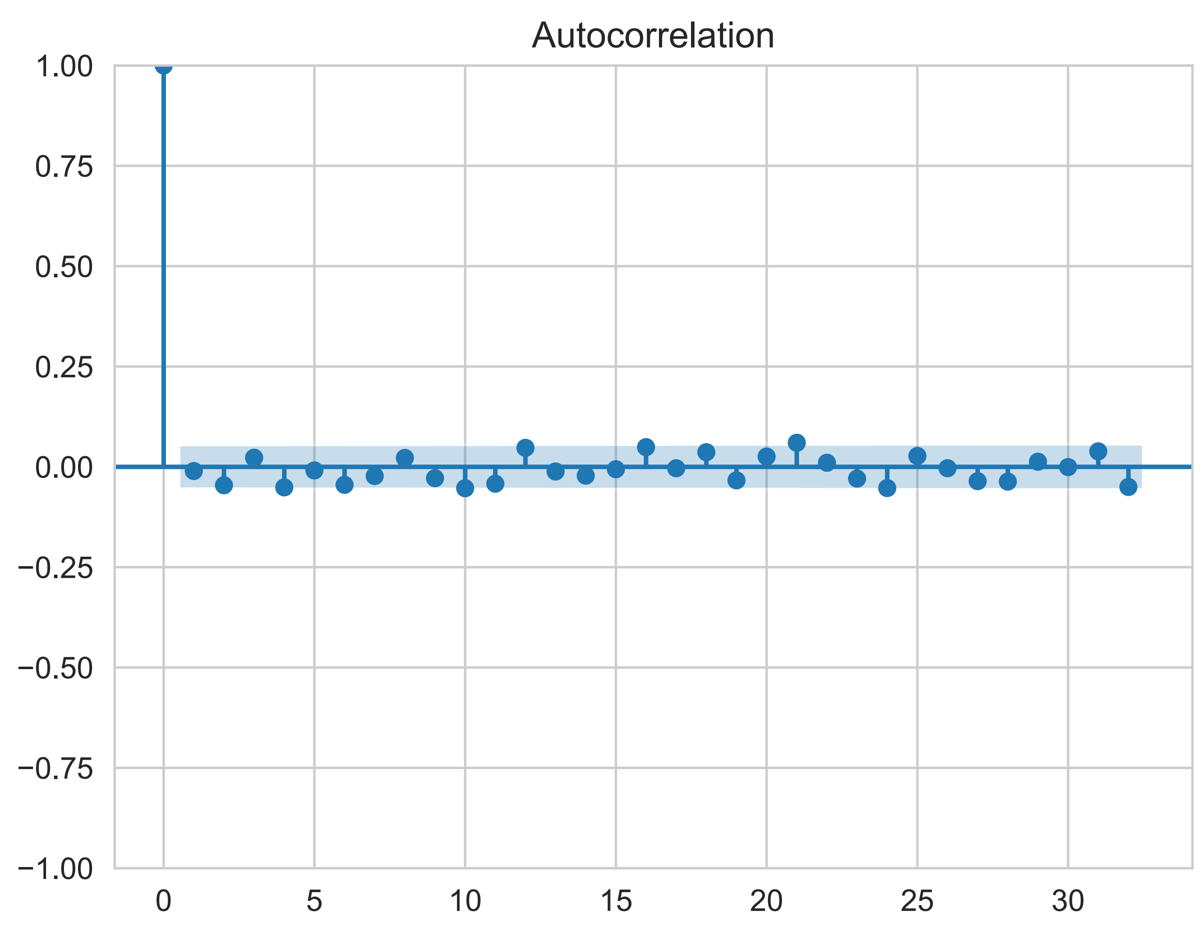
ACF & PACF AT STOCK MARKET

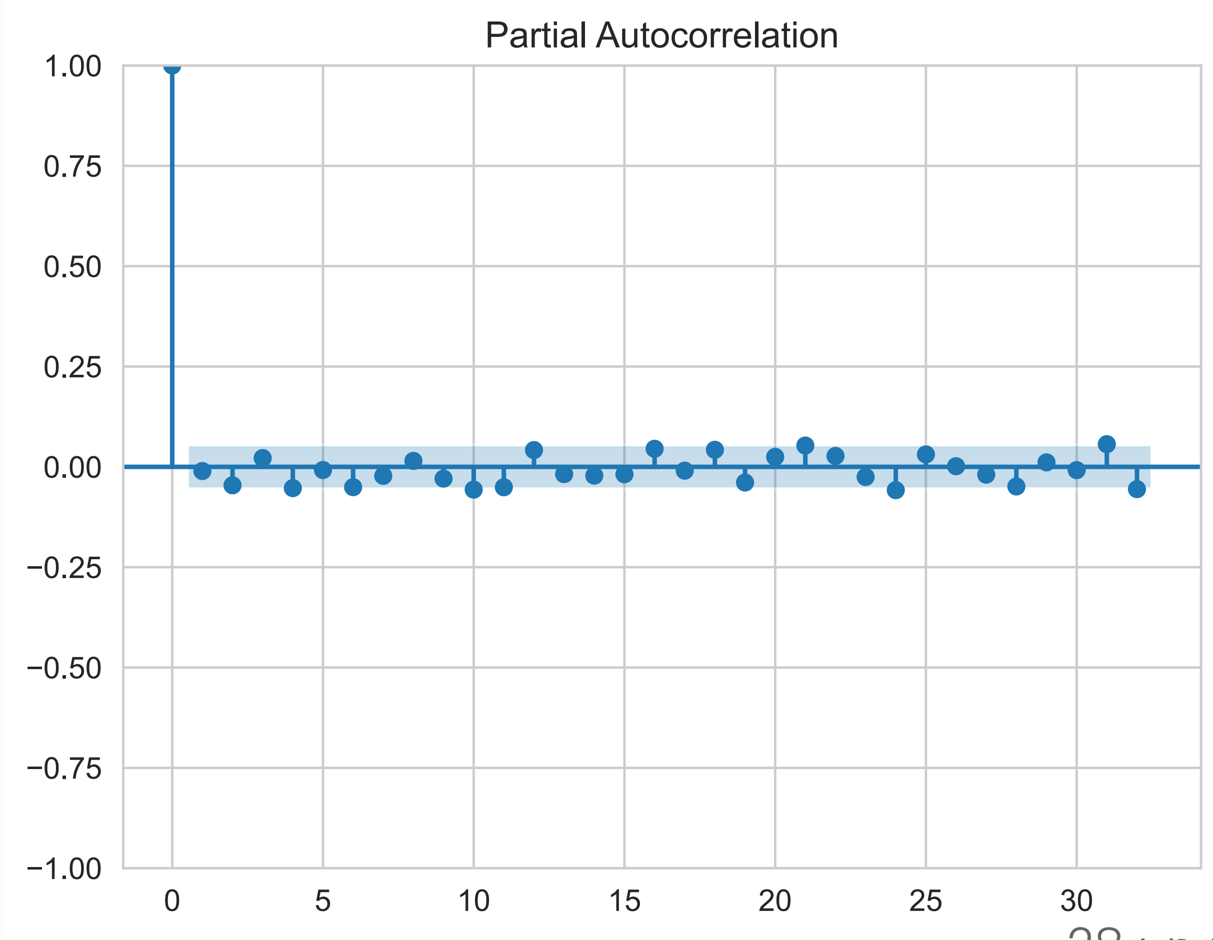


$(1-B)y_t$



No significant lags in both ACF and PACF





MODELLING METHODS

TRADITIONAL (STATISTICAL)

- Simple methods
- Regression methods
- ARMA family methods
- Exponential smoothing
- Kalman filter

MACHINE LEARNING

- Artificial neural networks
- Support vector machines
- Fuzzy logic
- Decision tree
- Genetic algorithm
- Knowledge-based expert systems
- LSTM

Autoregressive Models

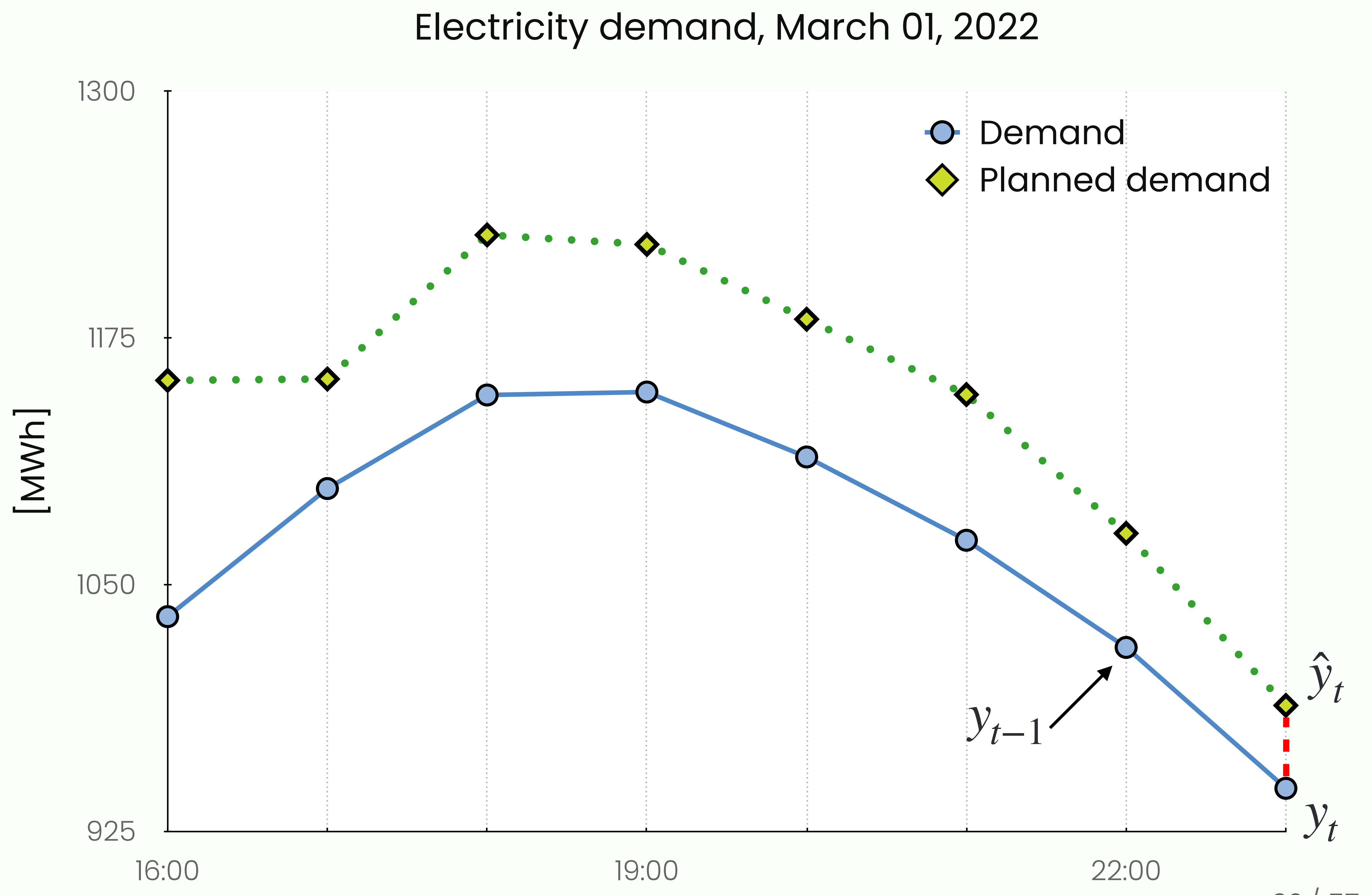
AUTOREGRESSION

The term *autoregression* indicates that it is a regression of the variable against itself (*past* values of the *same* variable).

AR(1) MODEL

For example, autoregressive model of order 1 (or AR(1)) can be written as

$$y_t = c + \phi y_{t-1} + \epsilon_t.$$

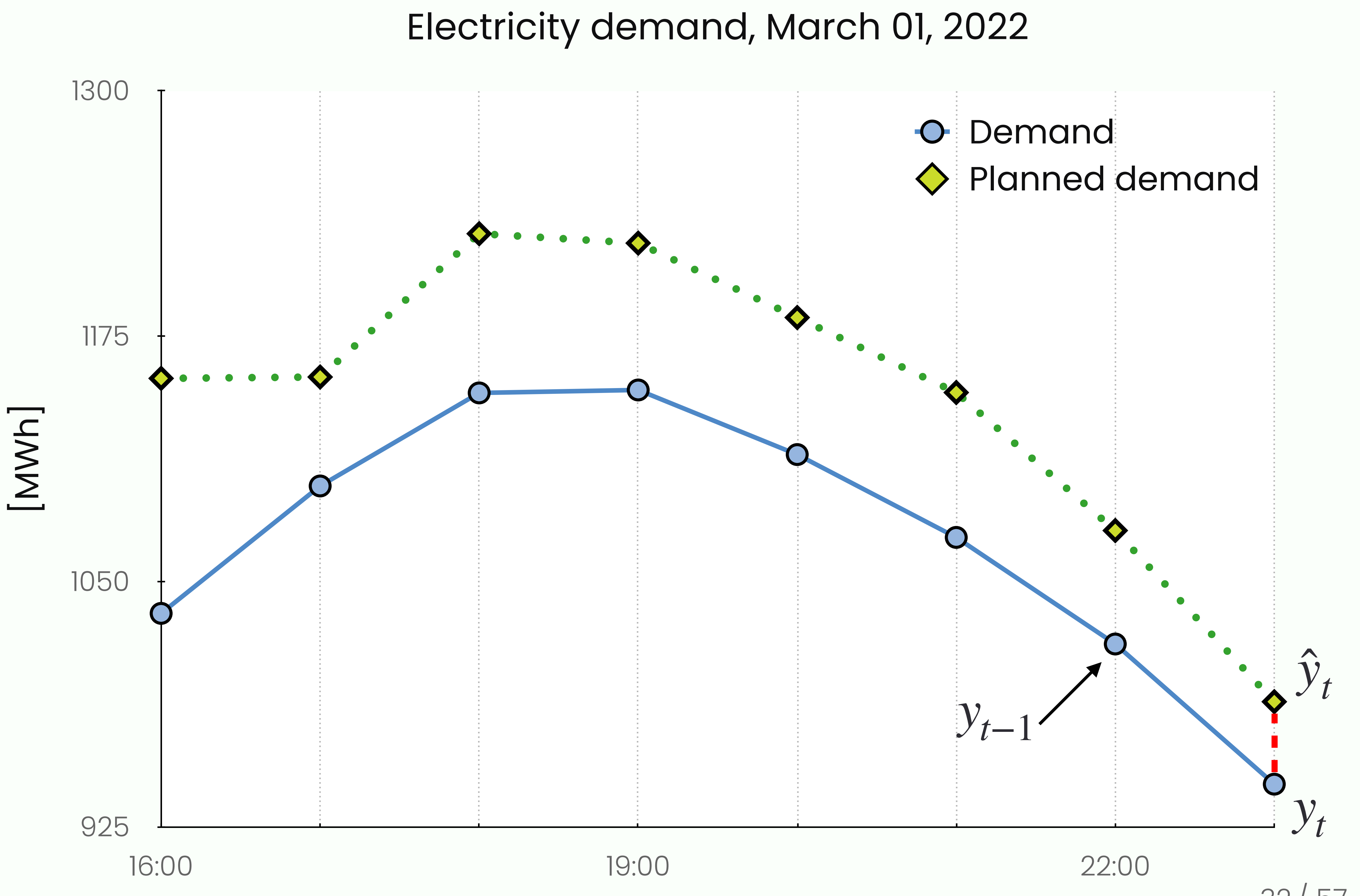


AR(1) MODEL

For example, autoregressive model of order 1 (or AR(1)) can be written as

$$y_t = c + \phi y_{t-1} + \epsilon_t.$$

target



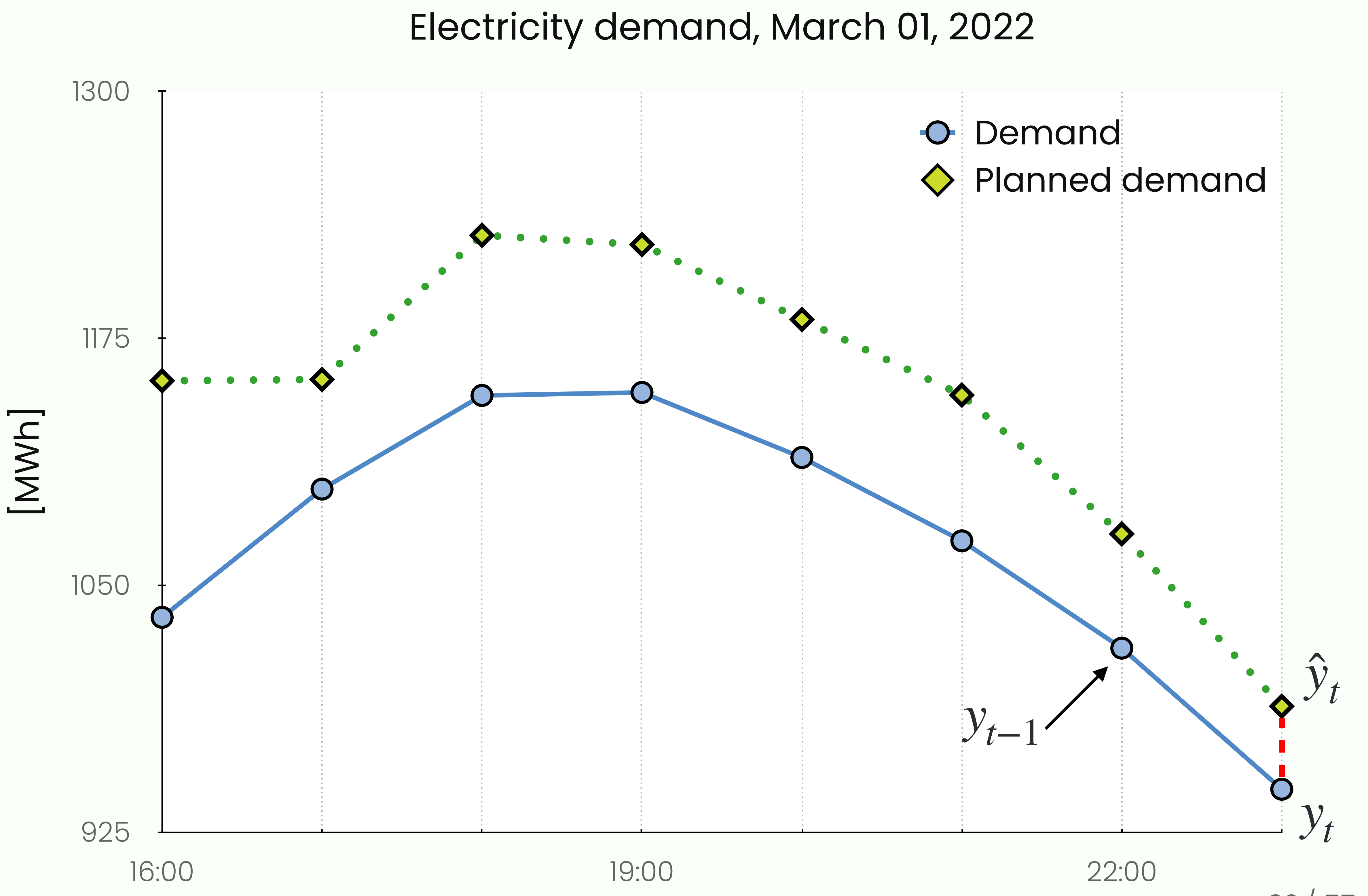
AR(1) MODEL

For example, autoregressive model of order 1 (or AR(1)) can be written as

intercept coefficient

$$y_t = c + \phi y_{t-1} + \epsilon_t.$$

target



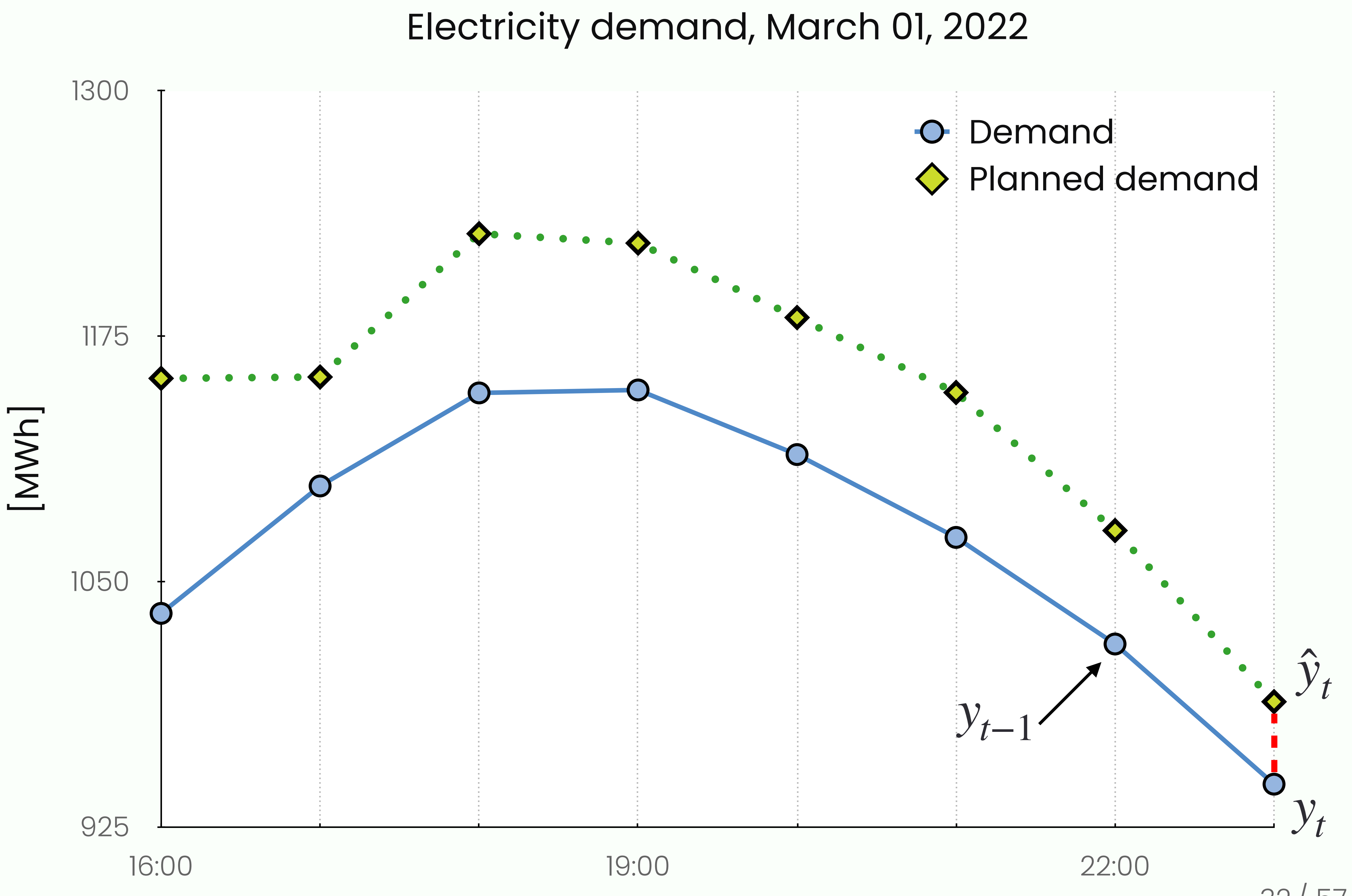
AR(1) MODEL

For example, autoregressive model of order 1 (or AR(1)) can be written as

intercept coefficient error

$$y_t = c + \phi y_{t-1} + \epsilon_t.$$

target



Practical observation (there are no perfect models):

$$\hat{y}_t = c + \phi y_{t-1} + \phi_2 y_{t-2}.$$

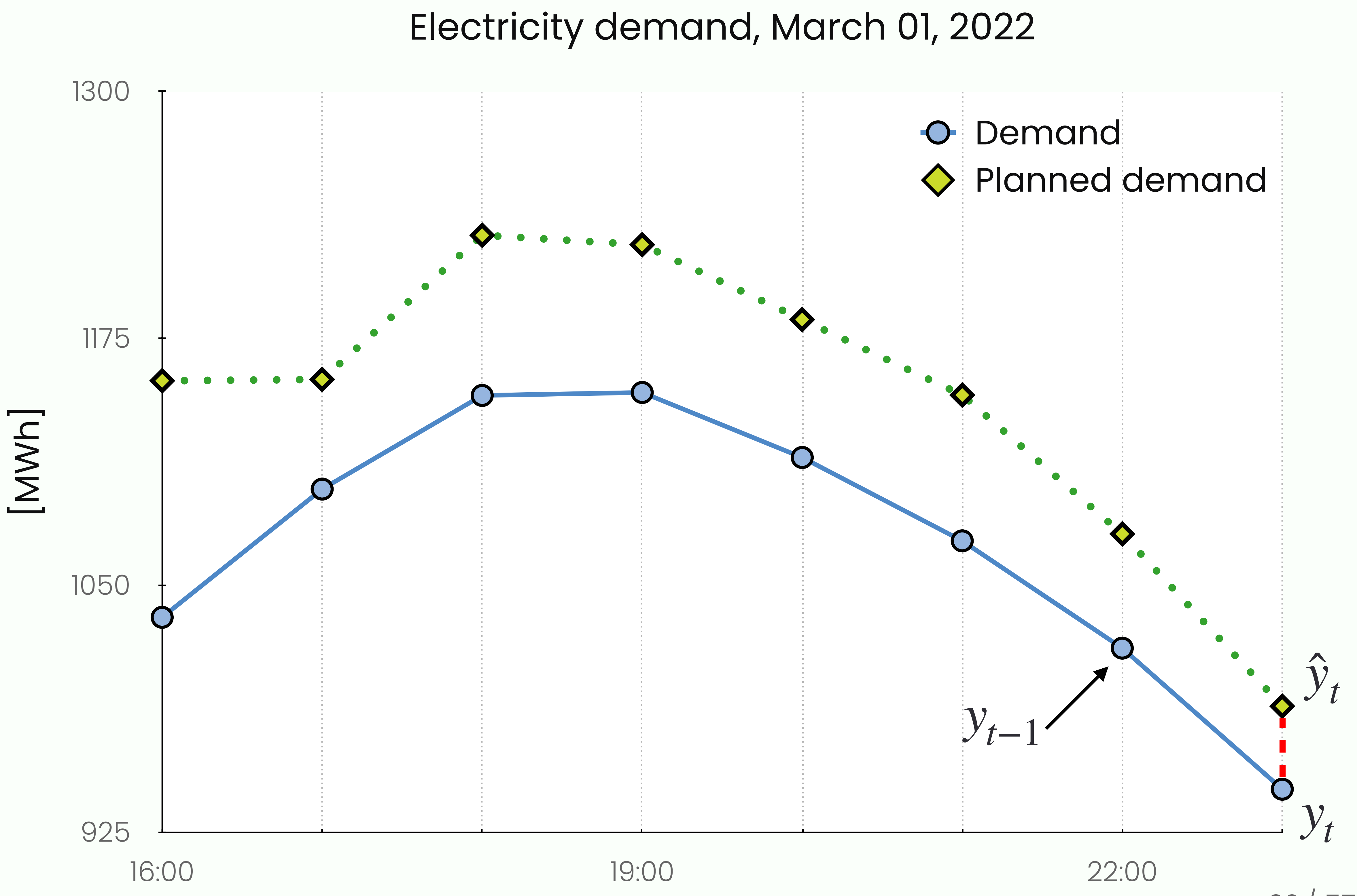
AR(1) MODEL

For example, autoregressive model of order 1 (or AR(1)) can be written as

$$\underbrace{y_t}_{\text{target}} = \overbrace{c}^{\text{intercept}} + \overbrace{\phi}^{\text{coefficient}} \underbrace{y_{t-1}}_{\text{lagged target}} + \overbrace{\epsilon_t}^{\text{error}}.$$

Practical observation (there are no perfect models):

$$\hat{y}_t = c + \phi y_{t-1} + \phi_2 y_{t-2}.$$



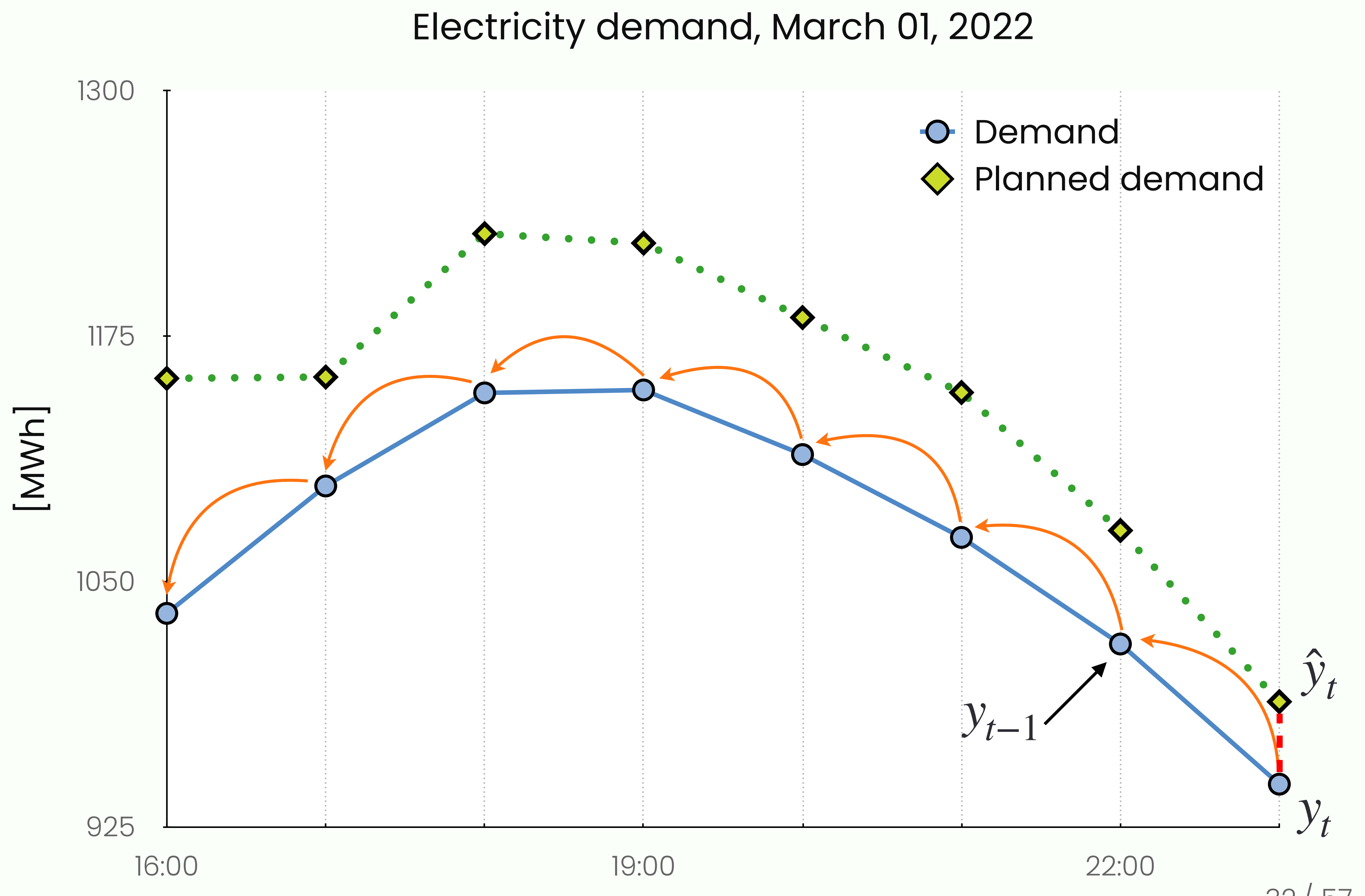
AR(1) MODEL

For example, autoregressive model of order 1 (or AR(1)) can be written as

intercept coefficient error

$$y_t = c + \phi y_{t-1} + \epsilon_t.$$

target lagged target



Practical observation (there are no perfect models):

$$\hat{y}_t = c + \phi y_{t-1} + \phi_2 y_{t-2}.$$

LONG MEMORY MODELS

Recursion goes back until the beginning of the series.

$$\boxed{y_t = c + \phi y_{t-1} + \epsilon_t}$$

$$y_{t-1} = c + \phi y_{t-2} + \epsilon_{t-1}$$

$$y_{t-2} = c + \phi y_{t-3} + \epsilon_{t-2}$$

LONG MEMORY MODELS

Recursion goes back until the beginning of the series.

$$\boxed{y_t = c + \phi y_{t-1} + \epsilon_t}$$

$$y_{t-1} = c + \phi y_{t-2} + \epsilon_{t-1}$$

$$y_{t-2} = c + \phi y_{t-3} + \epsilon_{t-2}$$

$$y_t = c + \phi(c + \phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t$$

LONG MEMORY MODELS

Recursion goes back until the beginning of the series.

$$\boxed{y_t = c + \phi y_{t-1} + \epsilon_t}$$

$$y_{t-1} = c + \phi y_{t-2} + \epsilon_{t-1}$$

$$y_{t-2} = c + \phi y_{t-3} + \epsilon_{t-2}$$

$$y_t = c + \phi(c + \phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t$$

$$\Big\downarrow \text{simplify}$$

$$y_t = c^* + \phi^2 y_{t-2} + \phi \epsilon_{t-1} + \epsilon_t$$

LONG MEMORY MODELS

Recursion goes back until the beginning of the series.

$$\boxed{y_t = c + \phi y_{t-1} + \epsilon_t}$$

$$y_{t-1} = c + \phi y_{t-2} + \epsilon_{t-1}$$

$$y_{t-2} = c + \phi y_{t-3} + \epsilon_{t-2}$$

$$y_t = c + \phi(c + \phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t$$

$$\Big\downarrow \text{simplify}$$

$$y_t = c^* + \phi^2 y_{t-2} + \phi\epsilon_{t-1} + \epsilon_t$$

$$\Big\downarrow$$

$$y_t = c^* + \phi^3 y_{t-3} + \phi^2\epsilon_{t-2} + \phi\epsilon_{t-1} + \epsilon_t$$

LONG MEMORY MODELS

Recursion goes back until the beginning of the series.

$$\boxed{y_t = c + \phi y_{t-1} + \epsilon_t}$$

$$y_{t-1} = c + \phi y_{t-2} + \epsilon_{t-1}$$

$$y_{t-2} = c + \phi y_{t-3} + \epsilon_{t-2}$$

$$y_t = c + \phi(c + \phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t$$

simplify

$$y_t = c* + \phi^2 y_{t-2} + \phi \epsilon_{t-1} + \epsilon_t$$

$$y_t = c* + \phi^3 y_{t-3} + \phi^2 \epsilon_{t-2} + \phi \epsilon_{t-1} + \epsilon_t$$

$$y_t = \frac{c}{1-\phi} + \phi^t y_1 + \phi^{t-1}\epsilon_2 + \phi^{t-2}\epsilon_3 + \cdots + \epsilon_t$$

LONG MEMORY MODELS

Recursion goes back until the beginning of the series.

$$\boxed{y_t = c + \phi y_{t-1} + \epsilon_t}$$

$$y_{t-1} = c + \phi y_{t-2} + \epsilon_{t-1}$$

$$y_{t-2} = c + \phi y_{t-3} + \epsilon_{t-2}$$

$$y_t = c + \phi(c + \phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t$$

simplify

$$y_t = c^* + \phi^2 y_{t-2} + \phi \epsilon_{t-1} + \epsilon_t$$

$$y_t = c^* + \phi^3 y_{t-3} + \phi^2 \epsilon_{t-2} + \phi \epsilon_{t-1} + \epsilon_t$$

$$y_t = \frac{c}{1-\phi} + \phi^t y_1 + \phi^{t-1}\epsilon_2 + \phi^{t-2}\epsilon_3 + \cdots + \epsilon_t$$

✓ Effects from past have little effect on the present if $|\phi| < 1$.

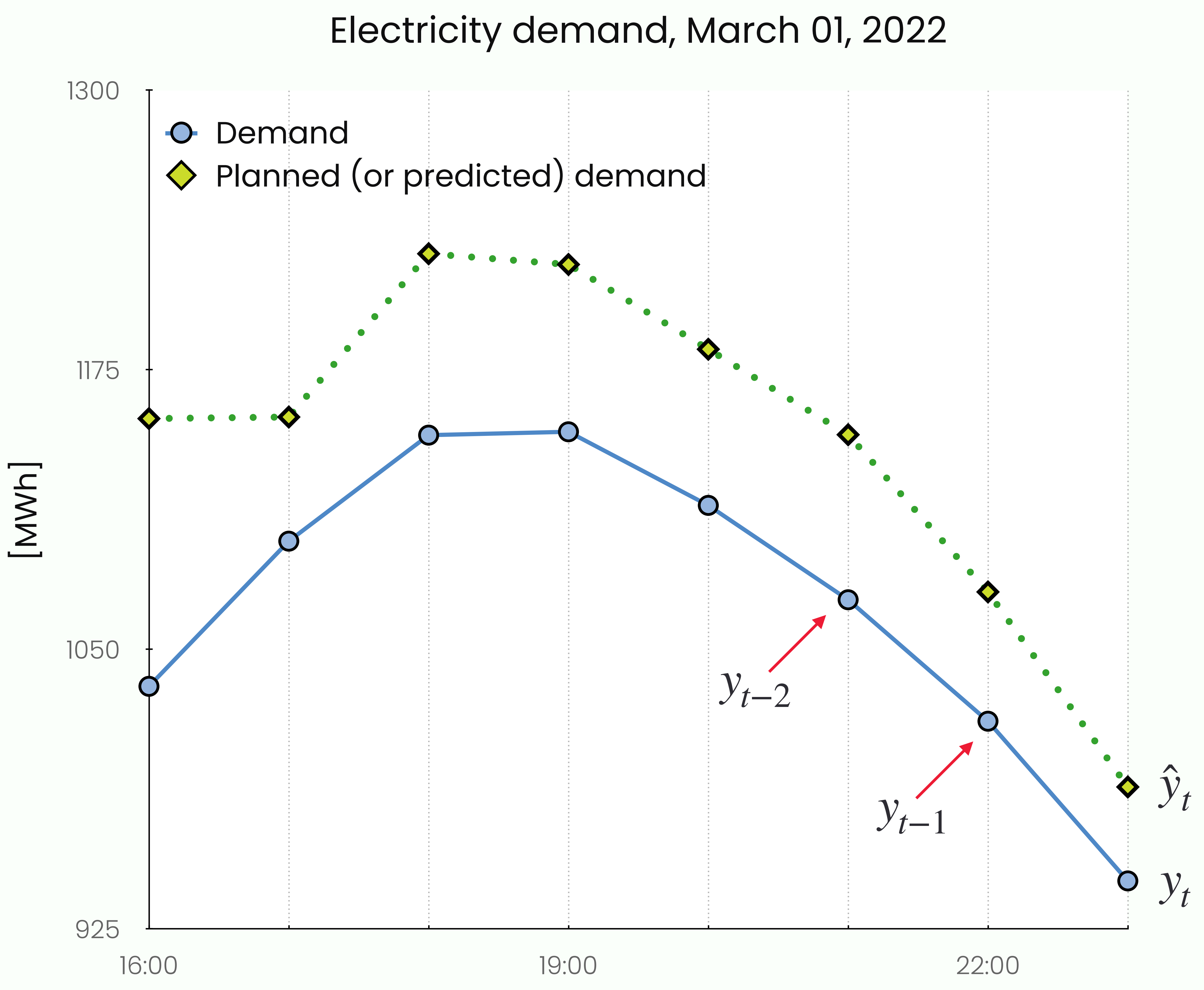
✓ Stationarity – the dependence of previous observations declines over time.

AR(2) MODEL

A linear function of 2 past values or AR(2) can be written as

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

lagged targets



AUTOREGRESSIVE (AR) MODEL

An autoregressive model of order p can be written as

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

where $\epsilon_t \sim wn(0, \sigma_w^2)$ is white noise, $\phi_1, \ldots, \phi_p$ are parameters, c is intercept.

We refer to this as an AR(p) model, an autoregressive model of order p.

MA(2) MODEL

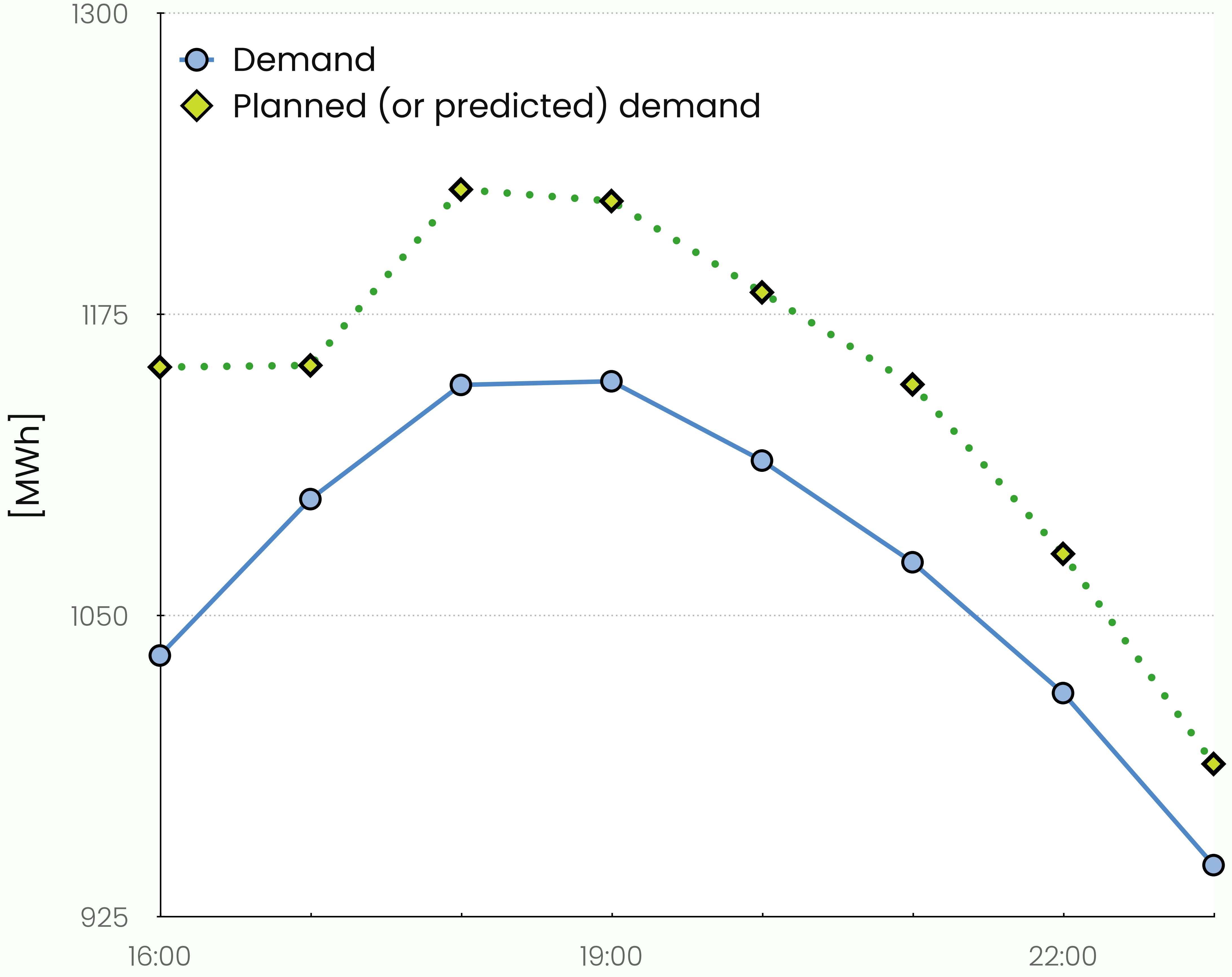
For example, moving average model MA(2) can be written as

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}.$$

MA(2) MODEL

For example, moving average model MA(2) can be written as

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}.$$



MA(2) MODEL

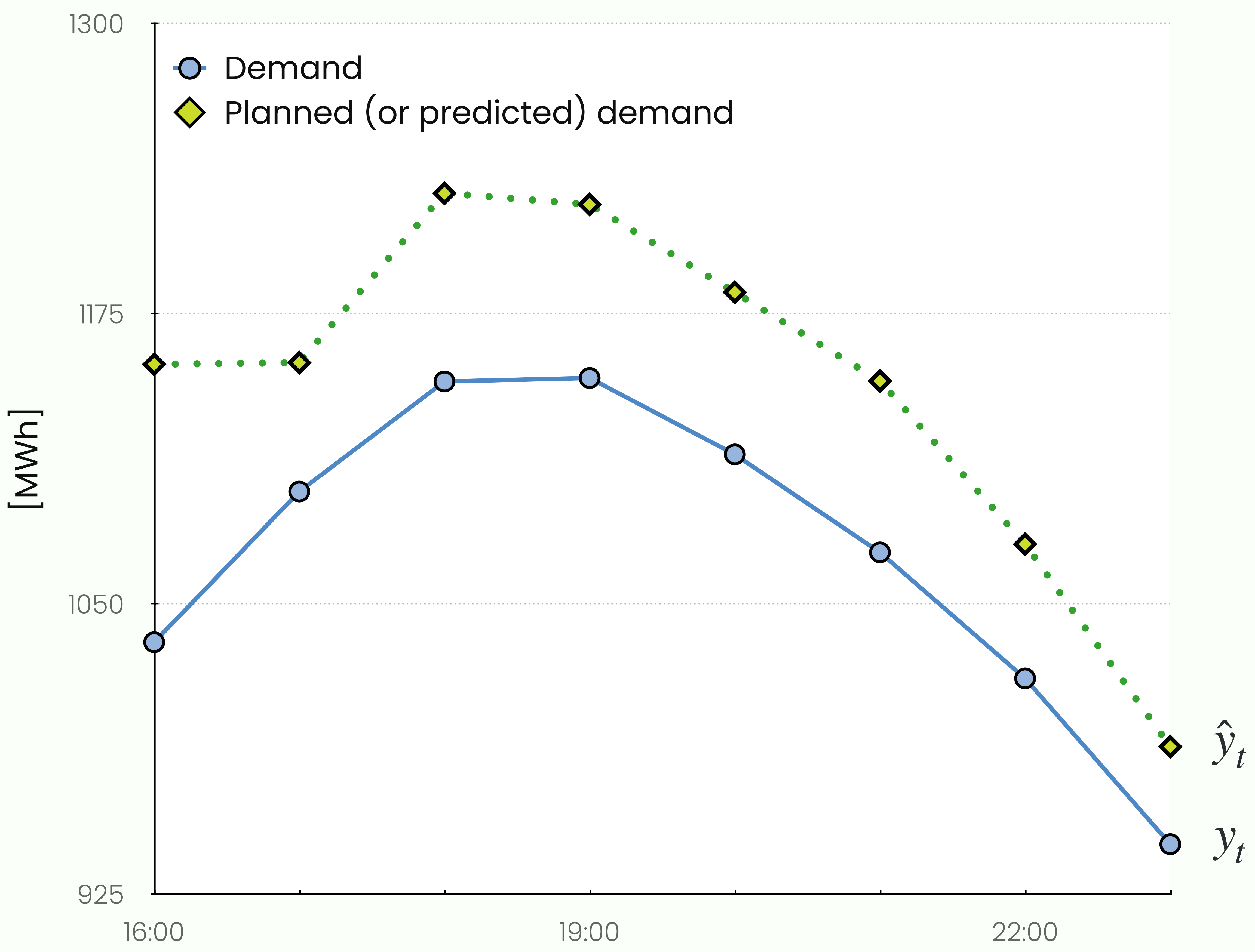
For example, moving average model MA(2) can be written as

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}.$$

MA helps to correct the error made in the previous time step.

Practical observation (we do not have access to the current error ϵ_t):

$$\hat{y}_t = \mu + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}.$$

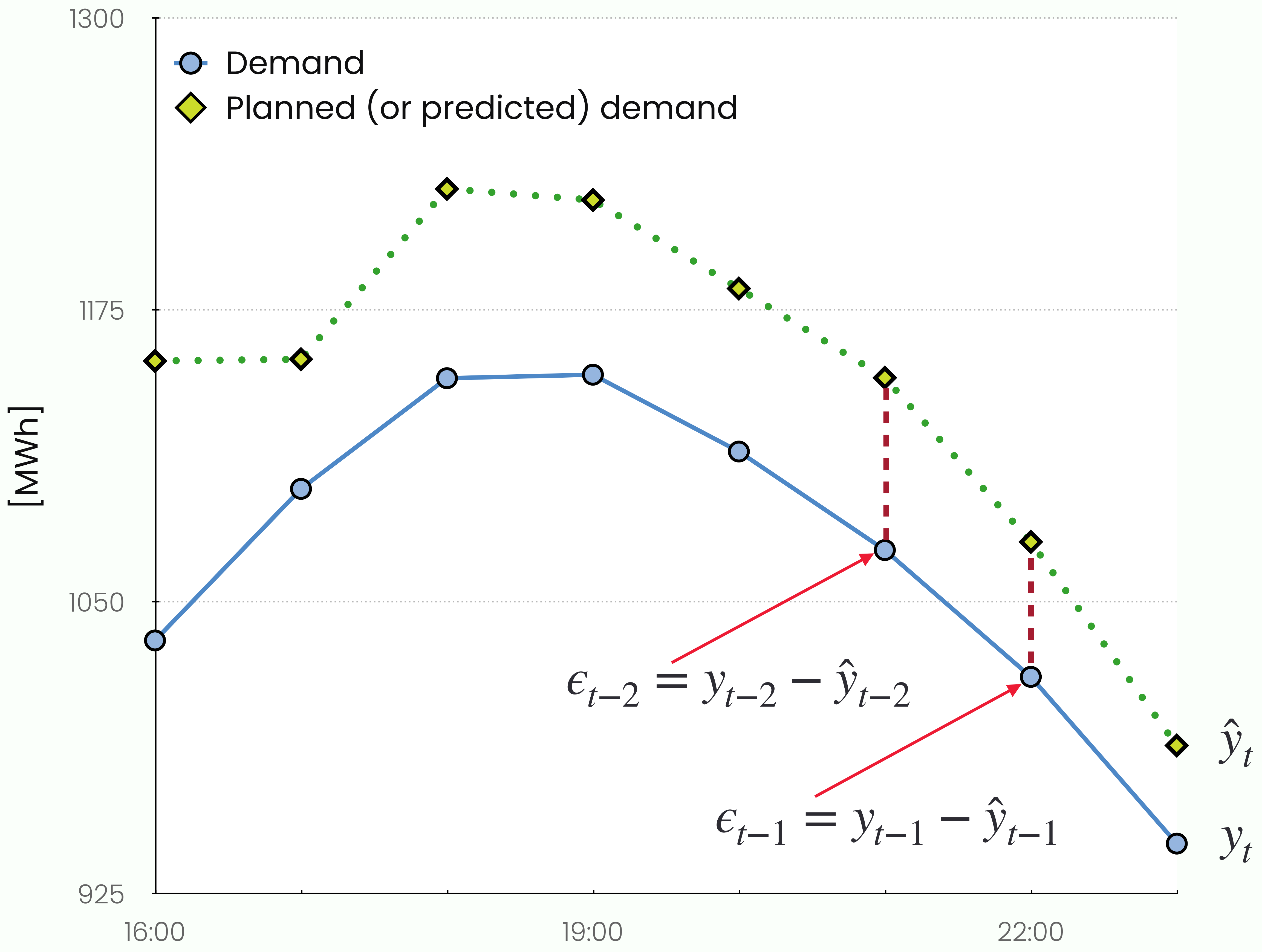


MA(2) MODEL

For example, moving average model MA(2) can be written as

$$y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}.$$

lagged errors



MA helps to correct the error made in the previous time step.

Practical observation (we do not have access to the current error ϵ_t):

$$\hat{y}_t = \mu + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}.$$

MOVING AVERAGE (MA) MODEL

Rather than using past values of the target variable in a regression, a moving average model uses past errors and can be written as

$$y_t = \mu + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q},$$

where $\epsilon_t \sim wn(0, \sigma_w^2)$ is white noise, $\theta_1, \ldots, \theta_q$ are parameters, μ is the mean of the series.

We refer to this as an MA(q) model, a moving average model of order q.

AUTOREGRESSIVE MOVING AVERAGE (ARMA)

ARMA(p, q)

AR(p)

AutoRegression

+

MA(q)

Moving Average

AR models a series based solely on the past values in the series (lags).

MA models a series based solely on the past errors in the series (error lags).

The family of methods is reach:

SARIMA, VARIMA, FARIMA, ARCH, VAR, etc.

ARMA: FULL MODEL

The full model can be written as

$$
\begin{aligned}
y_t &= c + \epsilon_t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \\
&= c + \epsilon_t + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \,.
\end{aligned}
$$

We call this an ARMA(p, q) model, where

- p is the number of lag observations,
- q is the order of the moving average part.

ARMA: EXAMPLE

For example, ARMA$(2,1)$ can be written as

$$\hat{y}_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \theta_1 \epsilon_{t-1}$$

or using backward shift (lag) notation as

$$\left(1 - \phi_1 B - \phi_2 B^2\right) \hat{y}_t = \left(1 + \theta_1 B\right) \epsilon_t.$$

ESTIMATING ARMA ORDERS

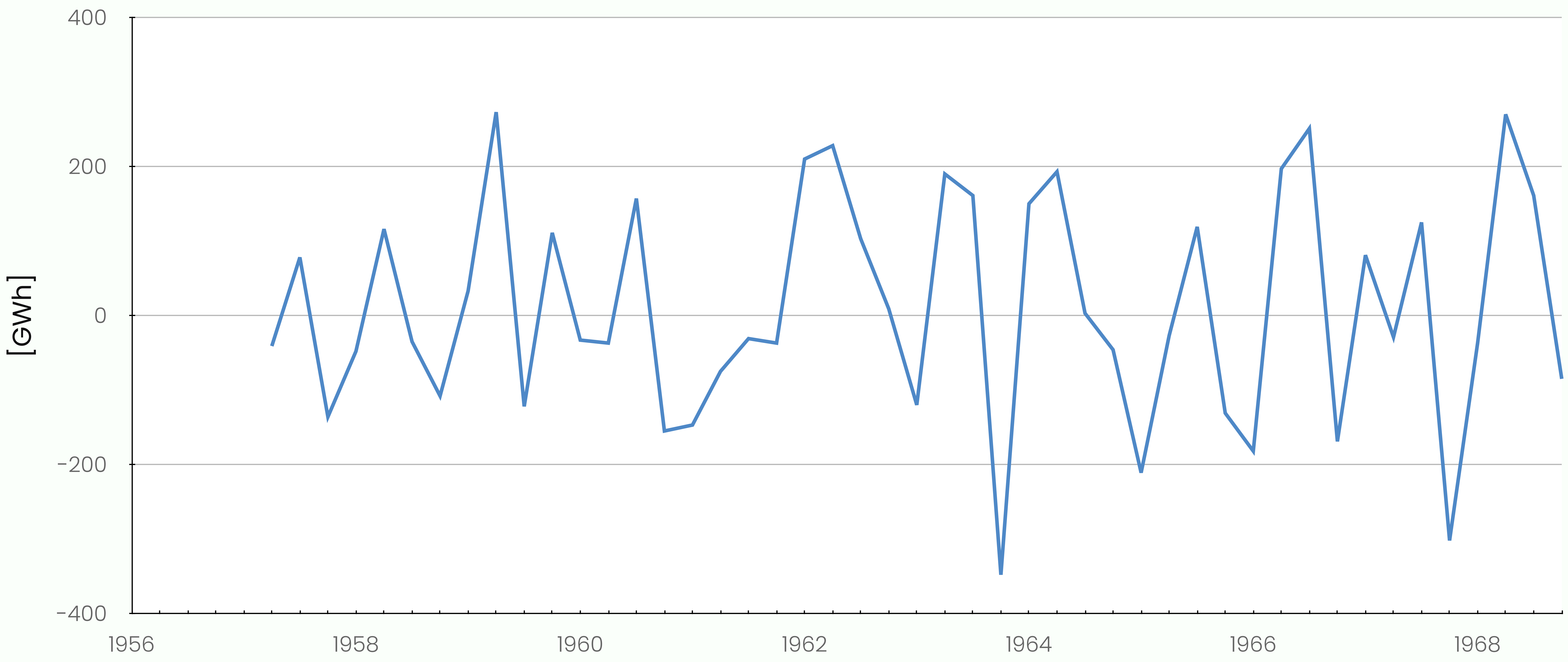
Note that the selection for p and q is not unique, and one can use:

- ACF and PACF plots
- information criteria (IC) (Akaike's IC or Bayesian IC)
- performance metrics (MSE or RMSE)

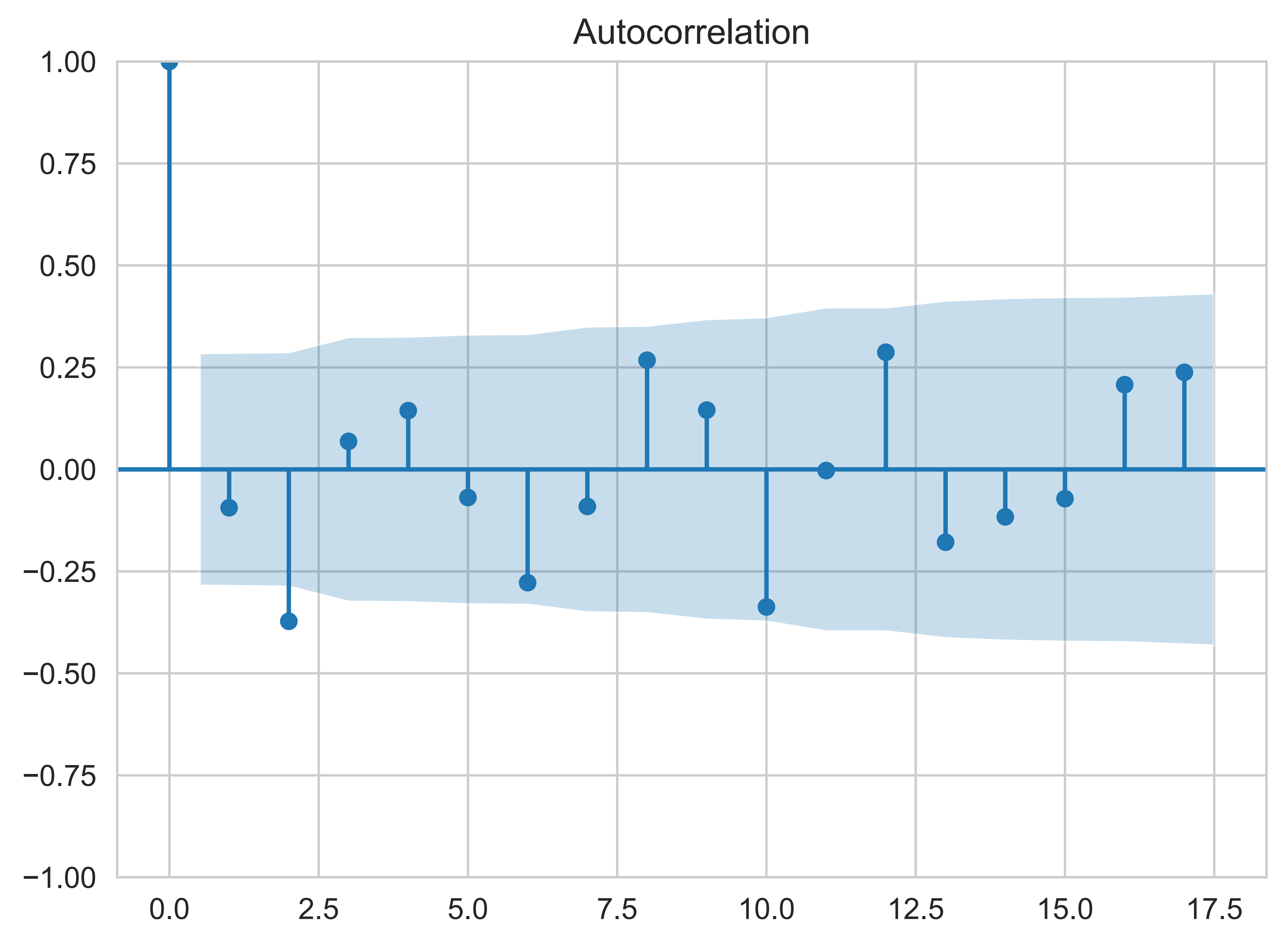
AR(p)/MA(q)	0	1	2	3	…
0					
1					
2					
3					
…					

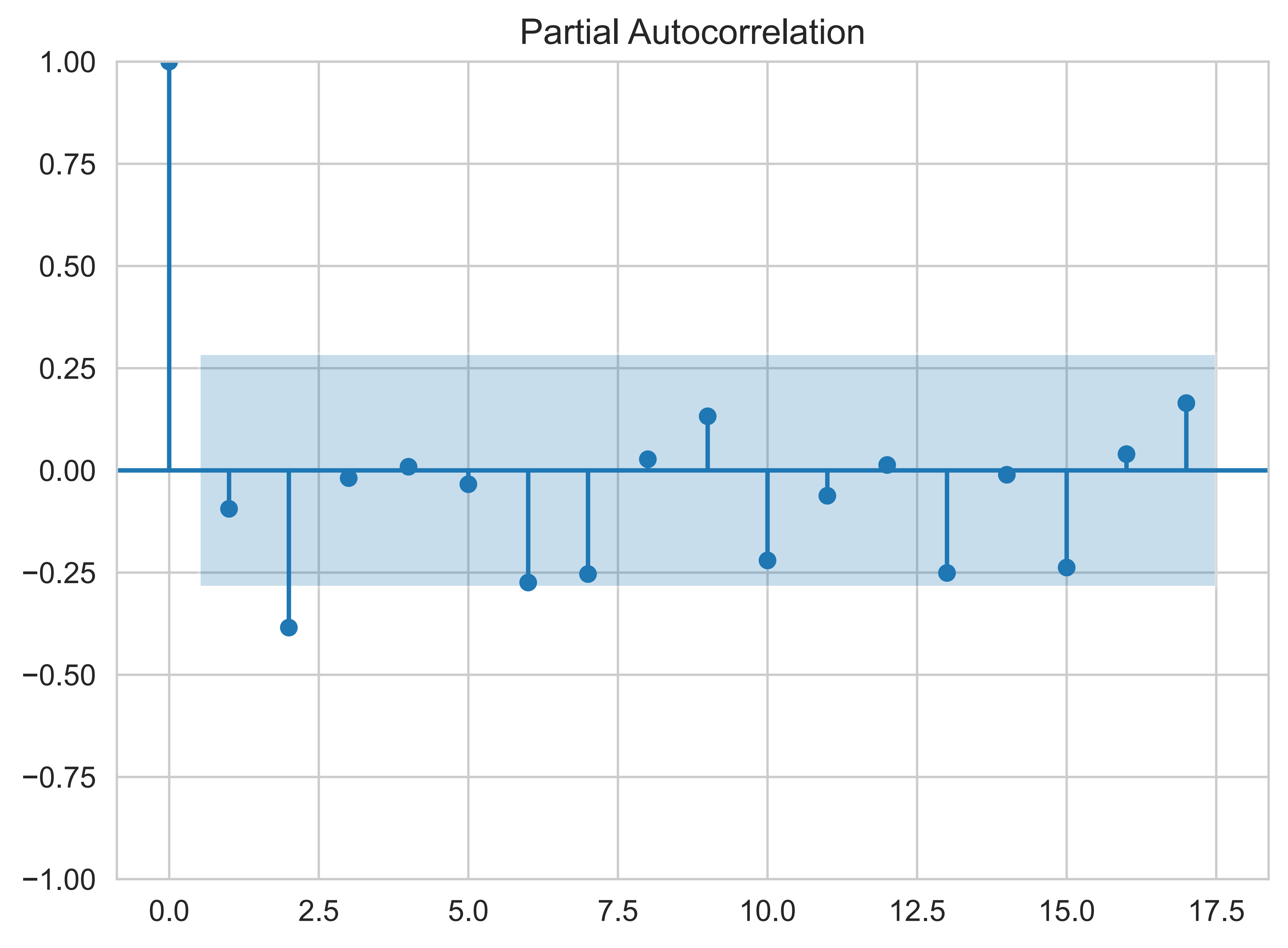
ILLUSTRATIVE EXAMPLE

Consider Australian electricity production data after seasonal and first difference transformation $(1-B)\left(1-B^4\right)y_t$.

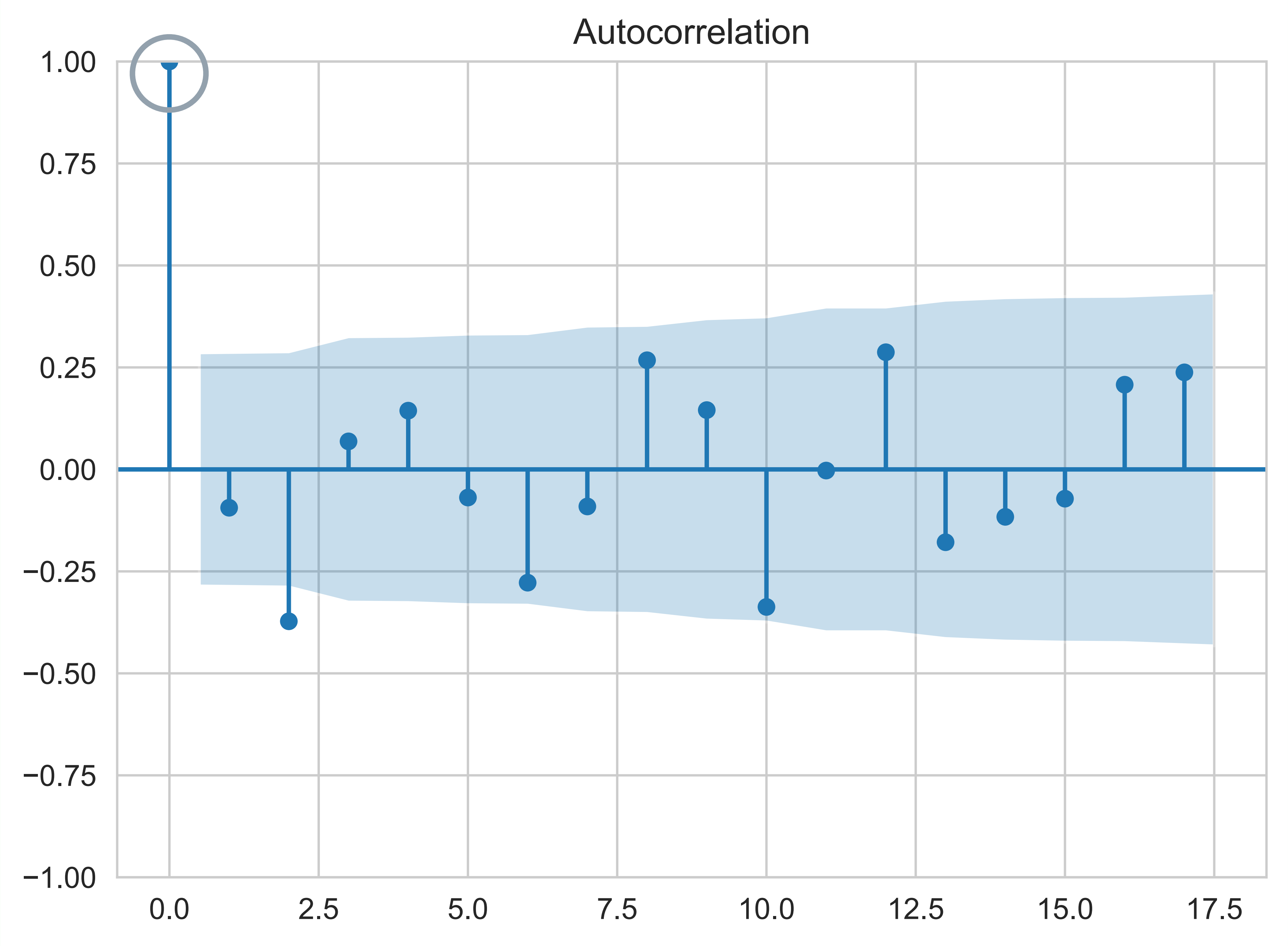


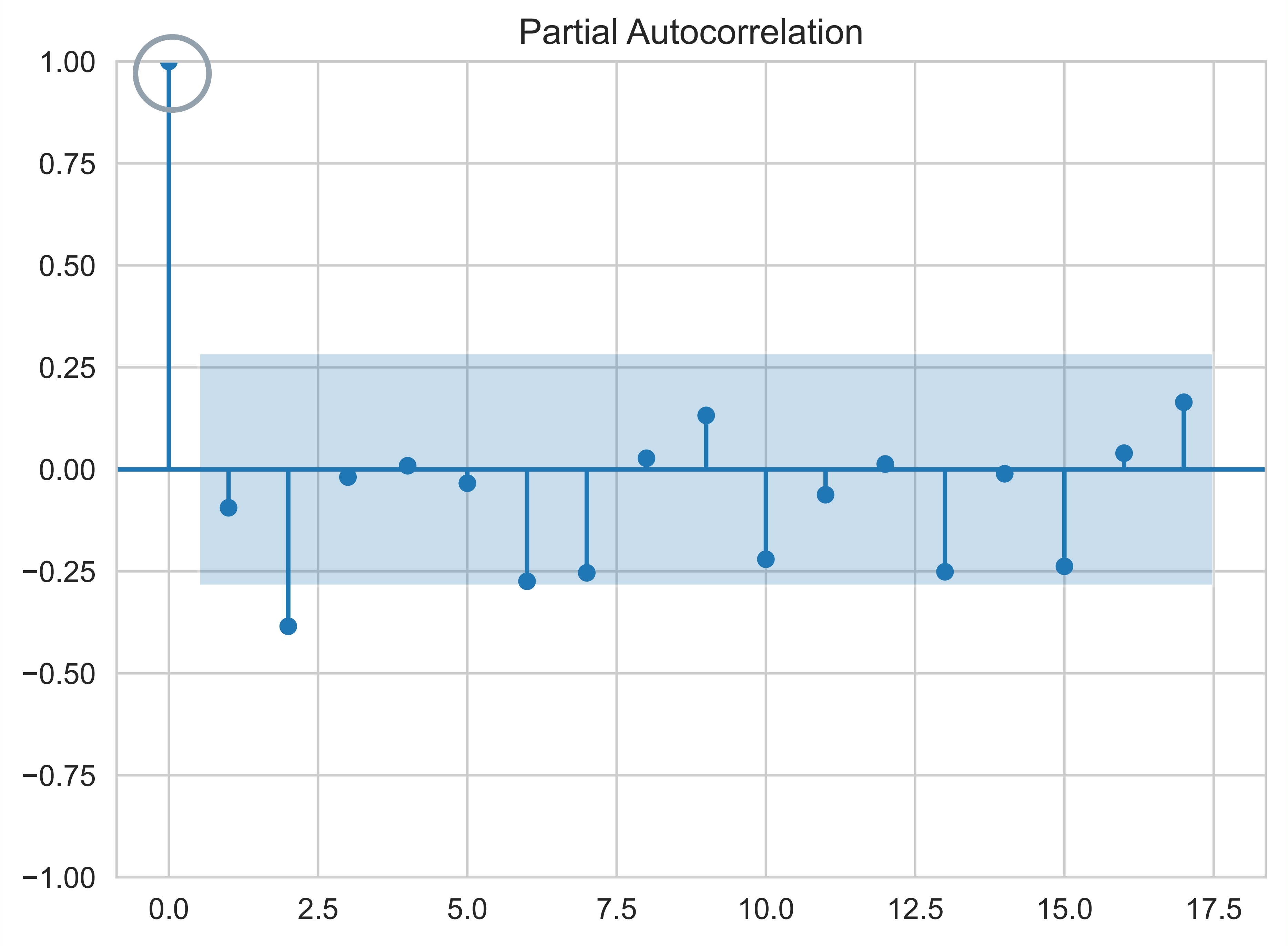
ACF & PACF


Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

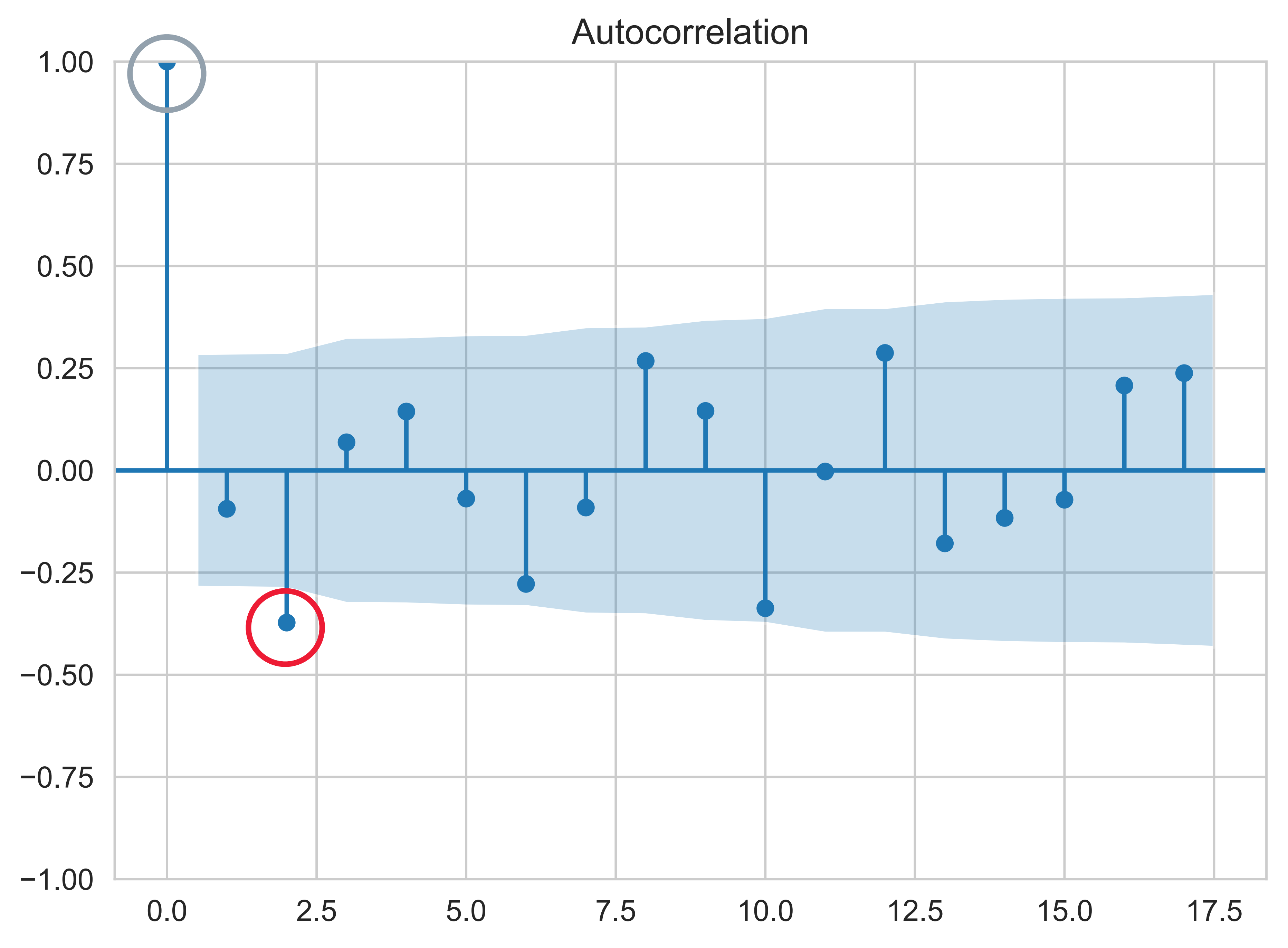

Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

ACF & PACF

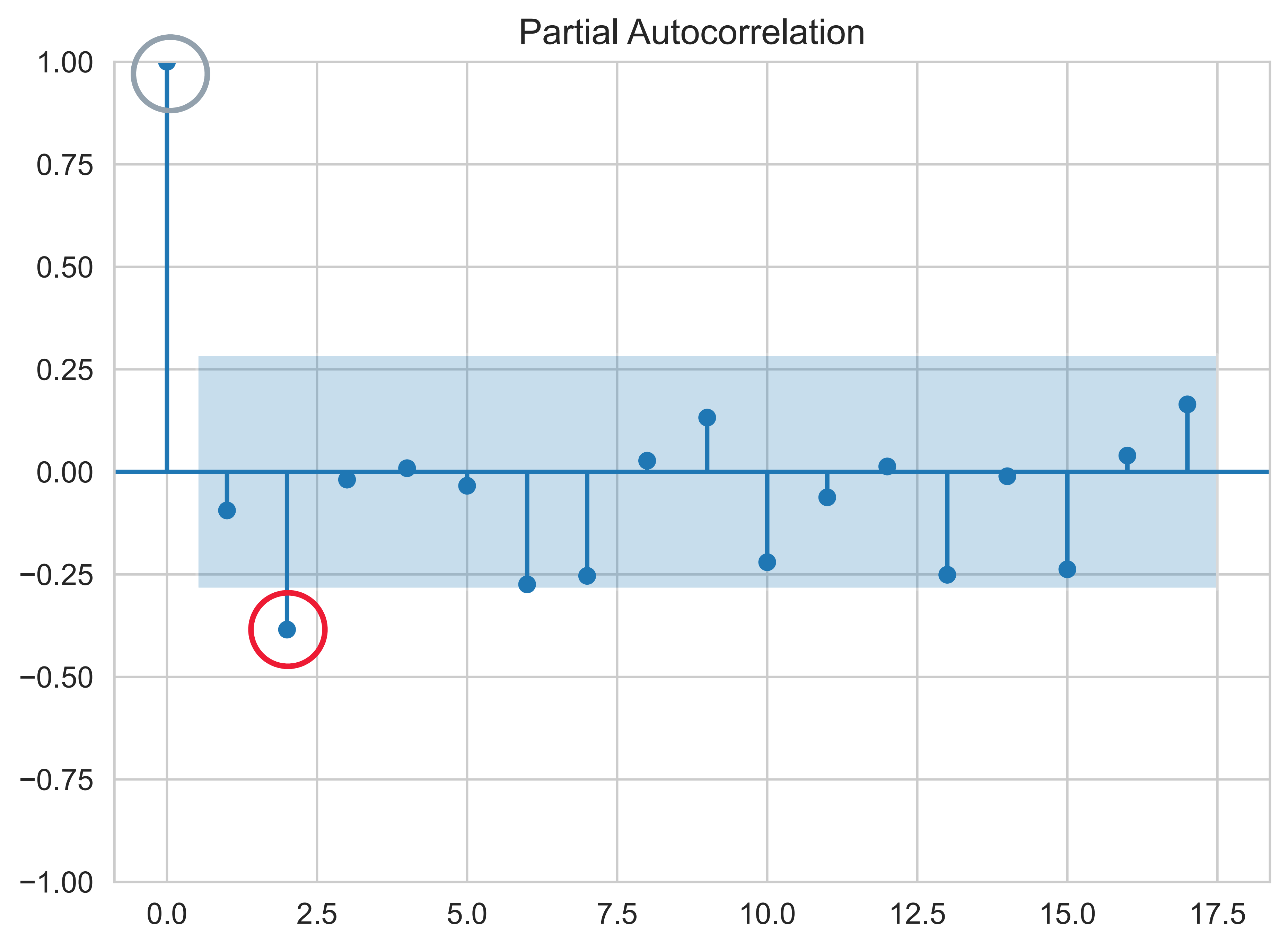

Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

ACF & PACF

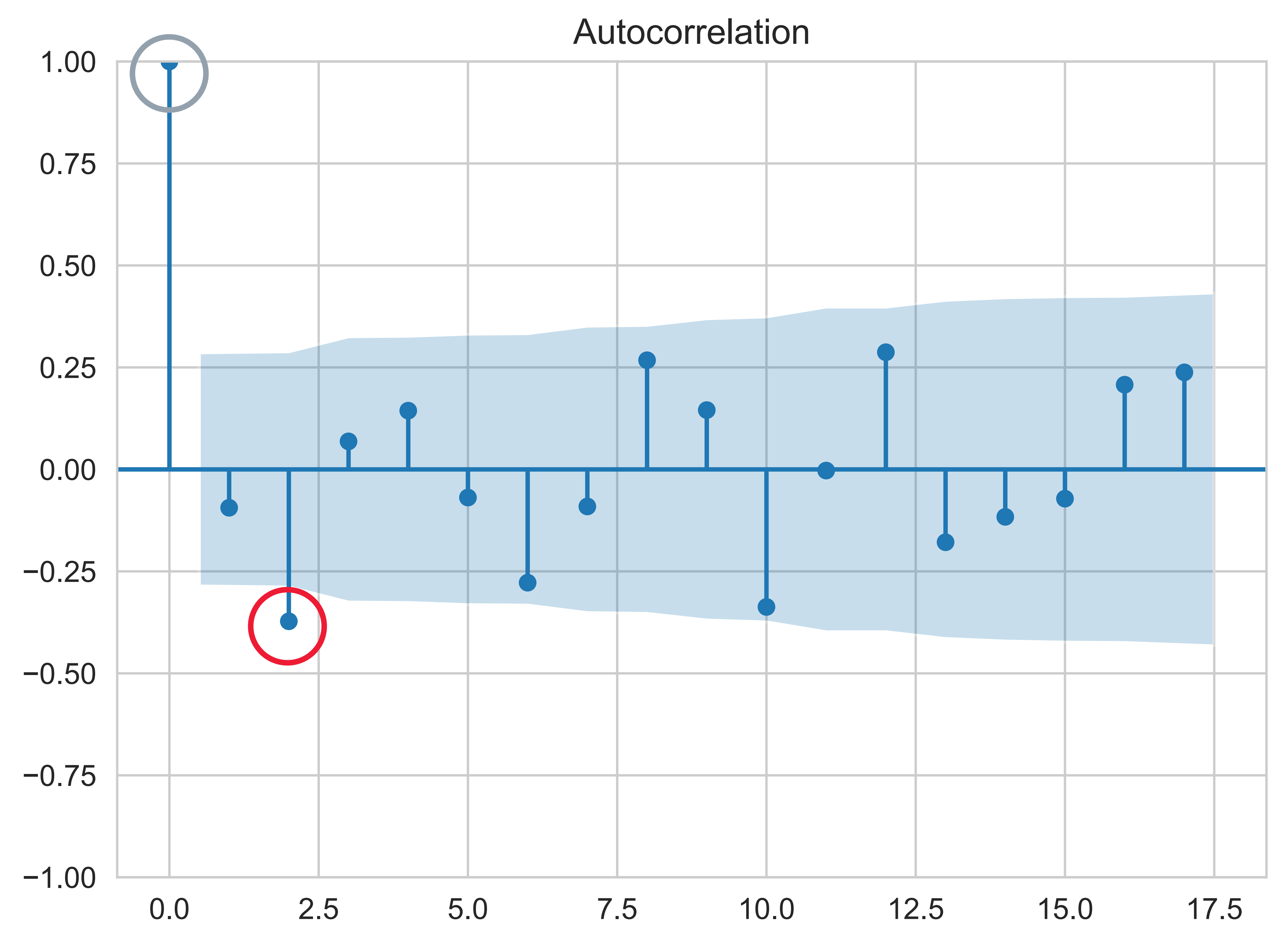


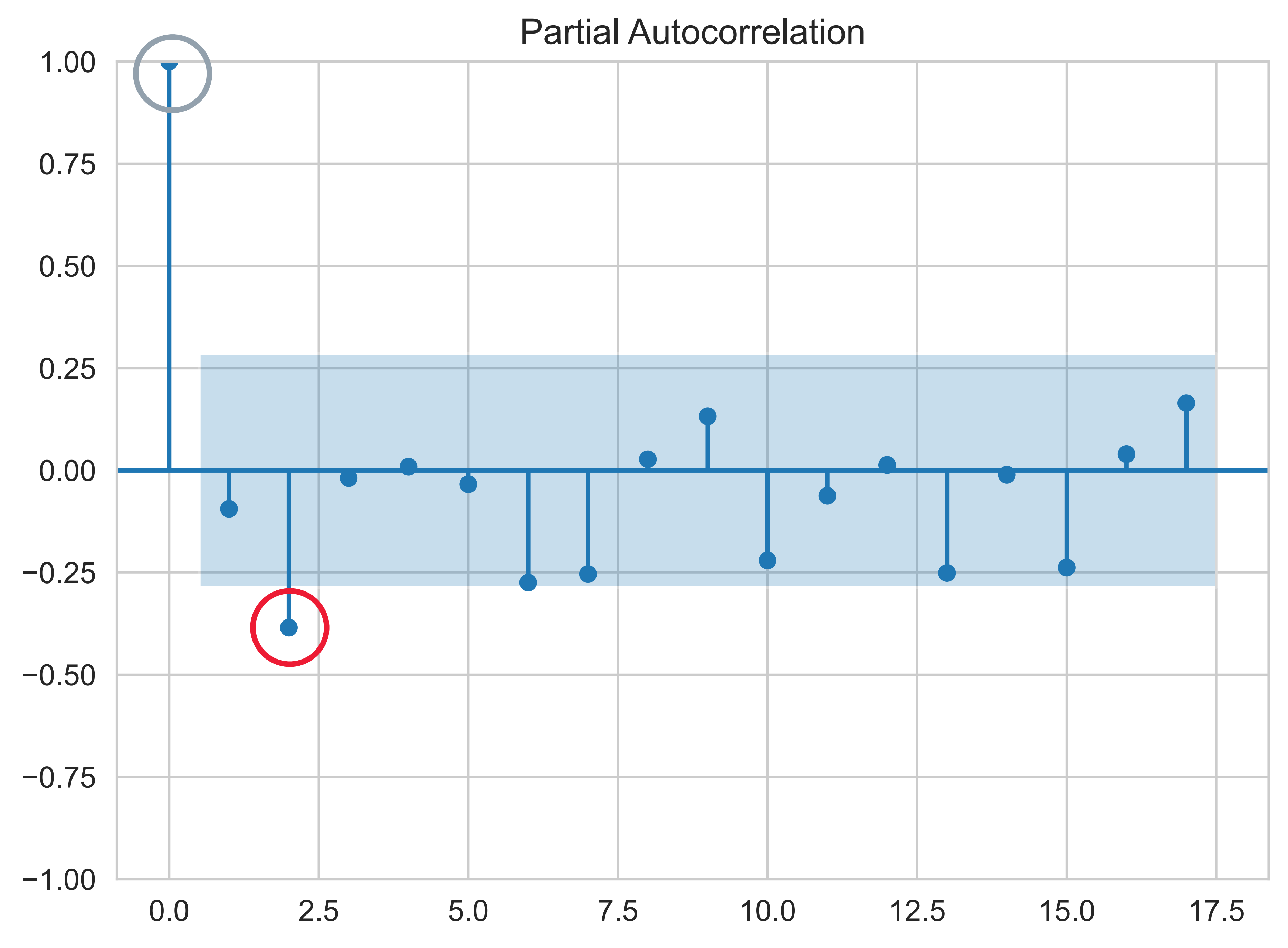
Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5



Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

ACF & PACF





Based on ACF q option is:

MA(2)

Based on PACF p option is:

AR(2)

ESTIMATION

```
                               SARIMAX Results
==============================================================================
Dep. Variable:            electricity   No. Observations:                   44
Model:                 ARIMA(2, 0, 2)   Log Likelihood                -277.270
Date:                Tue, 22 Jul 2025   AIC                            566.539
Time:                        15:12:58   BIC                            577.244
Sample:                    06-01-1957   HQIC                           570.509
                         - 03-01-1968
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.6923     18.744      0.250      0.802     -32.044      41.429
ar.L1          0.0975      0.217      0.449      0.653      -0.328       0.523
ar.L2         -0.9488      0.144     -6.567      0.000      -1.232      -0.666
ma.L1         -0.2016      0.333     -0.606      0.544      -0.853       0.450
ma.L2          0.8692      0.273      3.189      0.001       0.335       1.404
sigma2      1.791e+04   6557.285      2.732      0.006    5060.488    3.08e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.23   Jarque-Bera (JB):                 1.23
Prob(Q):                              0.63   Prob(JB):                         0.54
Heteroskedasticity (H):               1.39   Skew:                            -0.04
Prob(H) (two-sided):                  0.53   Kurtosis:                         2.18
===================================================================================
```

ESTIMATION

```
                               SARIMAX Results
==============================================================================
Dep. Variable:            electricity   No. Observations:                   44
Model:                 ARIMA(2, 0, 2)   Log Likelihood                -277.270
Date:                Tue, 22 Jul 2025   AIC                            566.539
Time:                        15:12:58   BIC                            577.244
Sample:                    06-01-1957   HQIC                           570.509
                         - 03-01-1968
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.6923     18.744      0.250      0.802     -32.044      41.429
ar.L1          0.0975      0.217      0.449      0.653      -0.328       0.523
ar.L2         -0.9488      0.144     -6.567      0.000      -1.232      -0.666
ma.L1         -0.2016      0.333     -0.606      0.544      -0.853       0.450
ma.L2          0.8692      0.273      3.189      0.001       0.335       1.404
sigma2      1.791e+04   6557.285      2.732      0.006    5060.488    3.08e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.23   Jarque-Bera (JB):                 1.23
Prob(Q):                              0.63   Prob(JB):                         0.54
Heteroskedasticity (H):               1.39   Skew:                            -0.04
Prob(H) (two-sided):                  0.53   Kurtosis:                         2.18
===================================================================================
```

ESTIMATION

```
                               SARIMAX Results
==============================================================================
Dep. Variable:            electricity   No. Observations:                   44
Model:                 ARIMA(2, 0, 2)   Log Likelihood                -277.270
Date:                Tue, 22 Jul 2025   AIC                            566.539
Time:                        15:12:58   BIC                            577.244
Sample:                    06-01-1957   HQIC                           570.509
                         - 03-01-1968
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.6923     18.744      0.250      0.802     -32.044      41.429
ar.L1          0.0975      0.217      0.449      0.653      -0.328       0.523
ar.L2         -0.9488      0.144     -6.567      0.000      -1.232      -0.666
ma.L1         -0.2016      0.333     -0.606      0.544      -0.853       0.450
ma.L2          0.8692      0.273      3.189      0.001       0.335       1.404
sigma2      1.791e+04   6557.285      2.732      0.006    5060.488    3.08e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.23   Jarque-Bera (JB):                 1.23
Prob(Q):                              0.63   Prob(JB):                         0.54
Heteroskedasticity (H):               1.39   Skew:                            -0.04
Prob(H) (two-sided):                  0.53   Kurtosis:                         2.18
===================================================================================
```

ESTIMATION

SARIMAX Results

Dep. Variable:	electricity	No. Observations:	44
Model:	ARIMA(2, 0, 2)	Log Likelihood	-277.270
Date:	Tue, 22 Jul 2025	AIC	566.539
Time:	15:12:58	BIC	577.244
Sample:	06-01-1957	HQIC	570.509
	- 03-01-1968		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
~~const~~	~~4.6923~~	18.744	0.250	0.802	-32.044	41.429
~~ar.L1~~	~~0.0975~~	0.217	0.449	0.653	-0.328	0.523
ar.L2	-0.9488	0.144	-6.567	0.000	-1.232	-0.666
~~ma.L1~~	~~0.2016~~	0.333	-0.606	0.544	-0.853	0.450
ma.L2	0.8692	0.273	3.189	0.001	0.335	1.404
sigma2	1.791e+04	6557.285	2.732	0.006	5060.488	3.08e+04

Ljung-Box (L1) (Q):	0.23	Jarque-Bera (JB):	1.23
Prob(Q):	0.63	Prob(JB):	0.54
Heteroskedasticity (H):	1.39	Skew:	-0.04
Prob(H) (two-sided):	0.53	Kurtosis:	2.18

ESTIMATION

SARIMAX Results

Dep. Variable:	electricity	No. Observations:	44
Model:	ARIMA(2, 0, 2)	Log Likelihood	-277.270
Date:	Tue, 22 Jul 2025	AIC	566.539
Time:	15:12:58	BIC	577.244
Sample:	06-01-1957	HQIC	570.509
	- 03-01-1968		
Covariance Type:	opg		

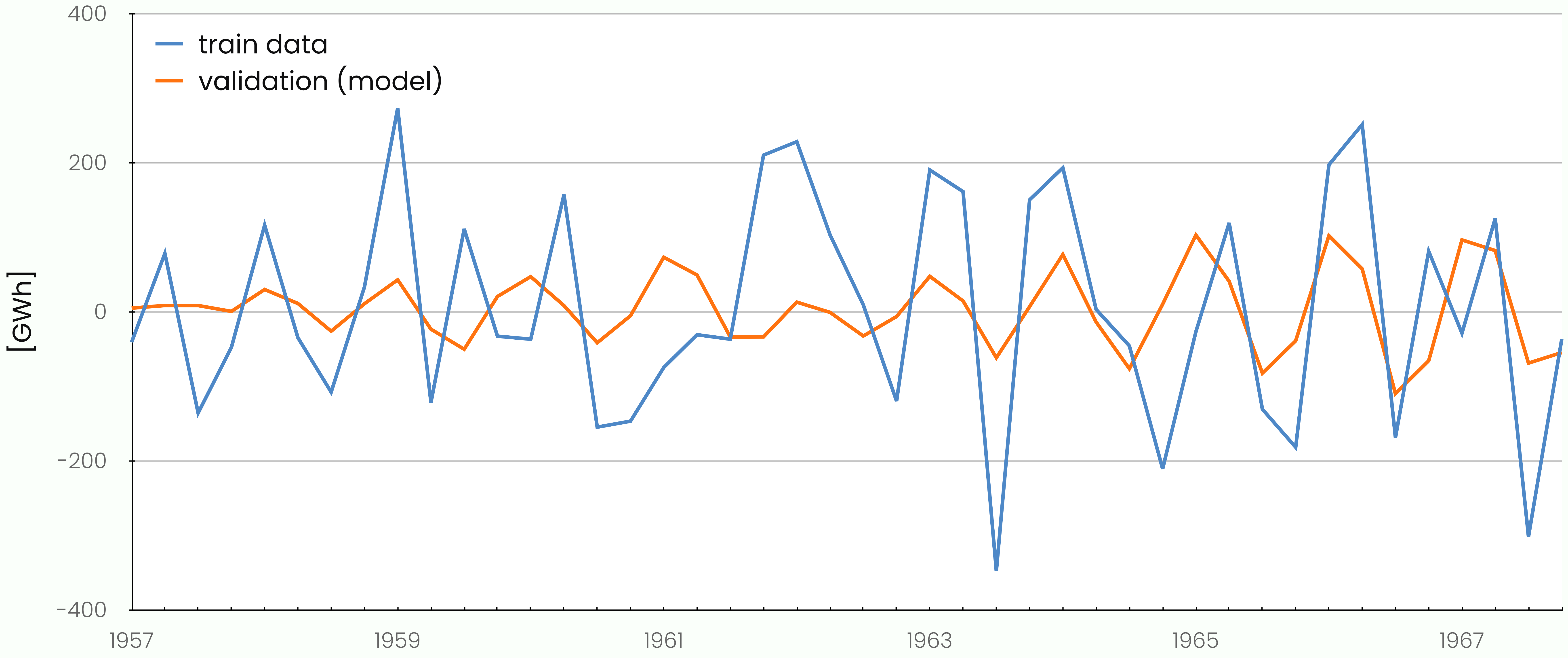
	coef	std err	z	P>|z|	[0.025	0.975]
~~const~~	~~4.6923~~	18.744	0.250	0.802	-32.044	41.429
~~ar.L1~~	~~0.0975~~	0.217	0.449	0.653	-0.328	0.523
ar.L2	-0.9488	0.144	-6.567	0.000	-1.232	-0.666
~~ma.L1~~	~~0.2016~~	0.333	-0.606	0.544	-0.853	0.450
ma.L2	0.8692	0.273	3.189	0.001	0.335	1.404
sigma2	1.791e+04	6557.285	2.732	0.006	5060.488	3.08e+04

Ljung-Box (L1) (Q):	0.23	Jarque-Bera (JB):	1.23
Prob(Q):	0.63	Prob(JB):	0.54
Heteroskedasticity (H):	1.39	Skew:	-0.04
Prob(H) (two-sided):	0.53	Kurtosis:	2.18

The final ARMA$(2,2)$ model (with B^1 lags and intercept excluded) reads as:

$$\hat{y}_t = -0.9488 y_{t-2} + 0.8692 \epsilon_{t-2}.$$

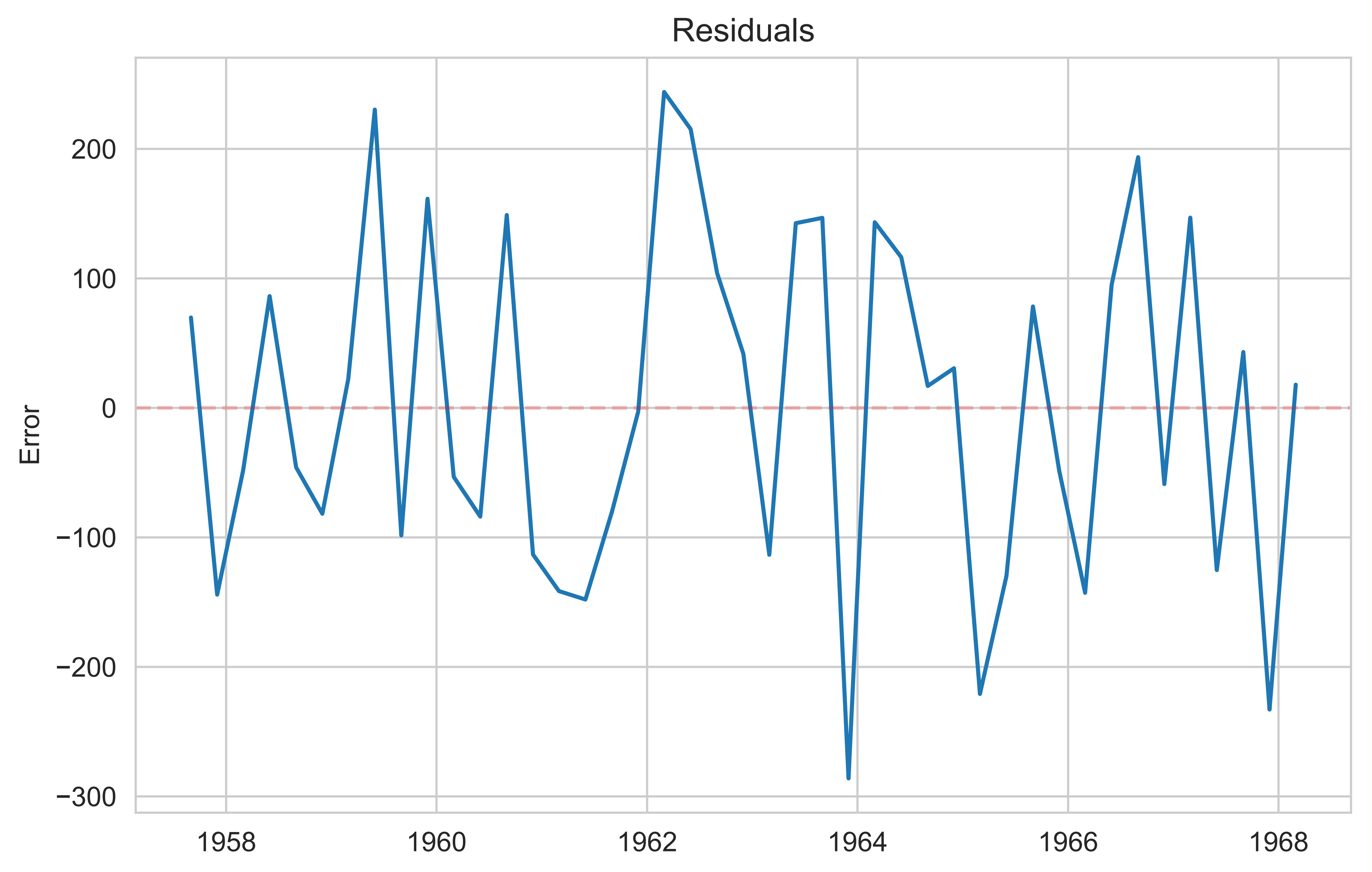
VALIDATION: TRAIN DATA

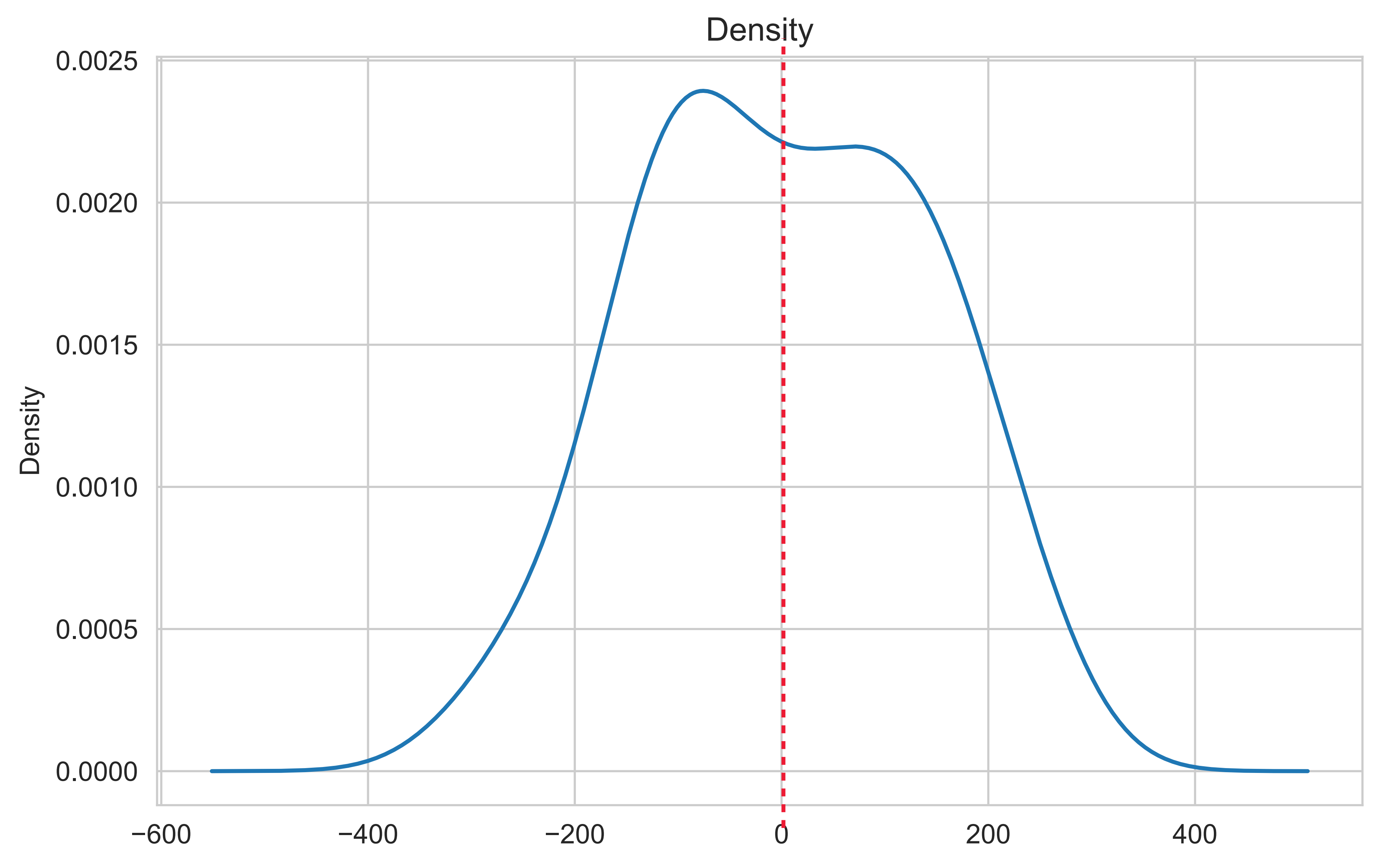


RMSE = 131.19

VALIDATION: RESIDUALS

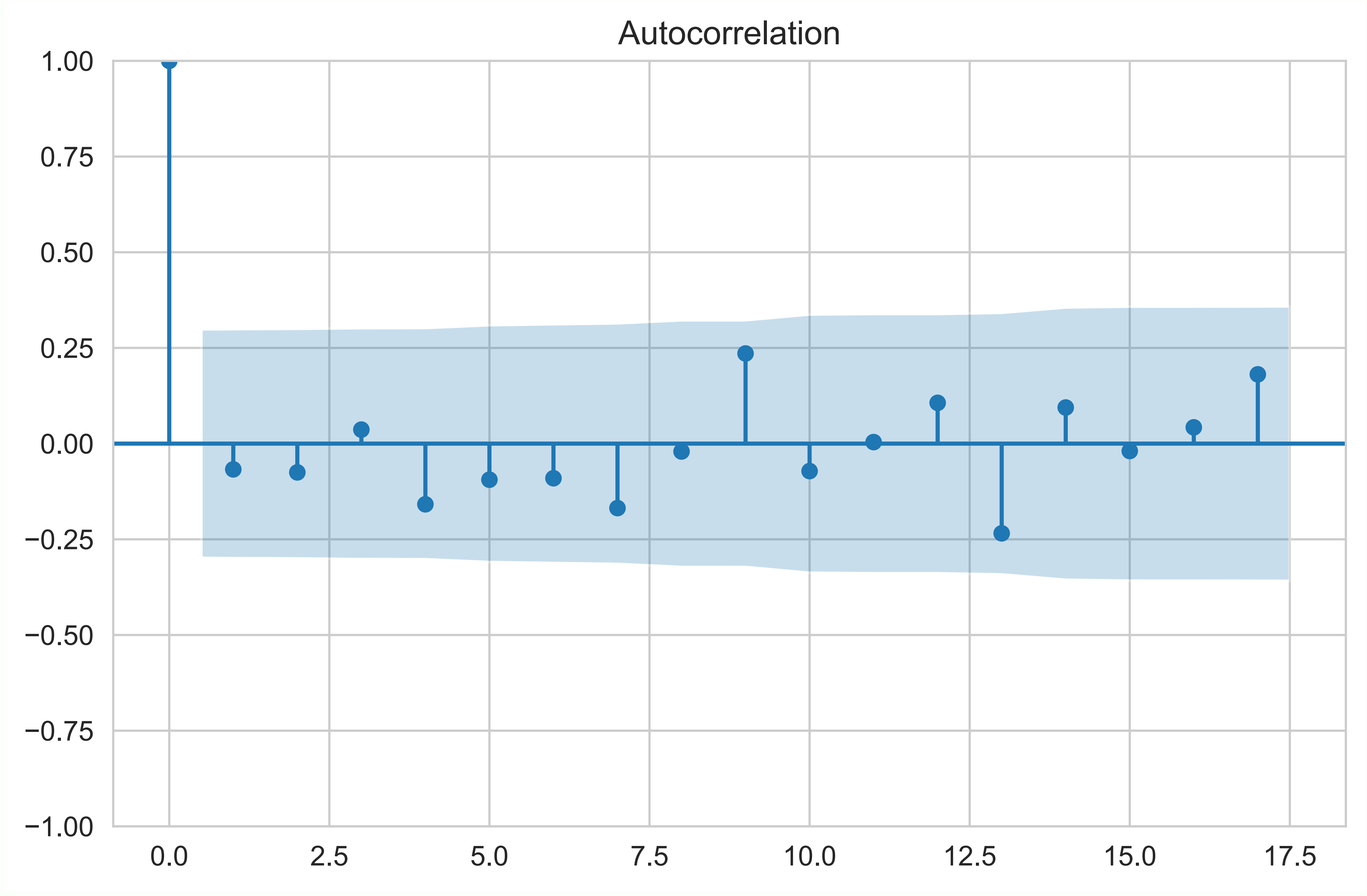
The residuals look random in general, and their density looks kind of normally distributed with a mean close to 0.

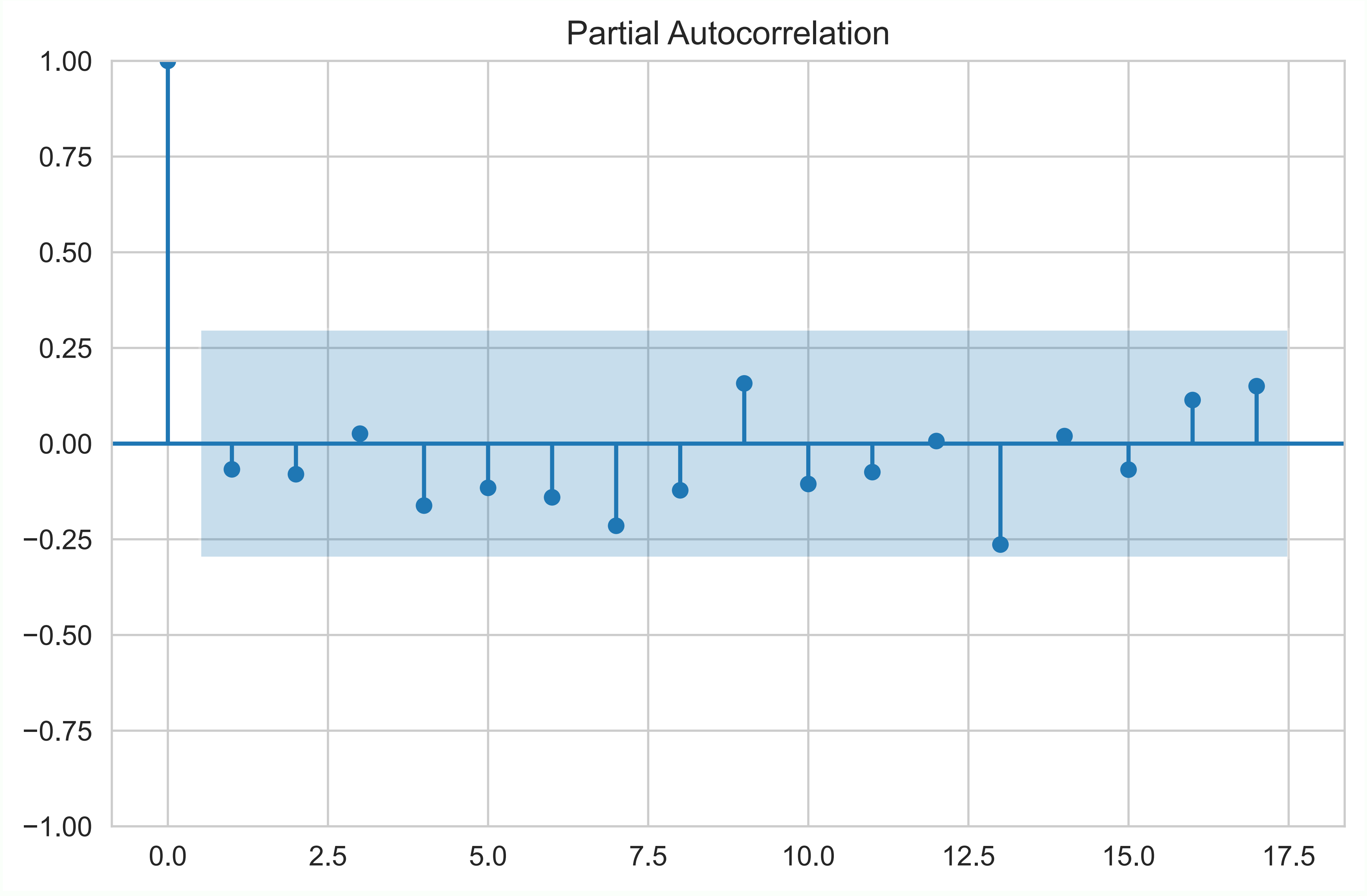




VALIDATION: RESIDUALS (2)

ACF and PACF of the residuals





These show that the residuals are close to white noise, and the model can be further utilised.

ARMA EXTENDED

$$\text{ARMA}(p, q)$$

ARMA EXTENDED

$$\text{ARMA}(p, q)$$

$$\big\downarrow \quad d \text{ is the degree of first differencing}$$

$$\text{ARIMA}(p, d, q)$$

SARIMA can handle only **one** seasonal component. If data contain more than one, use more advanced models.

ARMA EXTENDED

$$\text{ARMA}(p, q)$$

d is the degree of first differencing

$$\text{ARIMA}(p, d, q)$$

Seasonal: AR(P), first differencing (D), and MA(Q) parts

m is the number of observations per year

$$\text{SARIMA}(p, d, q)(P, D, Q)_m$$

SARIMA can handle only **one** seasonal component. If data contain more than one, use more advanced models.

Evaluating Time Series models

TIME SERIES VS CROSS-SECTIONAL

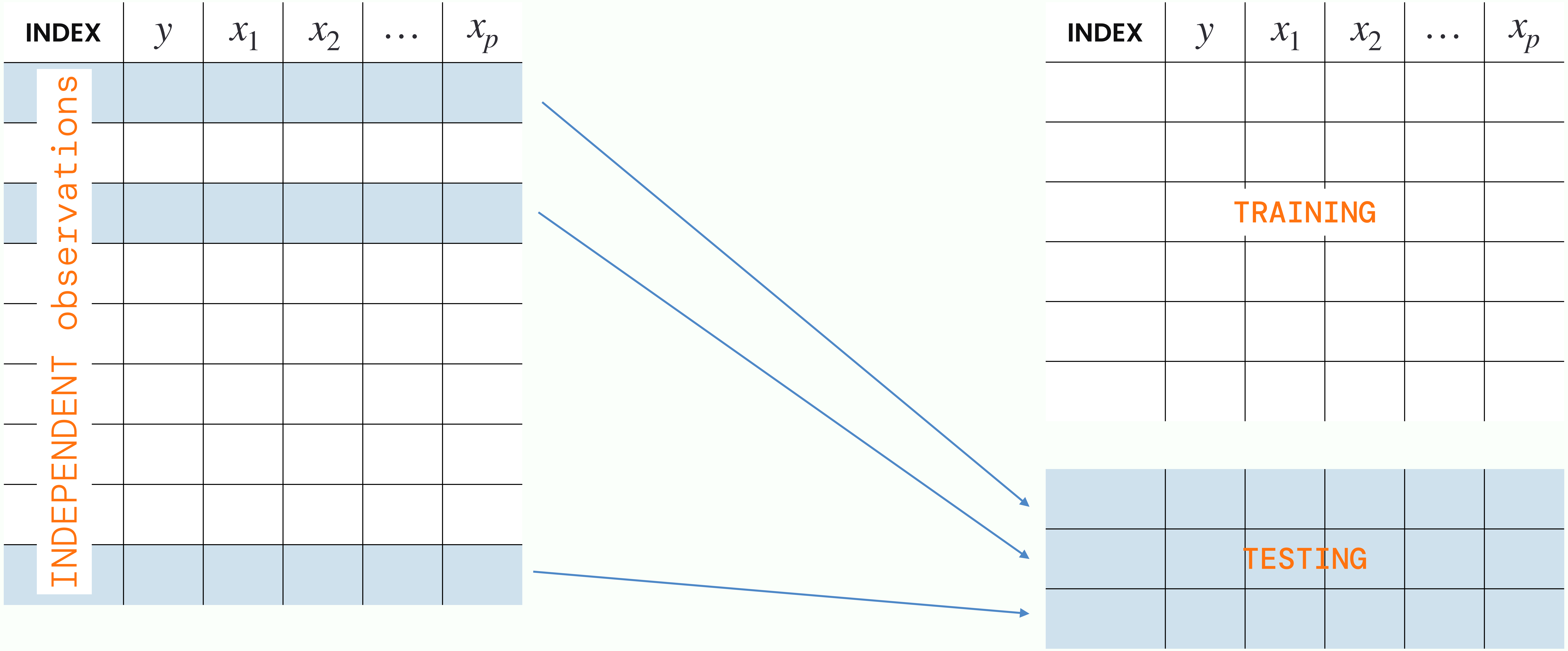
Time Series

Set of ordered data values *consistently* collected over an interval of time.

Cross-sectional

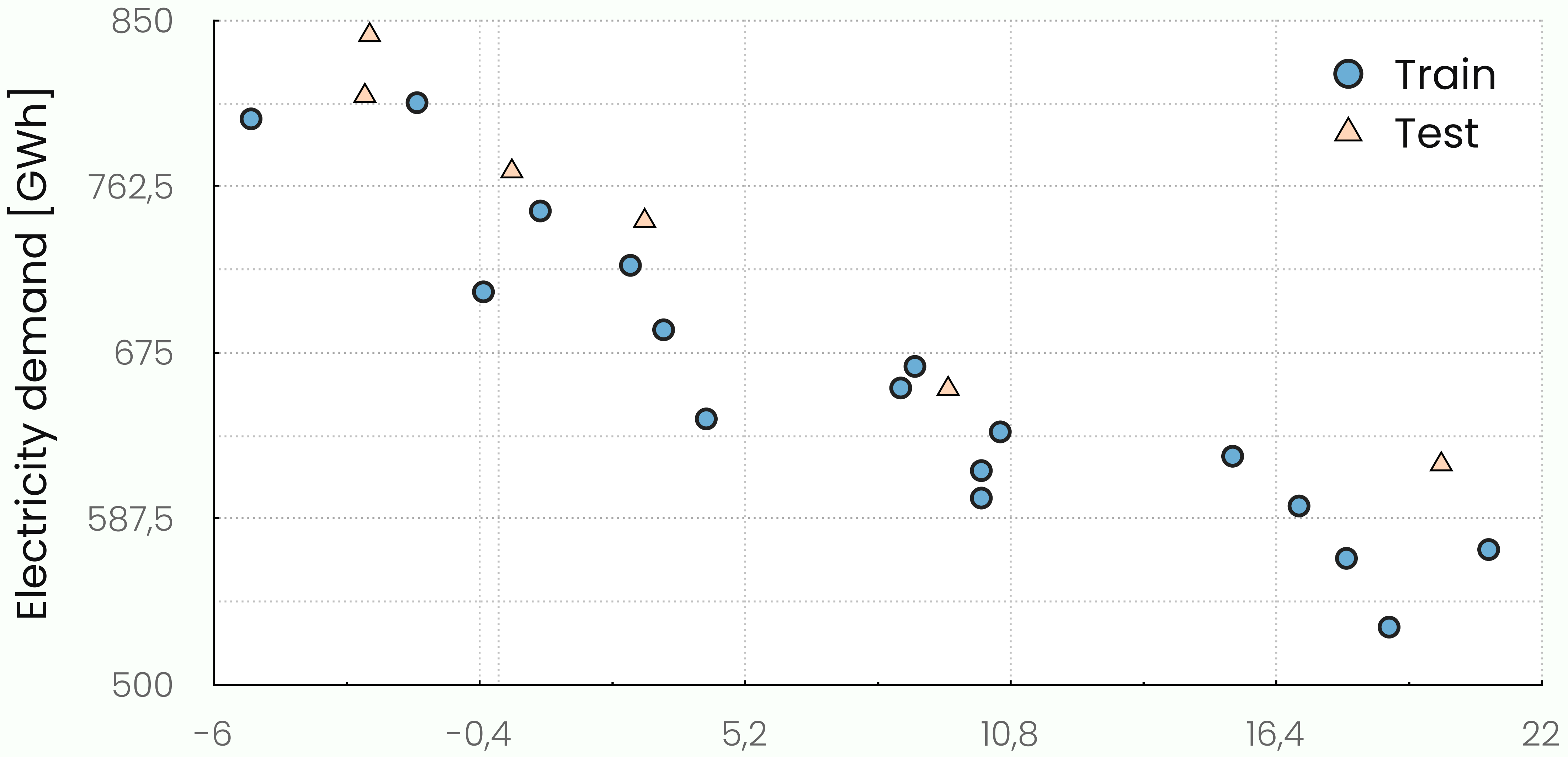
Set of data values collected at a *fixed* point in time, or where time is of *no significance*.

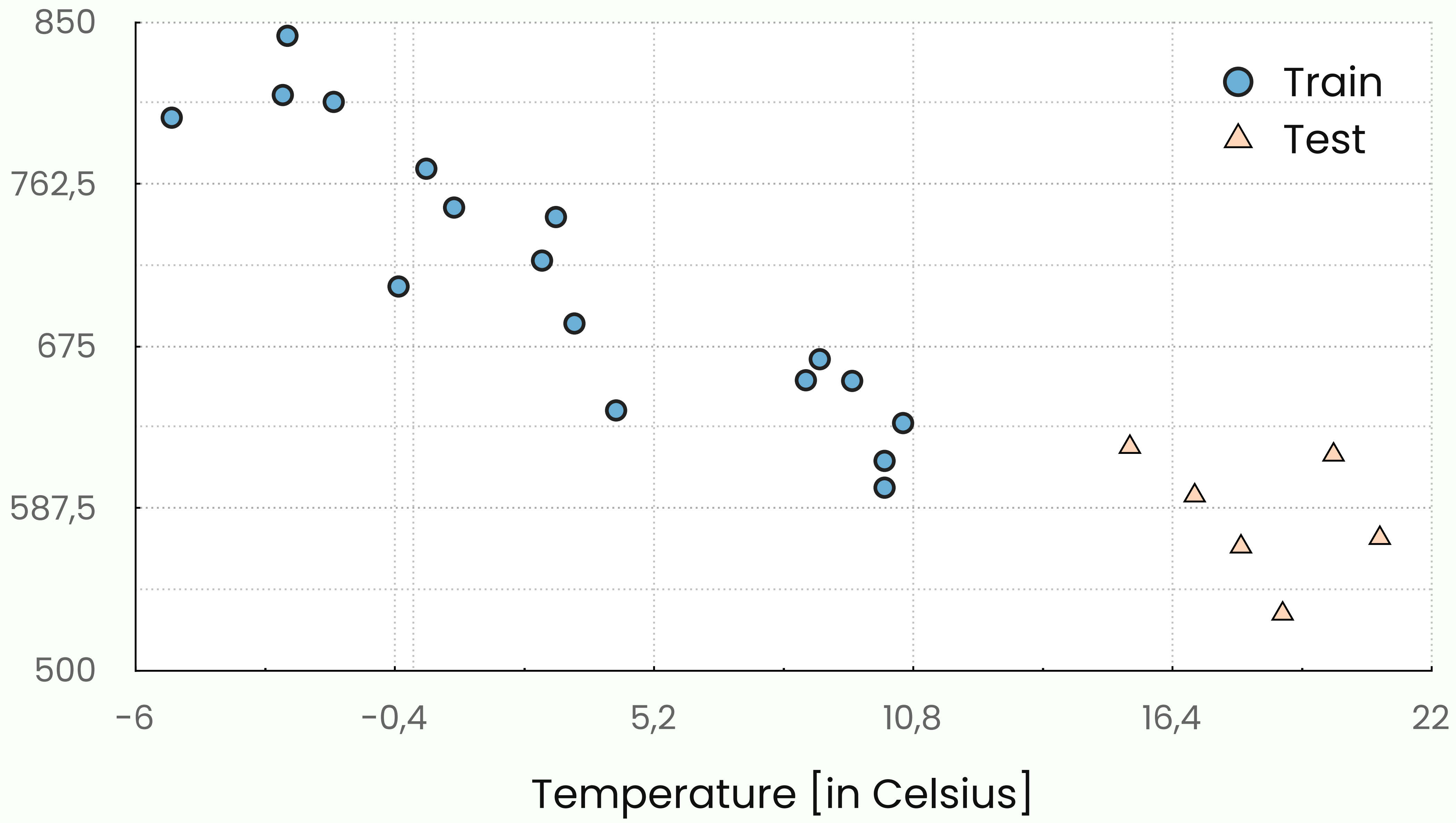
CS: RANDOM SPLIT FOR TRAINING/TESTING



(UN)LUCKY SPLIT?

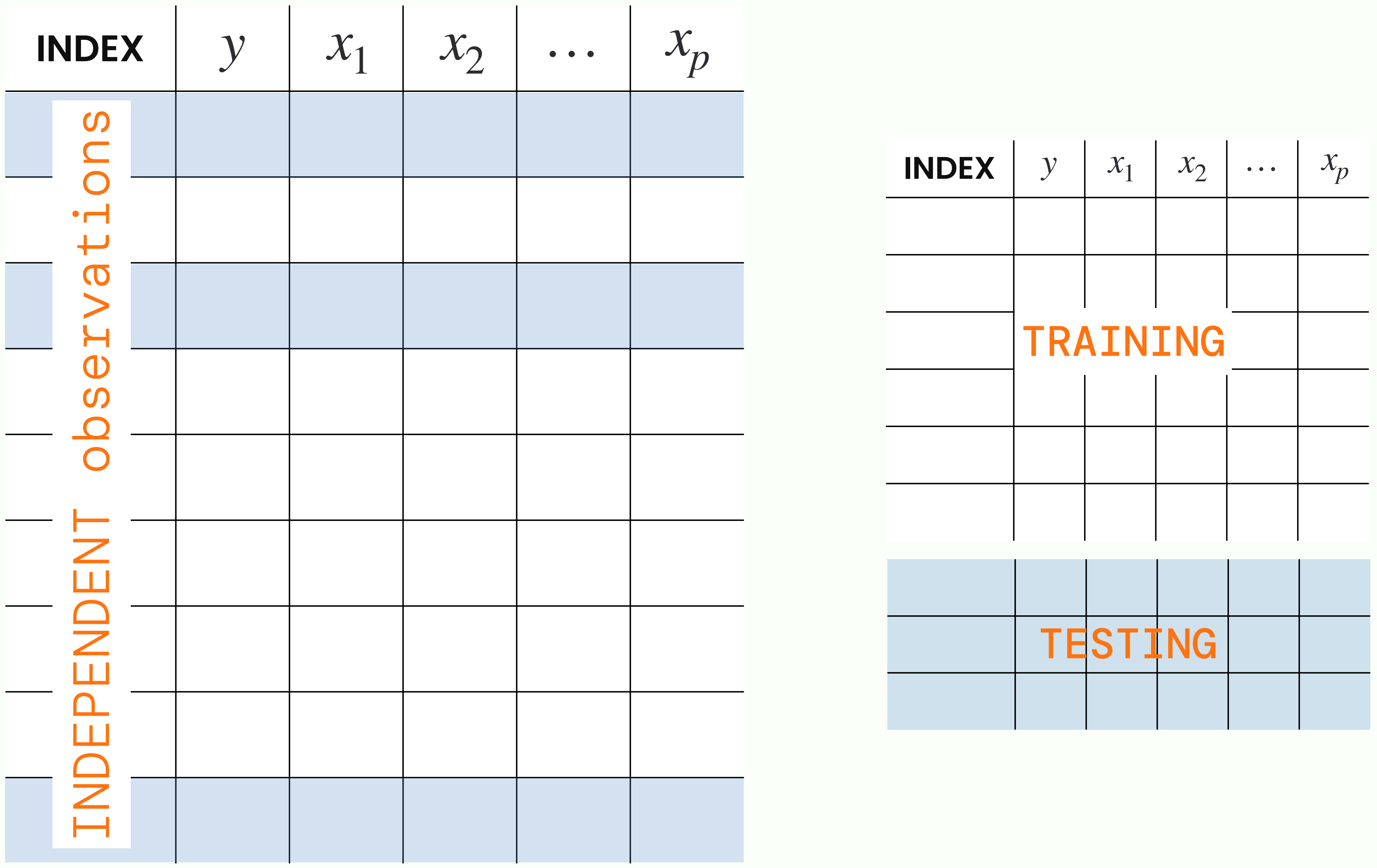
y	x_1
-2.7	842
…	…
0.3	771
…	…
3.1	745
…	…
9.5	656
…	…
-2.8	810



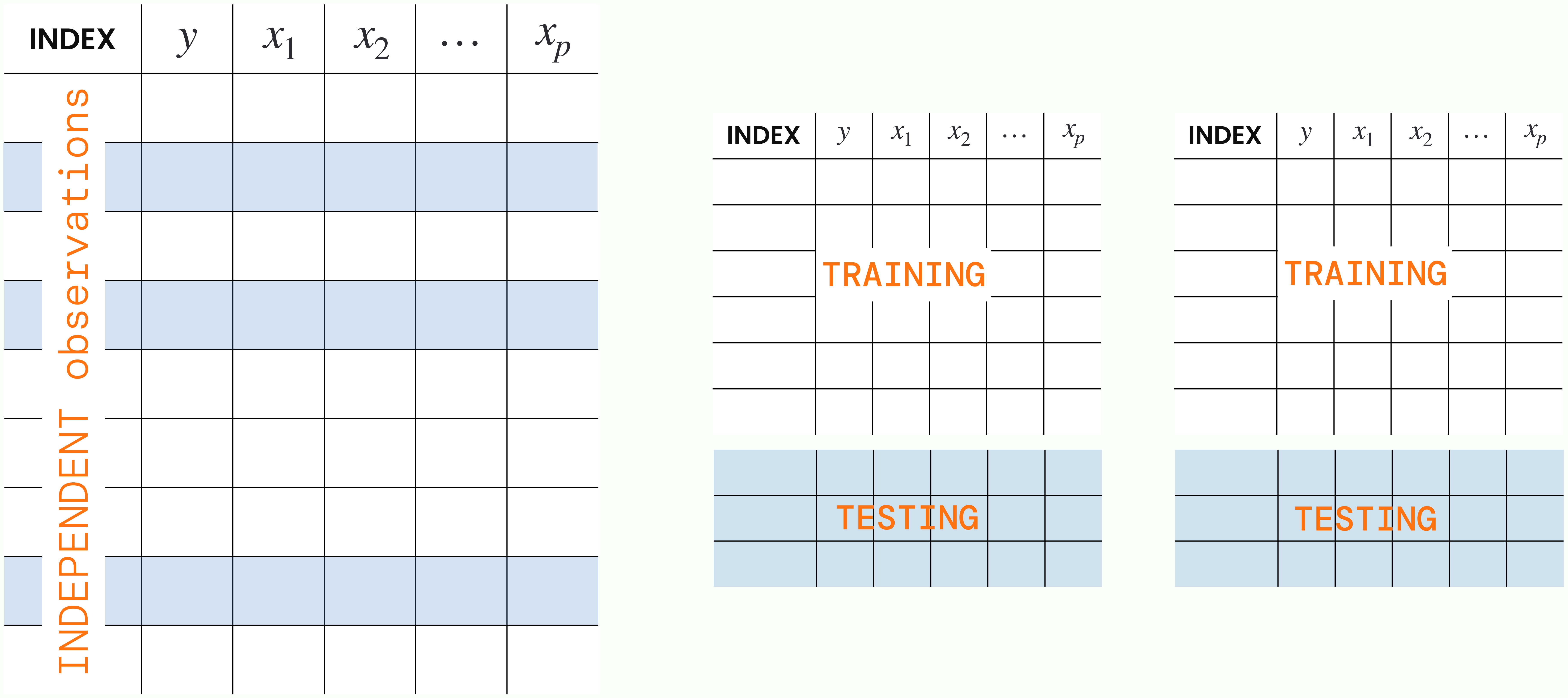


y	x_1
…	…
18.8	530
20.9	571
15.5	620
…	…
16.9	594
17.9	567
19.9	616
…	…

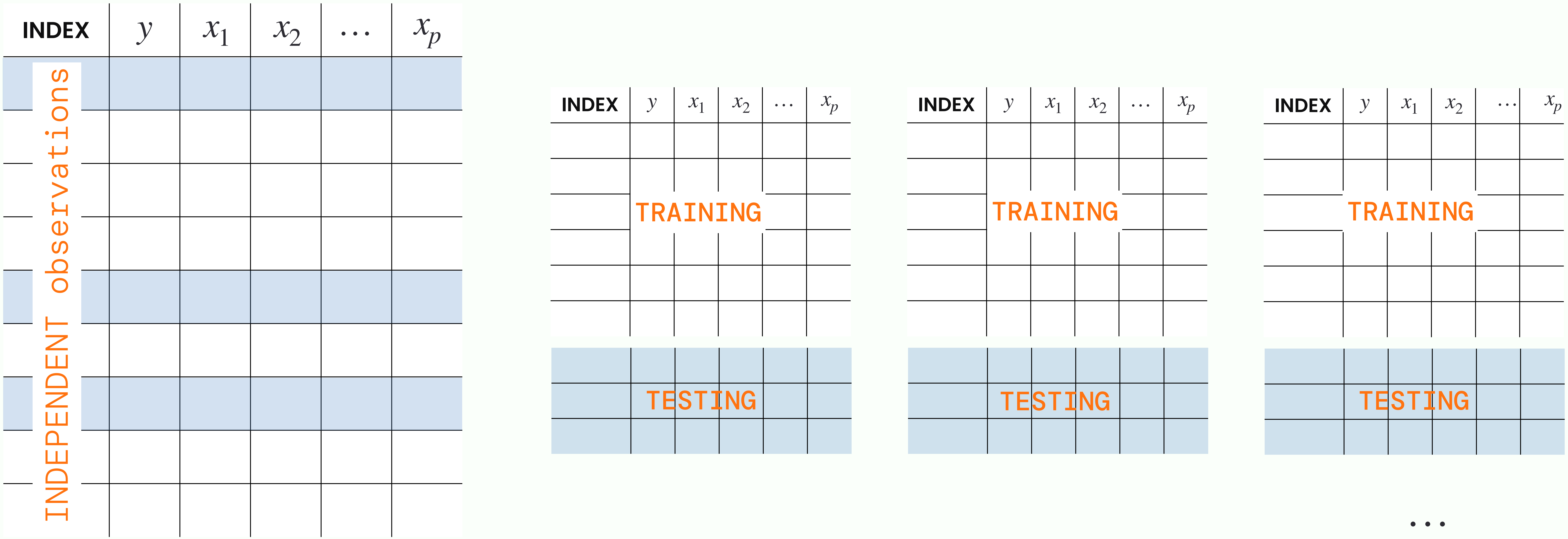
K-FOLD CROSS-VALIDATION



K-FOLD CROSS-VALIDATION

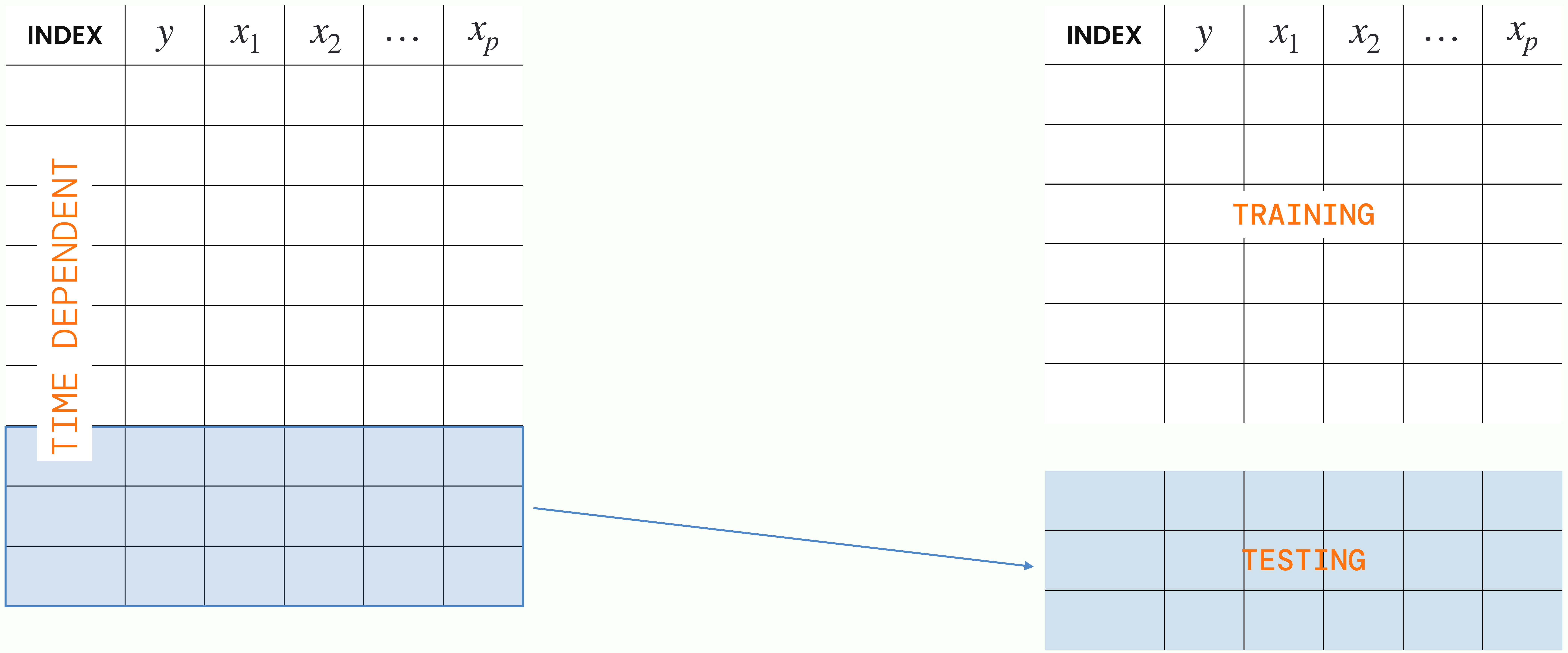


K-FOLD CROSS-VALIDATION

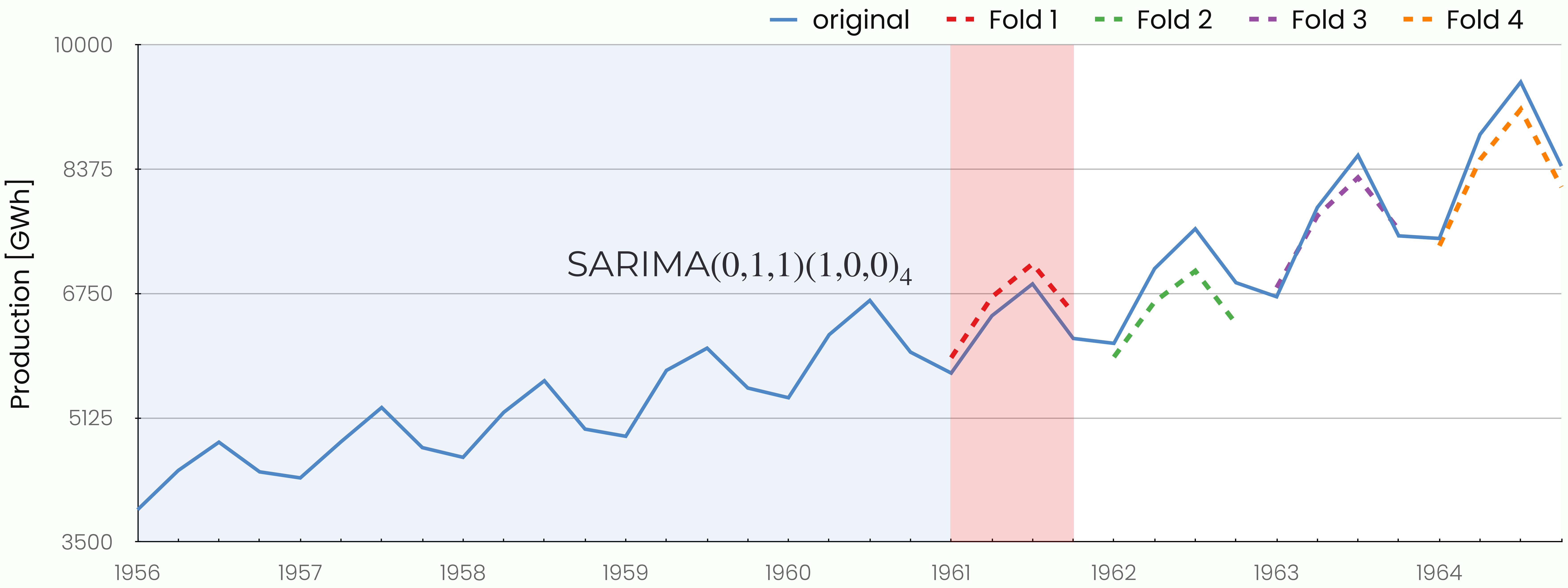


k-fold cross-validation

TS: ~~RANDOM~~ SPLIT FOR TRAINING/TESTING



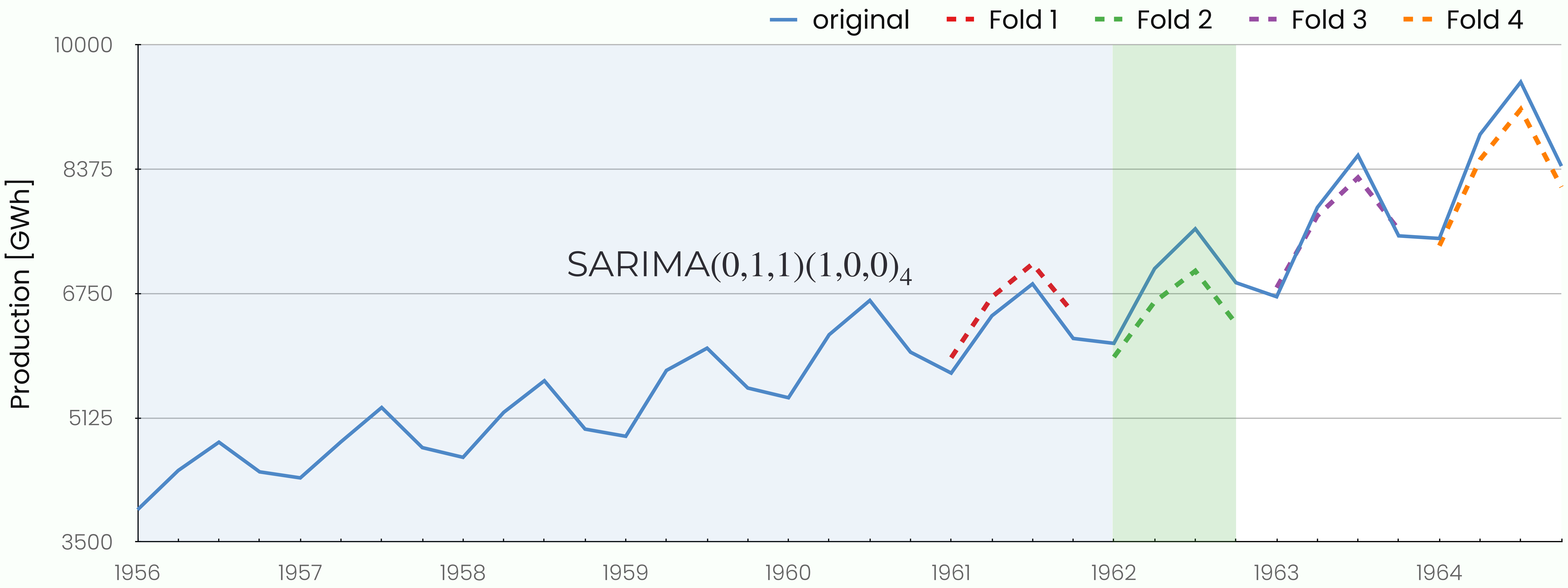
WALK–FORWARD VALIDATION



	Q1'56-Q4'60	Q1-Q4'61	Q1-Q4'62	Q1-Q4'63	Q1-Q4'64	SARIMA $(0,1,1)(1,0,0)_4$	SARIMA $(0,1,2)(1,0,0)_4$
Fold 1	train	test				4.13	2.62
Fold 2	train	train	test			6.02	3.69
Fold 3	train	train	train	test		1.91	3.14
Fold 4	train	train	train	train	test	3	4.66
avg MAPE						**3.76**	**3.53**

Alternative method is **sliding window validation** which uses fixed-size window for training.

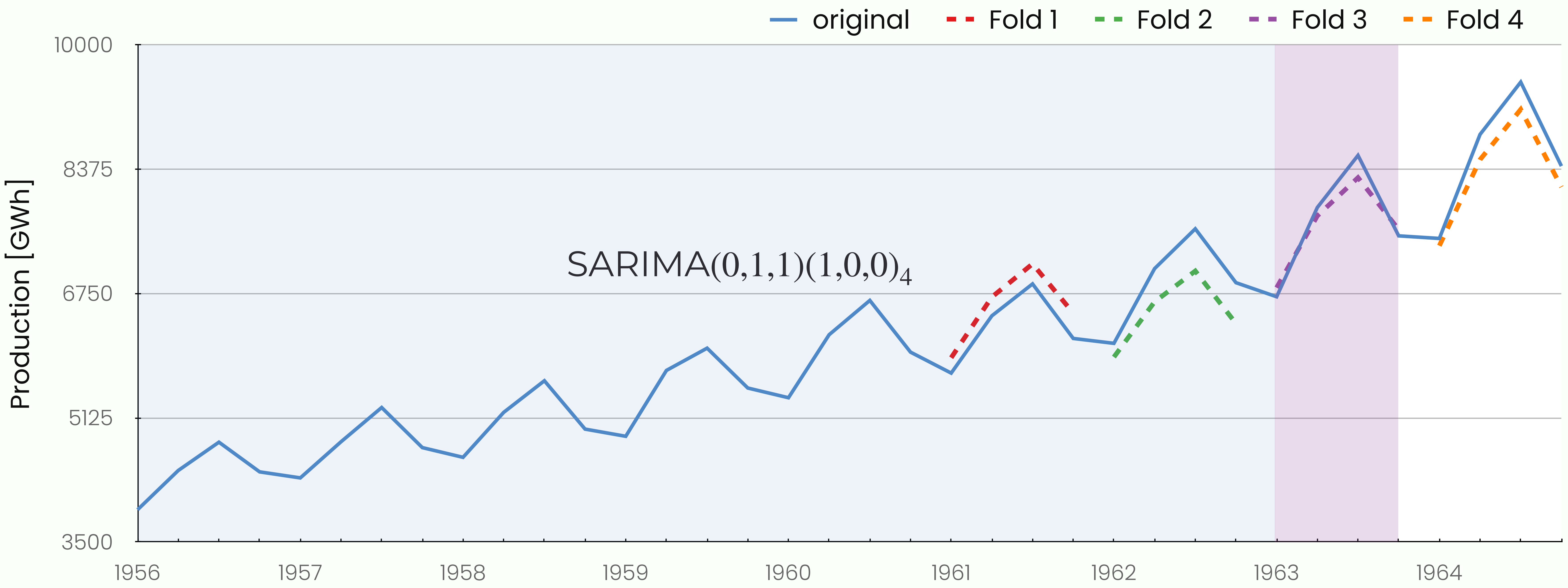
WALK–FORWARD VALIDATION



	Q1'56-Q4'60	Q1-Q4'61	Q1-Q4'62	Q1-Q4'63	Q1-Q4'64	SARIMA $(0,1,1)(1,0,0)_4$	SARIMA $(0,1,2)(1,0,0)_4$
Fold 1	train	test				4.13	2.62
Fold 2	train	train	test			6.02	3.69
Fold 3	train	train	train	test		1.91	3.14
Fold 4	train	train	train	train	test	3	4.66
avg MAPE						**3.76**	**3.53**

Alternative method is **sliding window validation** which uses fixed-size window for training.

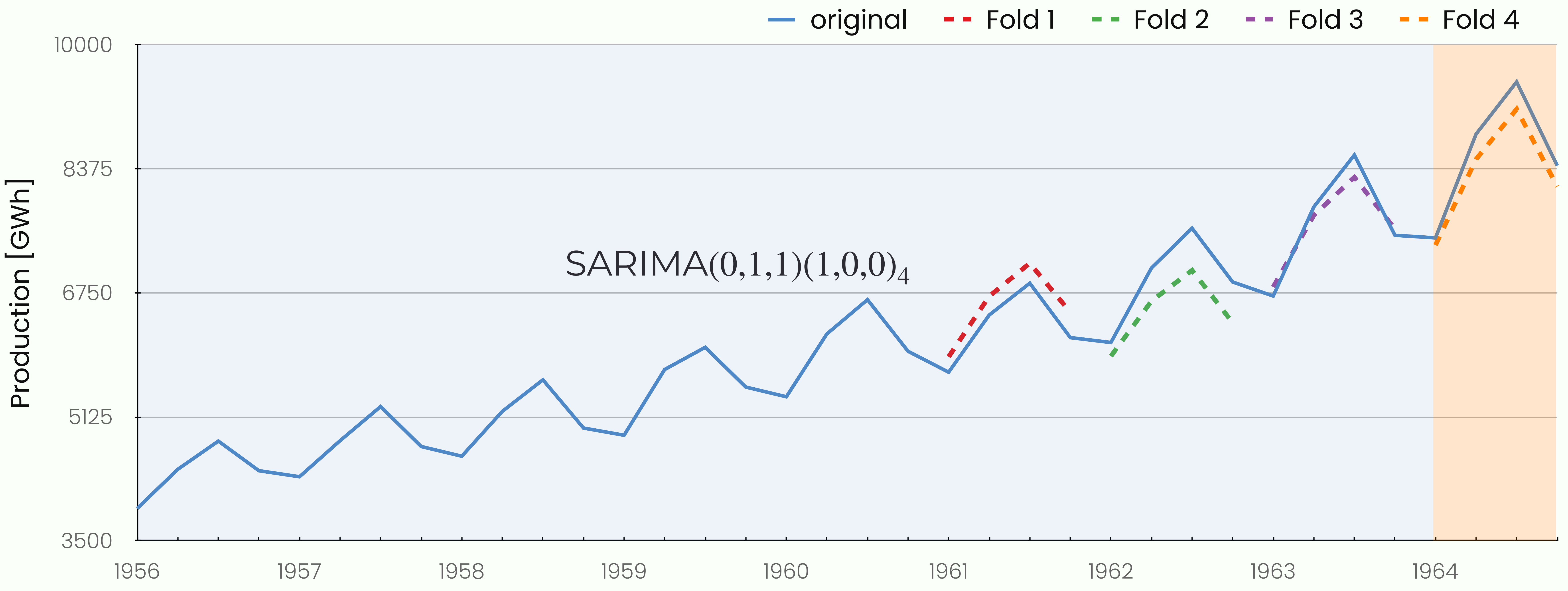
WALK–FORWARD VALIDATION



	Q1'56–Q4'60	Q1–Q4'61	Q1–Q4'62	Q1–Q4'63	Q1–Q4'64	SARIMA $(0,1,1)(1,0,0)_4$	SARIMA $(0,1,2)(1,0,0)_4$
Fold 1	train	test				4.13	2.62
Fold 2	train	train	test			6.02	3.69
Fold 3	train	train	train	test		1.91	3.14
Fold 4	train	train	train	train	test	3	4.66
avg MAPE						**3.76**	**3.53**

Alternative method is **sliding window validation** which uses fixed-size window for training.

WALK-FORWARD VALIDATION



	Q1'56-Q4'60	Q1-Q4'61	Q1-Q4'62	Q1-Q4'63	Q1-Q4'64	SARIMA $(0,1,1)(1,0,0)_4$	SARIMA $(0,1,2)(1,0,0)_4$
Fold 1	train	test				4.13	2.62
Fold 2	train	train	test			6.02	3.69
Fold 3	train	train	train	test		1.91	3.14
Fold 4	train	train	train	train	test	3	4.66
		avg MAPE				**3.76**	**3.53**

Alternative method is **sliding window validation** which uses fixed-size window for training.

HOME ACTIVITIES & BRAIN EXERCISE

- ✳ Try other orders for ARMA model and compare results.
- ✳ Try different lengths of a dataset.
- ✳ Try other transformations (incl. variation stabilisation) and compare results.
- ✳ Given a time series y_t of the length T. Compare the results calculated by

$$r_i = \text{ACF}(y_{t-i})$$

and

$$\text{Pearson's coefficient } r_i = \operatorname{corr}(y_t, y_{t-i})$$

for an arbitrary $i \in \{1,\ldots,T-1\}$.

Thank you!

Questions?