

# Time Series Analysis

## Foundations II

**Dr. Ludovic Amruthalingam**  
ludovic.amruthalingam@hslu.ch

**Informatik**



# Outline

- Stationarity
- Visual inspection
- Statistical tests
- Non-linear transforms
- Decomposition
- Differencing

# Stationarity

**Strict stationarity:** the joint distribution of any collection of the TS RVs is identical to that of the shifted RVs:

$$P[X_{t_1} = x_{t_1}, \dots, X_{t_k} = x_{t_k}] = P[X_{t_1+h} = x_{t_1+h}, \dots, X_{t_k+h} = x_{t_k+h}] \quad \forall k \geq 1, \forall t_1, \dots, t_k, \forall h \geq 0$$

**(Weak) stationarity:** both mean and variance are **constant** over time, while the auto-covariance depends **only on the lag**  $h = |s - t|$  between observations :

$$\gamma(t + h, t) = \text{cov}(X_{t+h}, X_t) = \text{cov}(X_h, X_0) = \gamma(h, 0) = \gamma(h) \quad \forall h, t$$

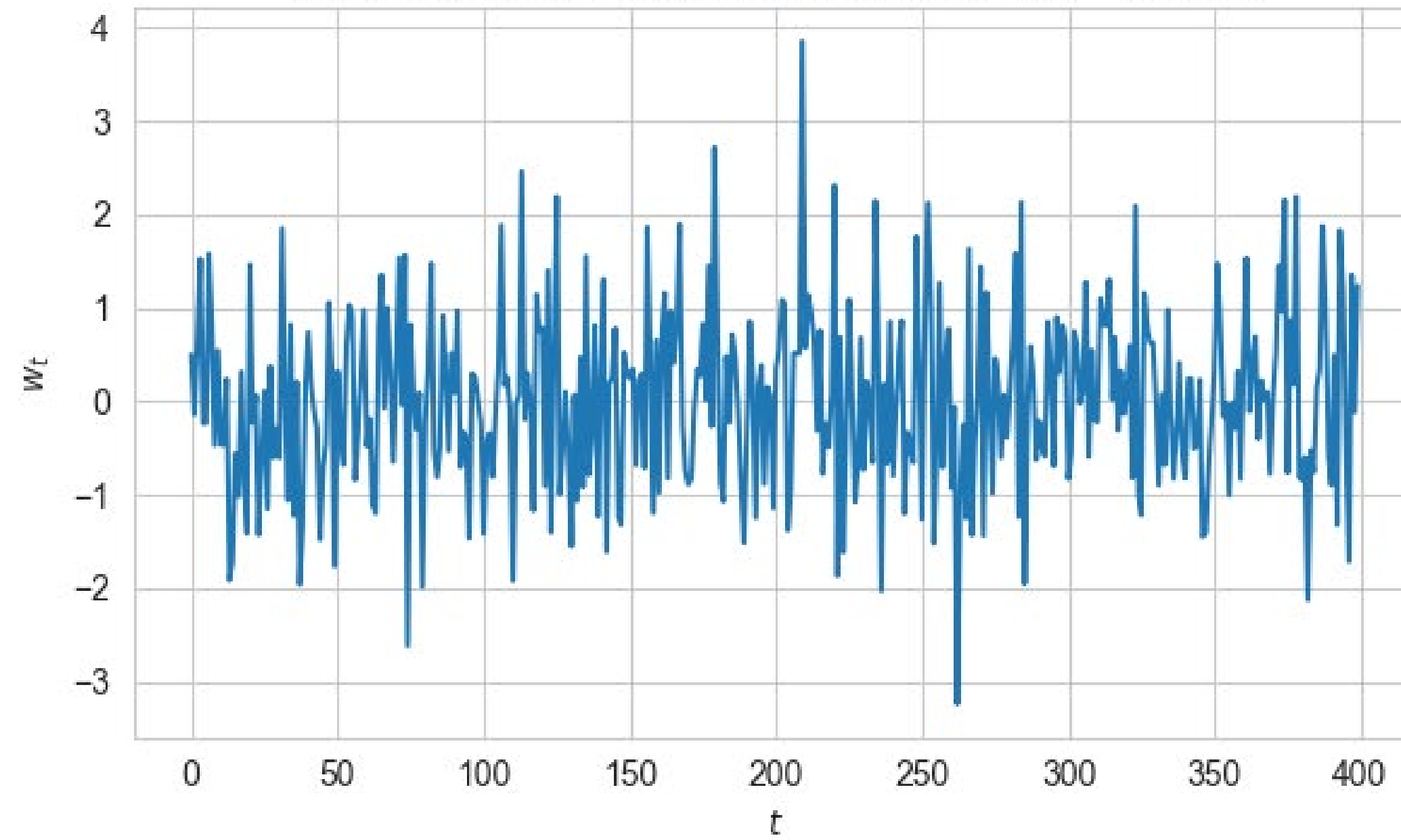
The **auto-correlation** only depends on the **shift** since

$$\rho(t + h, t) = \frac{\gamma(t + h, t)}{\sqrt{\text{Var}(X_{t+h})\text{Var}(X_t)}} = \frac{\gamma(h, 0)}{\sqrt{\text{Var}(X_h)\text{Var}(X_0)}} = \frac{\gamma(h)}{\gamma(0)} = \rho(h)$$

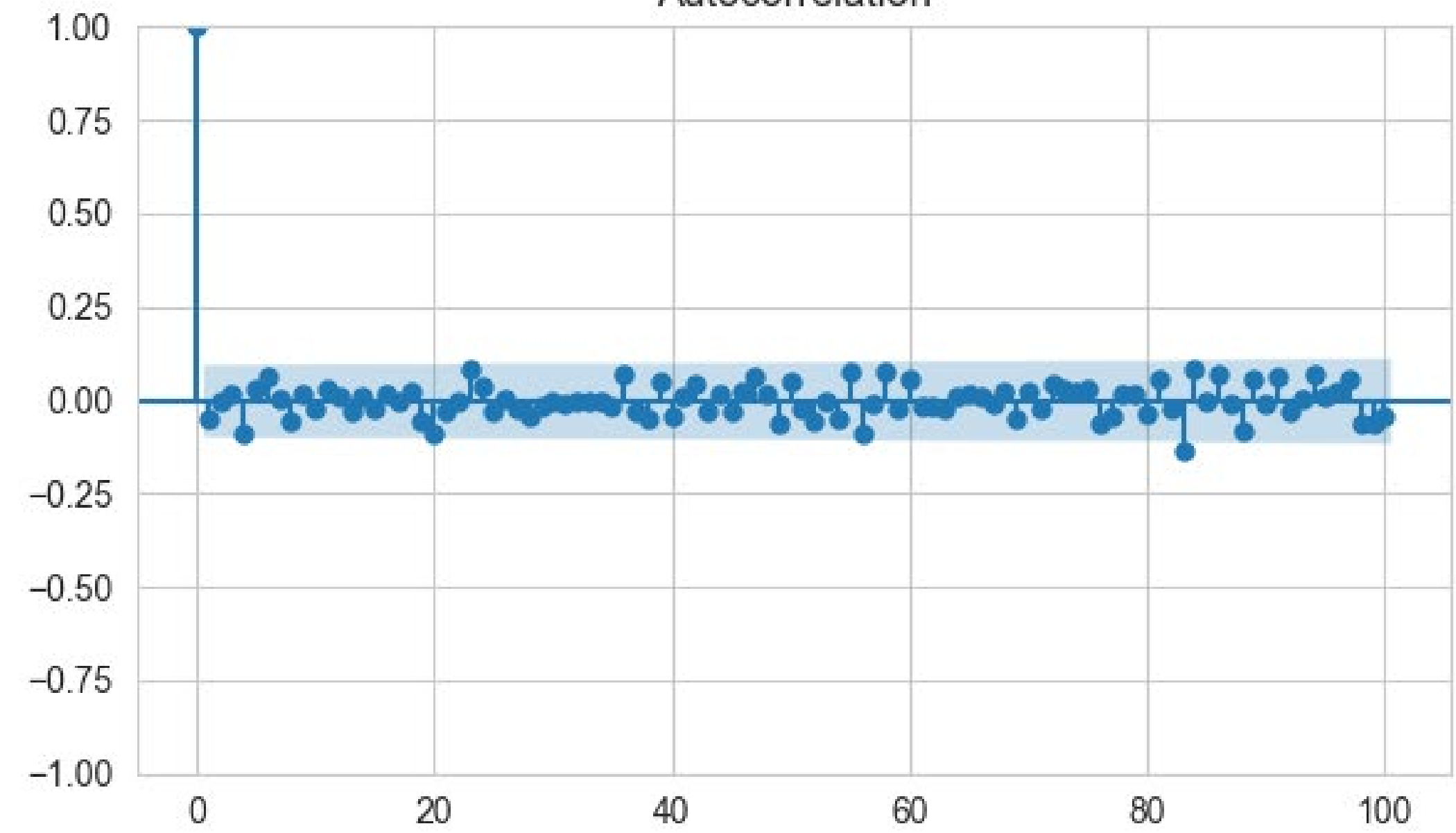
A **white noise process is stationary** since  $W_t \sim WN(0, \sigma^2)$  and  $\gamma_{W_t}(s, t) = 0 \quad \forall s \neq t$ .

Strict stationarity implies stationarity, but the reverse is **not** true in general (Gaussian processes are an exception).

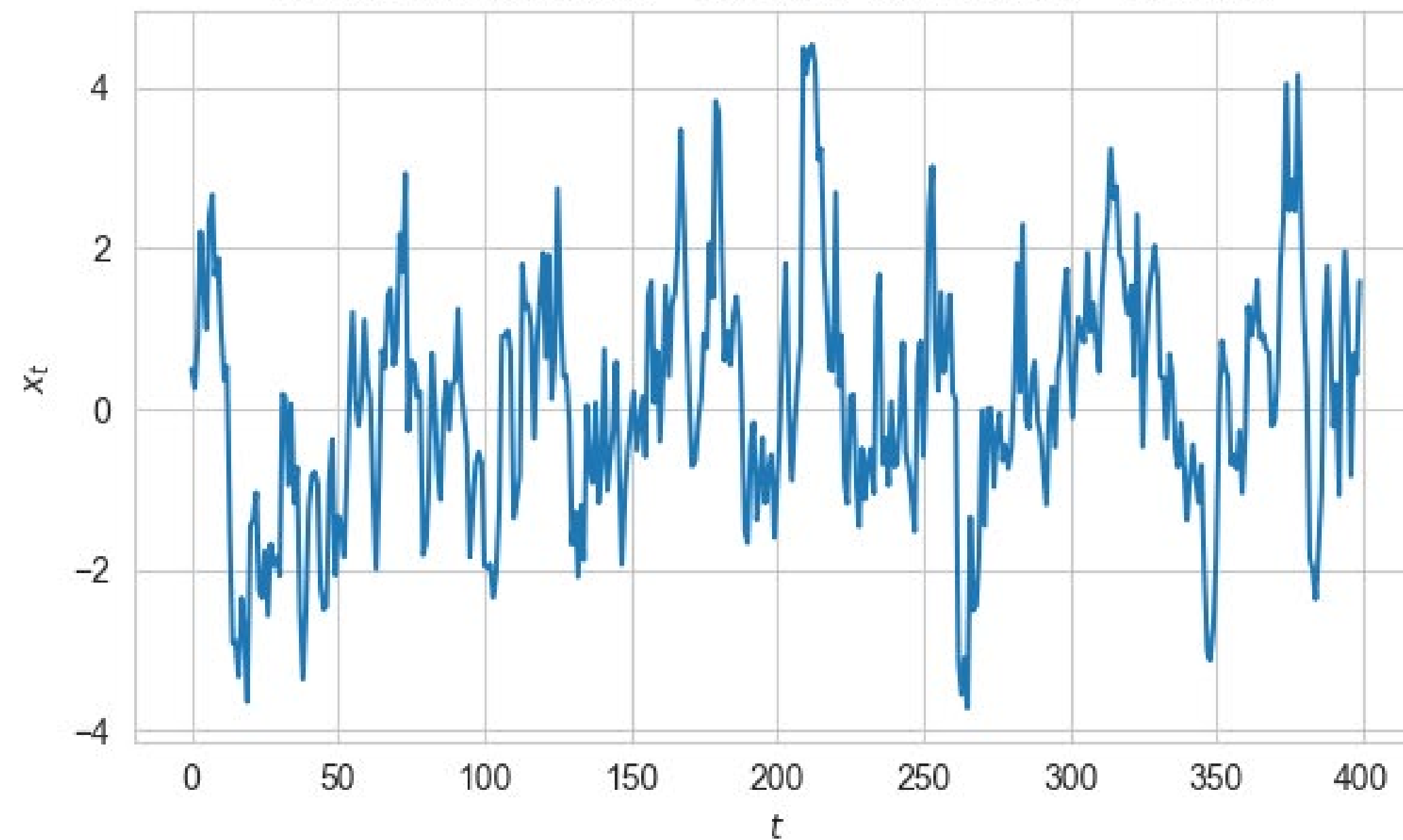
Strict Stationarity: Gaussian White Noise  $W_t \sim \mathcal{N}(0, 1)$



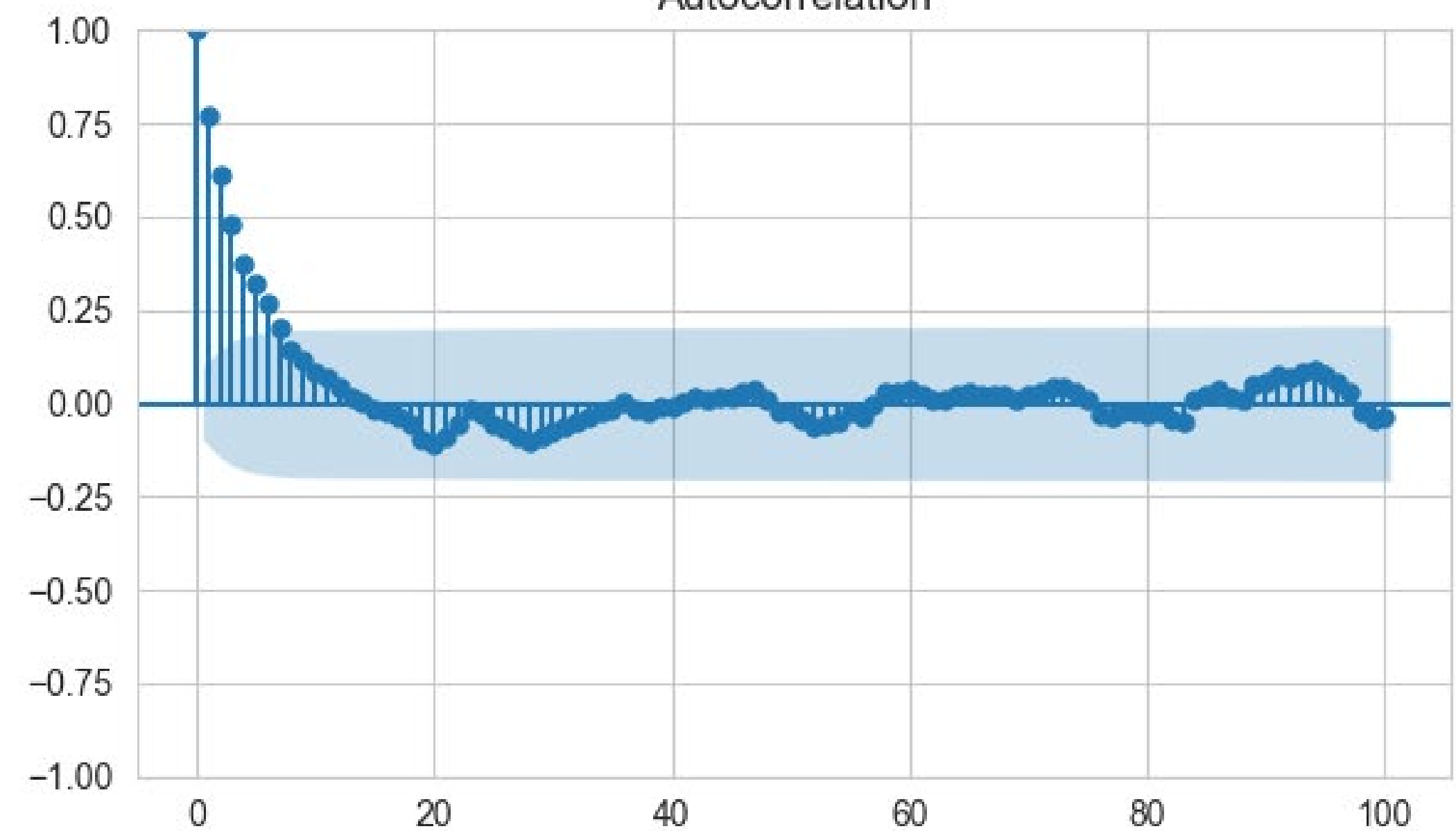
Autocorrelation



Weak Stationarity:  $X_t = 0.8X_{t-1} + W_t$  with  $W_t \sim \mathcal{N}(0, 1)$



Autocorrelation



# Estimating the statistical properties of a time series

Typically, only a **single realization** of a TS is available. With non-stationary TS,

- Actual mean, variance and covariance (statistical properties) may **vary over time**.
- **Cannot** rely on past values to **estimate future** relationships.

With a stationary TS, its **statistical properties are constant over time** and can be estimated from past values.

Sample mean	Sample variance	Sample auto-covariance	Sample auto-correlation
$\hat{\mu} = \frac{1}{n} \sum_{i=0}^n x_i$	$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=0}^n (x_i - \hat{\mu})^2$	$\hat{\gamma}(h) = \frac{1}{n} \sum_{i=0}^{n- h } (x_{i+ h } - \hat{\mu})(x_i - \hat{\mu})$	$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$

TS analysis typically starts with removing trends, seasonality, and heteroskedasticity to get **stationary residuals**, which can then be modeled.

# Checking stationarity – Visual inspection of time plots

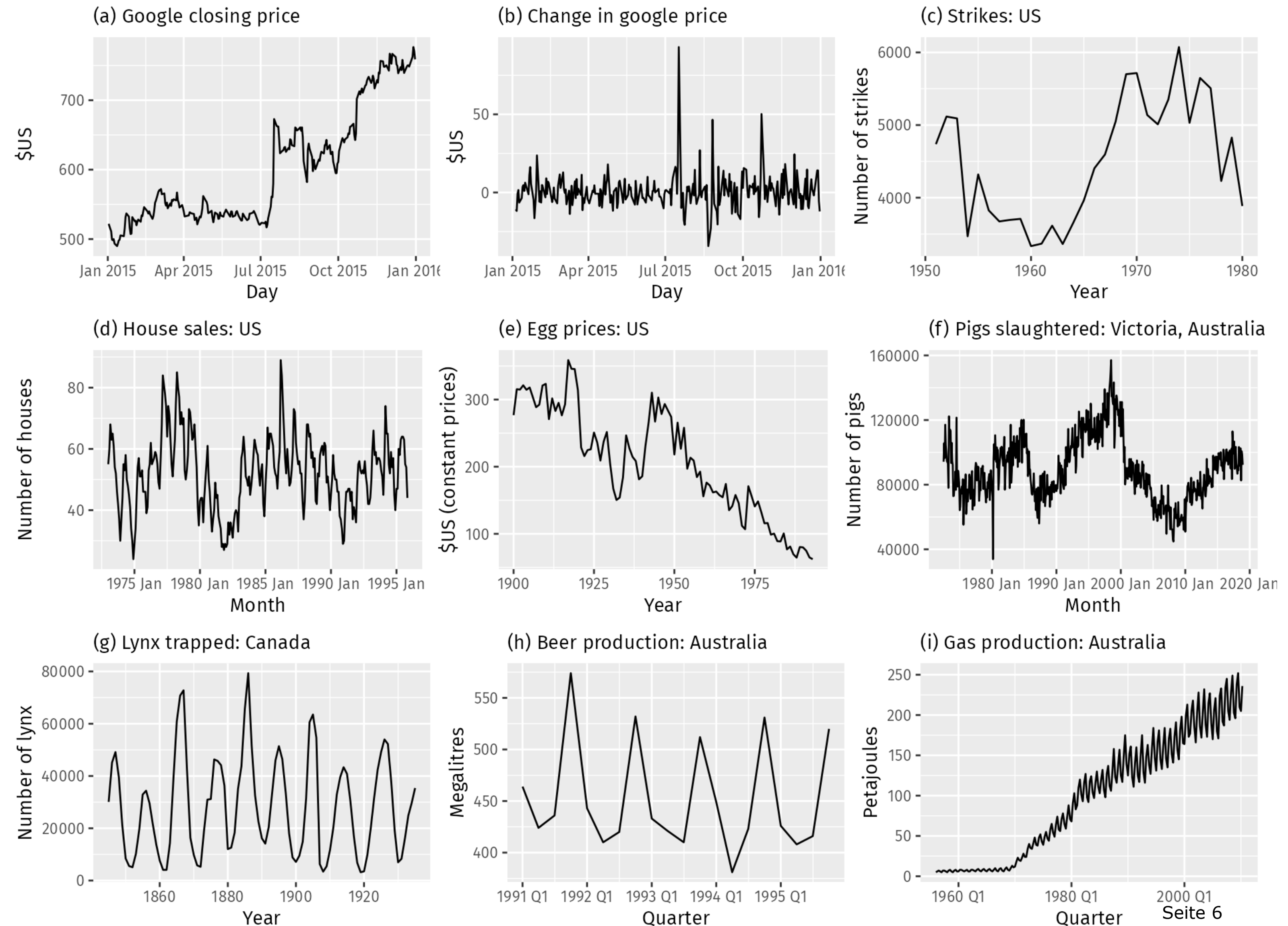
Which of these TS are stationary?

## Non-stationary patterns

- Trends
- Seasonality with trend
- Cycles with trend
- Changing variance
- Apparent breaks in the data

## Cyclic patterns can be stationary:

ups and downs are random but stable in distribution (time-invariant).

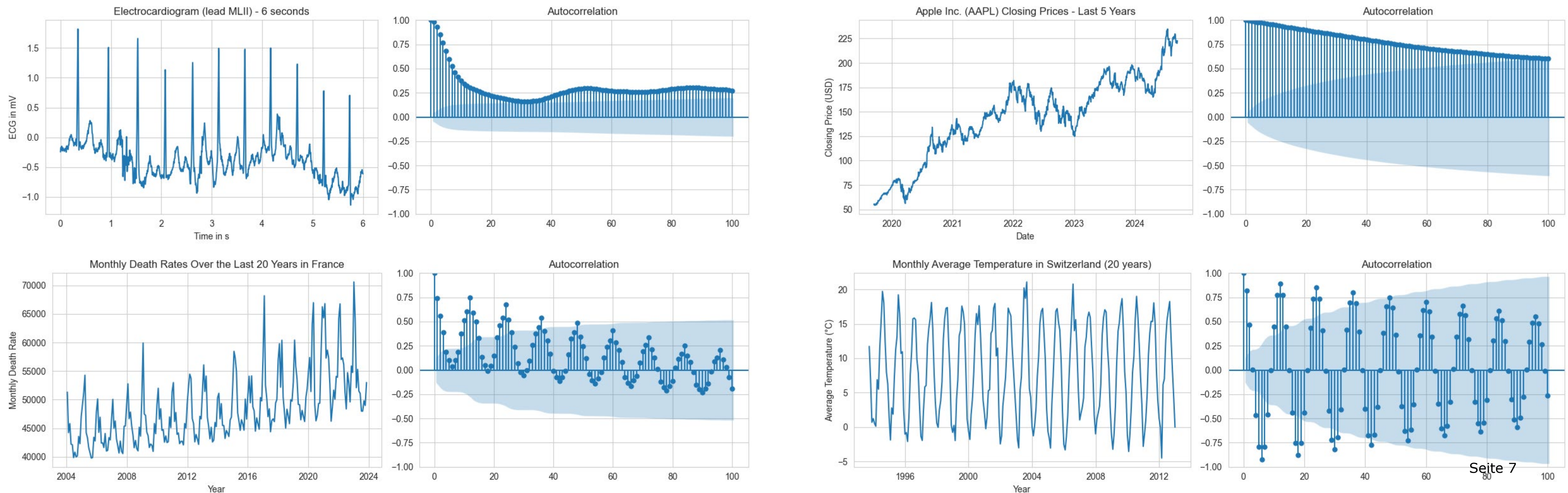




# Checking stationarity – Visual inspection of correlograms

Correlogram: in stationary TS, the auto-correlations typically **drop off quickly** as the lag increases. If the coefficients drop quickly but then show a pattern, it typically indicates **weak stationarity**.

ACF estimates are biased: shrunken towards zero for higher lags → consider only the **first  $\sim n/4$  lags**.



# Checking stationarity – Statistical tests

Null hypothesis ( $H_0$ ): TS has a **unit root** e.g., random walk  $x_t = 1 \cdot x_{t-1} + w_t$

- TS value is strongly dependent on its previous value => time-dependent mean/variance, slow-decaying  $\rho(h)$ .
- Augmented Dickey-Fuller (ADF) test, Phillips-Perron (PP) test

$H_0$ : TS is **trend stationary** (removing trend leaves a stationary TS)

- Detect whether TS is mean-reverting, has constant variance and auto-correlation over time.
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.

Tests for structural breaks: Zivot-Andrews (ZA) test, Chow test, Bai-Perron test.

Statistical tests focus on specific stationarity aspects, but do **not** test stationarity in a broad sense.

They struggle to detect general forms of non-stationarity → **couple with visual analysis.**



# Achieving stationarity

To make a TS stationary, apply either of the following steps (if necessary).

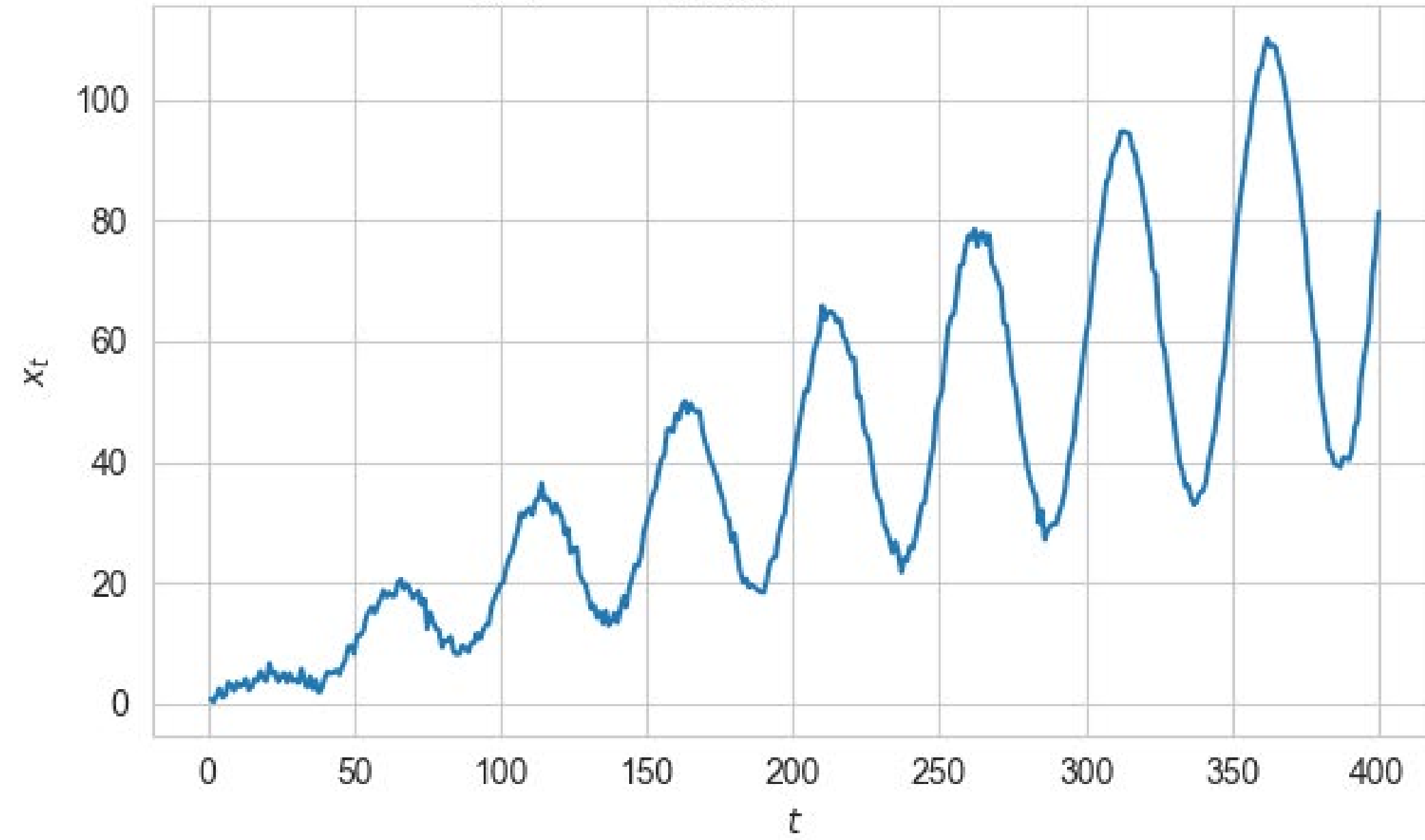
1. **Non-linear transformations** to stabilize variance
  - Log transform
  - Power transforms
  - Box-Cox transform
2. **Decomposition** to isolate trend and seasonal components
  - STL decomposition
3. **Differencing**
  - On the decomposition remainder if the residuals are non-stationary.
  - On the TS directly when decomposition is either difficult (e.g., weak patterns) or not needed.

# Non-linear transformations

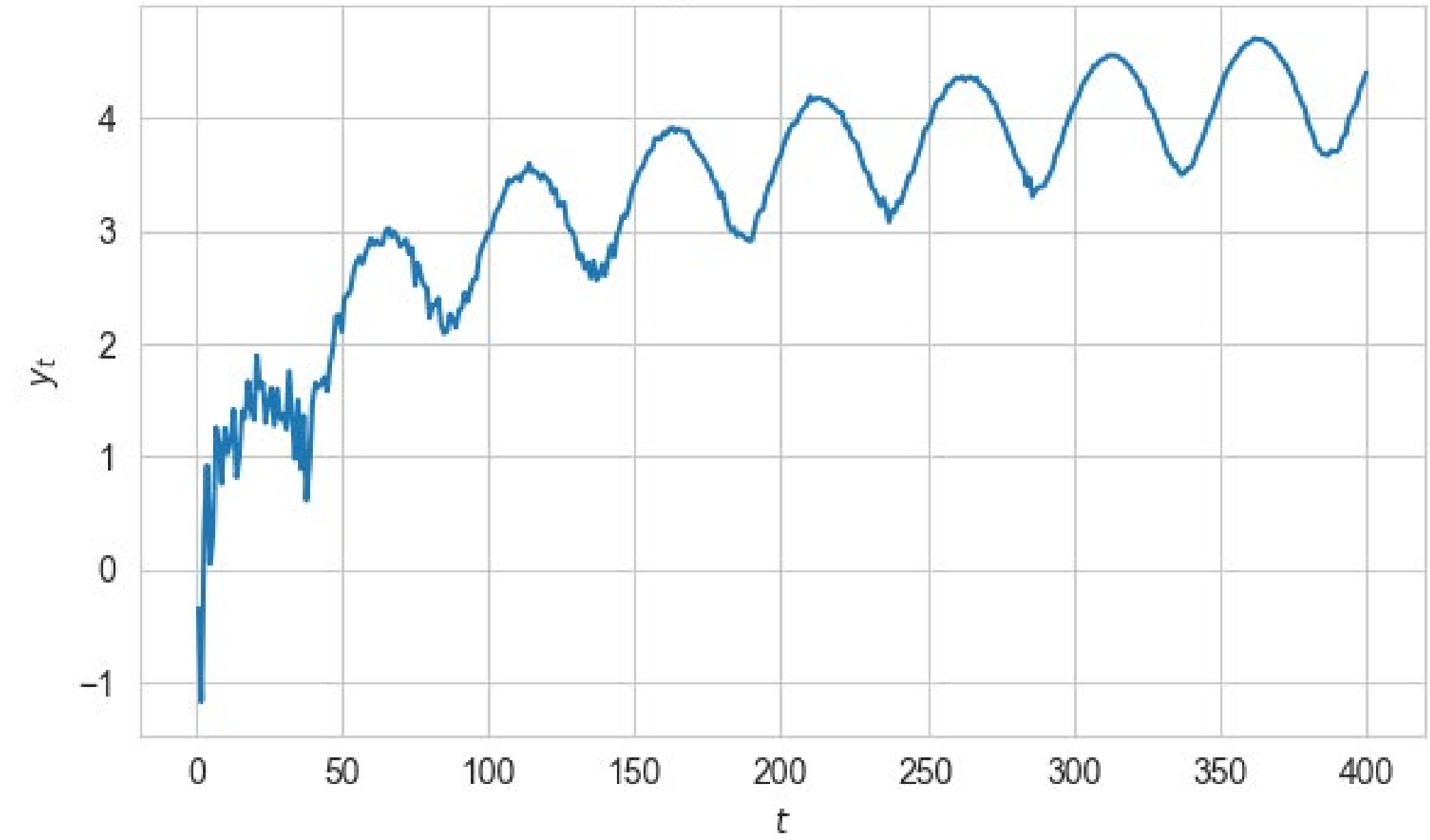
Goal: **stabilize** variance.

Log transform	Power transform	Box-Cox transform
$y_t = \log(x_t)$	$y_t = x_t^\lambda$	$y_t = \begin{cases} \log(x_t) & \text{if } \lambda = 0 \\ \frac{(\text{sign}(x_t) x_t ^\lambda - 1)}{\lambda} & \text{if } \lambda \neq 0 \end{cases}$
<ul style="list-style-type: none"><li>• Interpretable</li><li>• Use on skewed data with long right tail</li><li>• Converts multiplicative relationships into additive</li></ul>	<ul style="list-style-type: none"><li>• Lack interpretability</li><li>• Typically, <math>\lambda = \frac{1}{2}</math> or <math>\frac{1}{3}</math></li><li>• Use on count data (Poisson distribution)</li><li>• Reduce impact of large values</li></ul>	<ul style="list-style-type: none"><li>• Flexible transformation</li><li>• Determine optimal <math>\lambda</math> for each TS</li></ul>

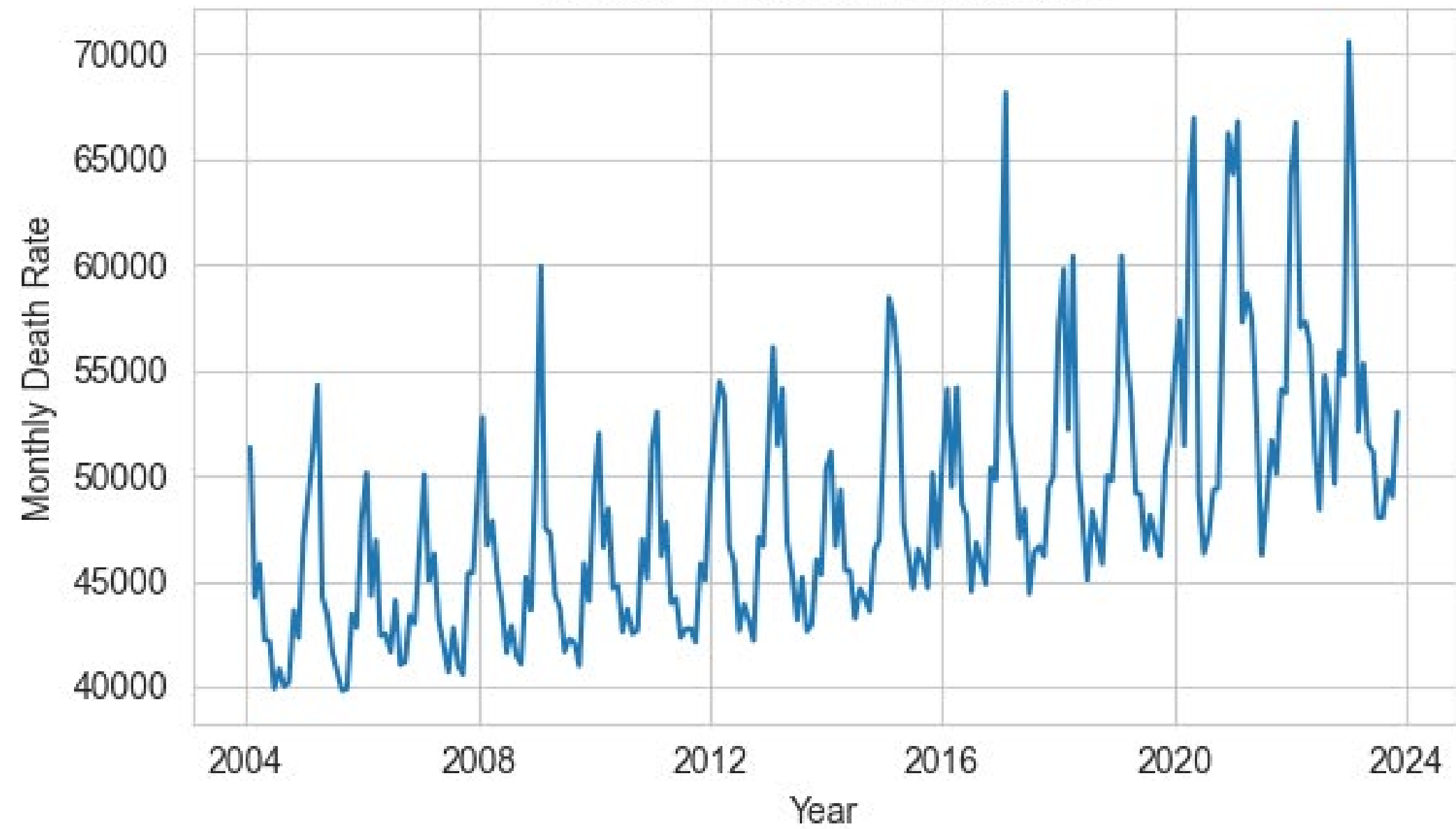
$$x_t = \frac{t}{10} \left( 2 + \sin\left(\frac{2\pi t}{50}\right) \right) + w_t \text{ with } w_t \sim \mathcal{N}(0, 1)$$



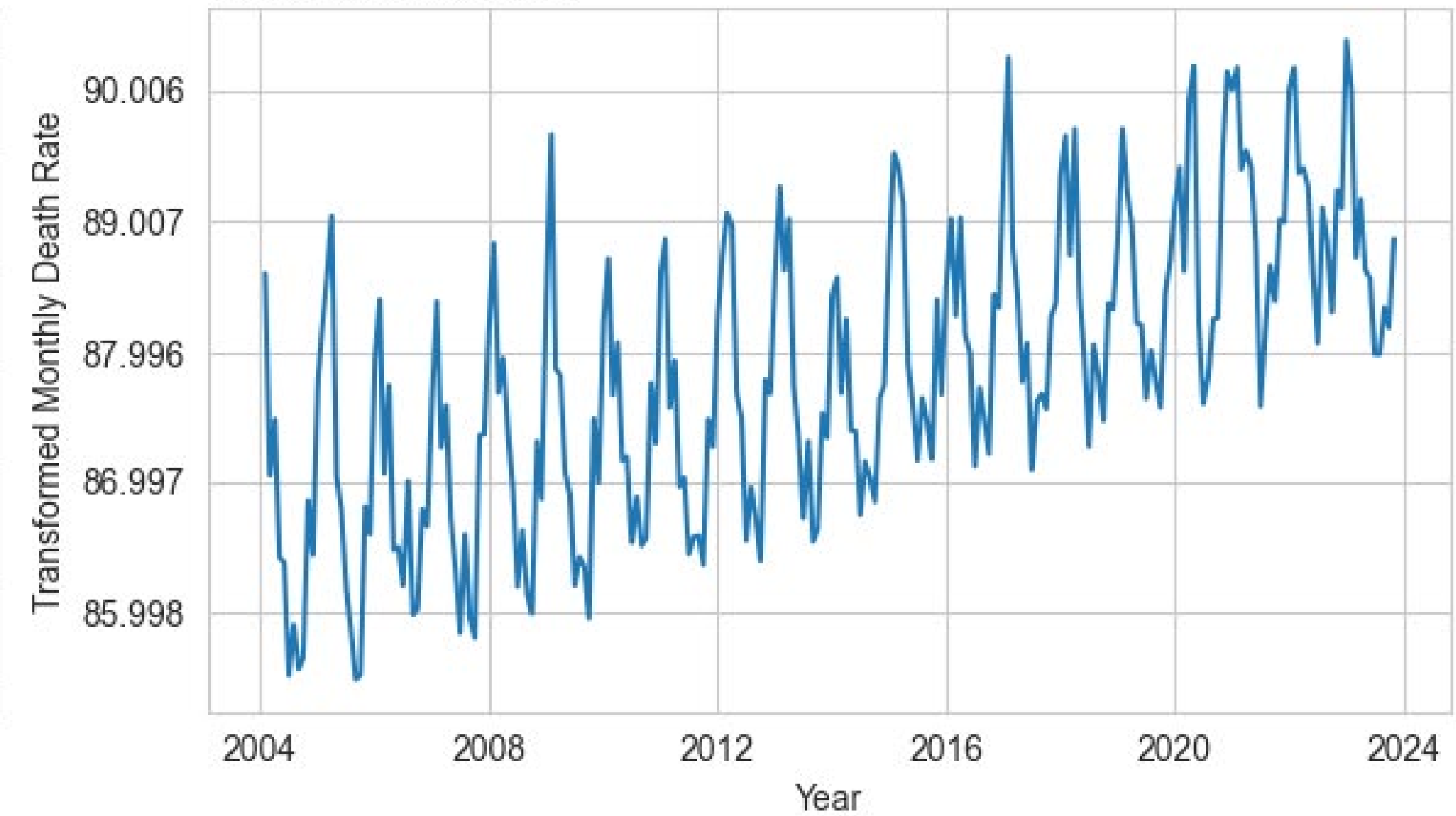
$$y_t = \log(x_t)$$



Monthly Death Rates in France



Box-Cox-transformed ( $\lambda = -2.78$ ) Monthly Death Rates in France  
 $1e-14 + 3.603226817e-1$



# Recap – Decomposition

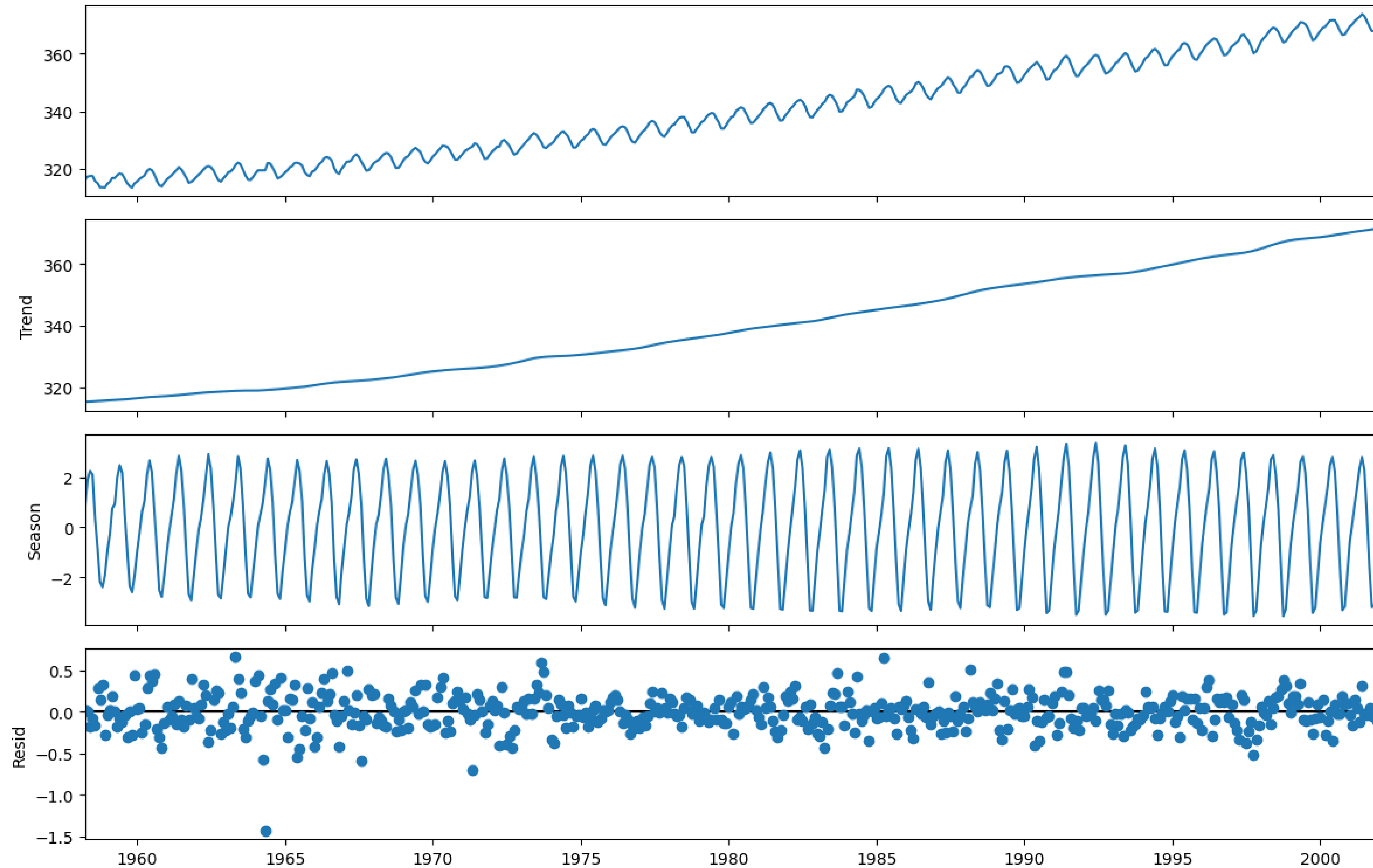
Separate TS into three components:  $X_t = U_t + S_t + R_t$

- Trend/trend-cycle  $U_t$ : increasing/decreasing pattern in the data.
- Seasonality  $S_t$ : repeating pattern with roughly fixed period.
- Remainder  $R_t$ : everything remaining (**residuals**).

Given a time series realization  $\{x_1, x_2, \dots, x_n\}$ :

1. Estimate trend component  $\hat{u}_t$ 
  - LOESS – locally estimated scatterplot smoothing (STL)
2. Detrend time series  $\hat{d}_t = x_t - \hat{u}_t$
3. Estimate seasonality component  $\hat{s}_t$ 
  - LOESS – locally estimated scatterplot smoothing (STL)
4. Deseasonalize TS to estimate the remainder  $\hat{r}_t = \hat{d}_t - \hat{s}_t$

STL decomposition: Mauna Loa Monthly Atmospheric CO2 Data





# Differencing

Studying a TS value **changes** rather than its values:  $\nabla X_t = X_t - X_{t-1}$  (**first-order differencing**)

$\nabla_h$  is called the **lag-h difference operator**:  $\nabla_h X_t = X_t - X_{t-h}$  with  $\nabla_1$  denoted as  $\nabla$ .

Differencing does **not** obtain explicit estimate of trend, season and remainder.

Differencing can **remove the trend** component: consider a TS with linear trend  $X_t = \alpha + \beta t + W_t$

Then  $\nabla X_t = \alpha + \beta t + W_t - (\alpha + \beta(t-1) + W_{t-1}) = \beta + W_t - W_{t-1}$

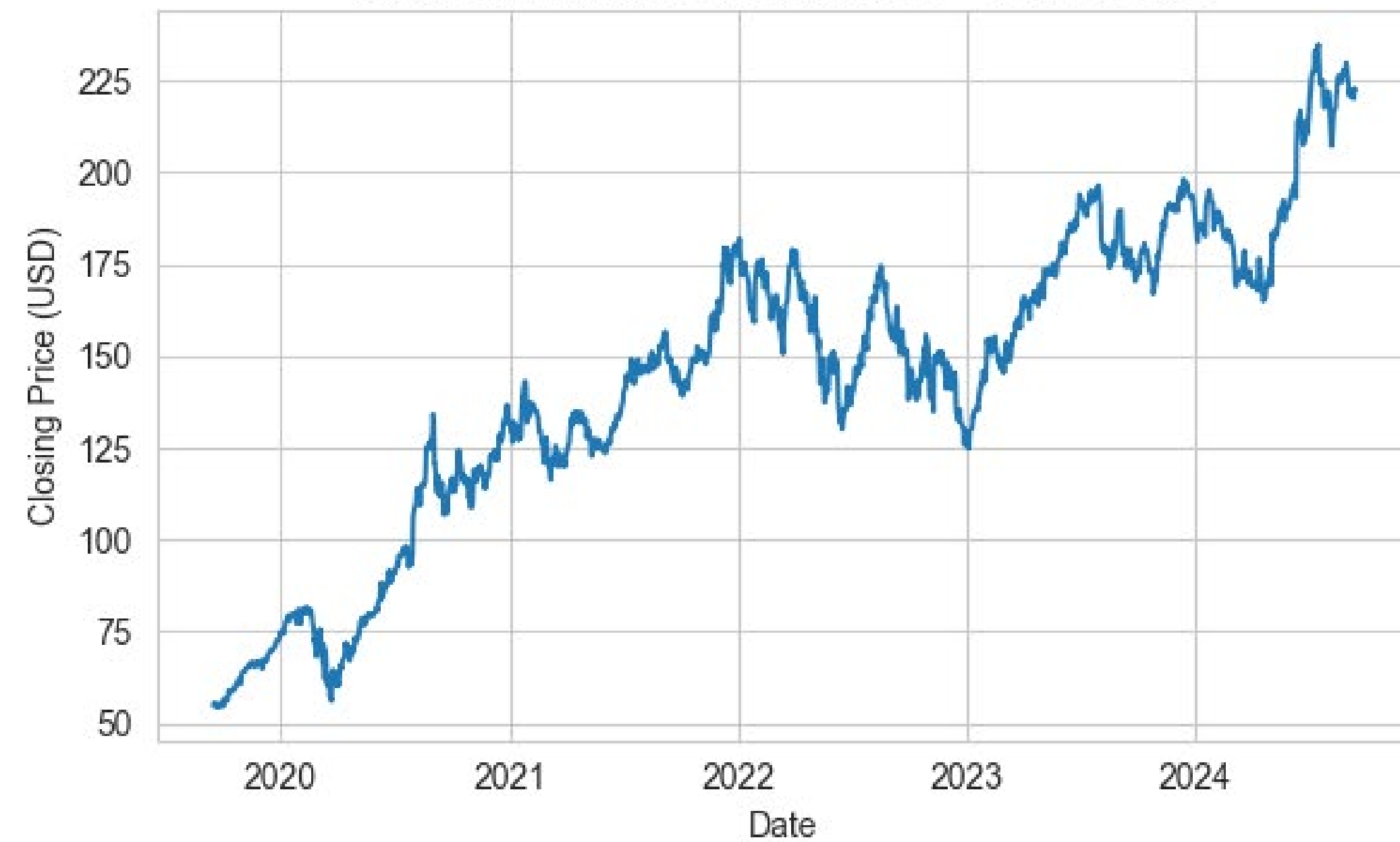
Differencing can remove the seasonal component: consider  $X_t = A \sin\left(\frac{2\pi t}{P}\right) + W_t$

Then  $\nabla_P X_t = X_t - X_{t-P} = A \sin\left(\frac{2\pi t}{P}\right) + W_t - \left(A \sin\left(\frac{2\pi(t-P)}{P}\right) + W_{t-P}\right) = W_t - W_{t-P}$  (**seasonal differencing**)

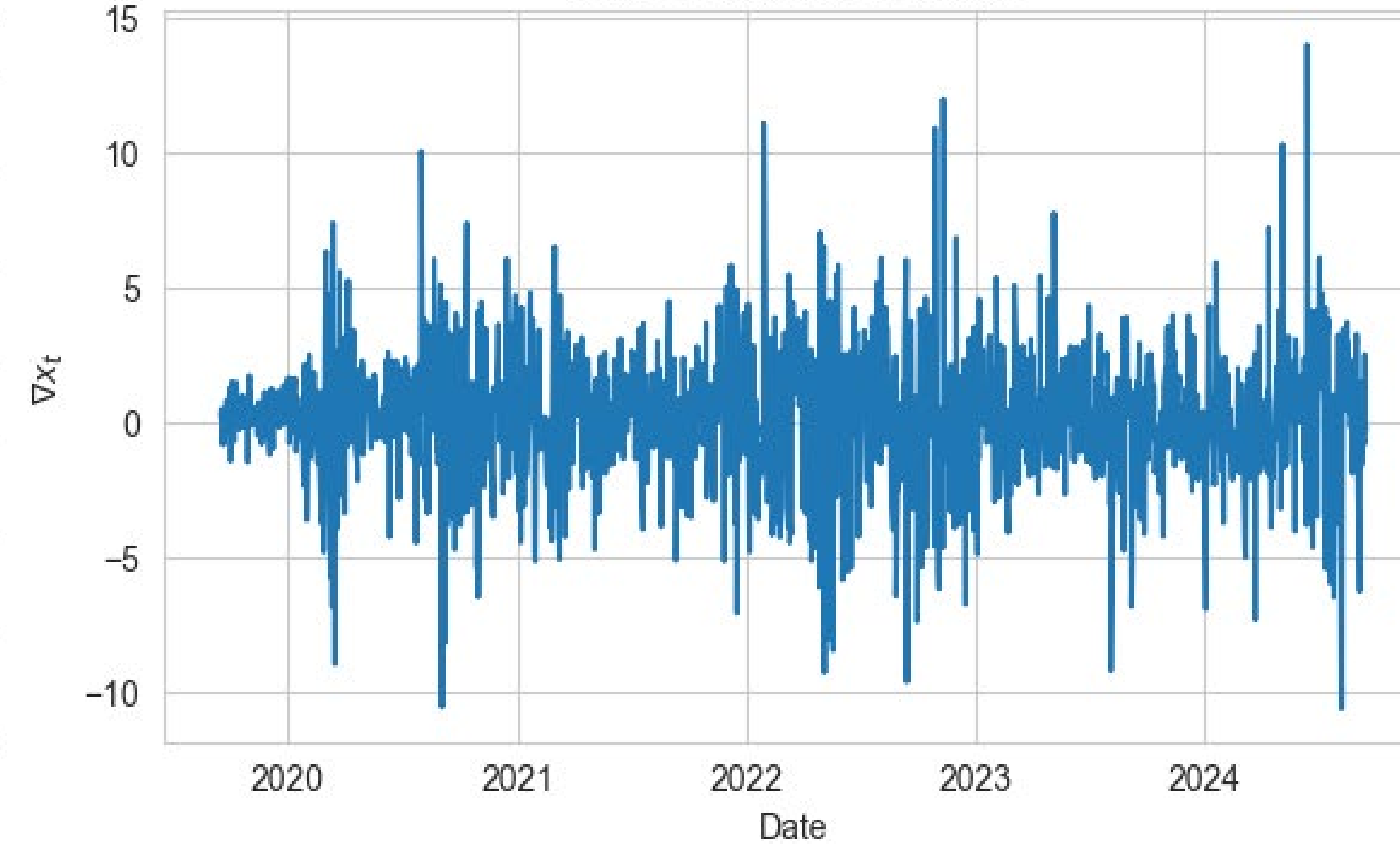
Differencing introduces **artificial dependencies**:  $\text{cov}(\nabla X_t, \nabla X_{t-1}) = \text{cov}(W_t - W_{t-1}, W_{t-1} - W_{t-2}) = -\text{cov}(W_{t-1}, W_{t-1}) \neq 0$

→ Keep differences **interpretable**: first-order, seasonal differencing.

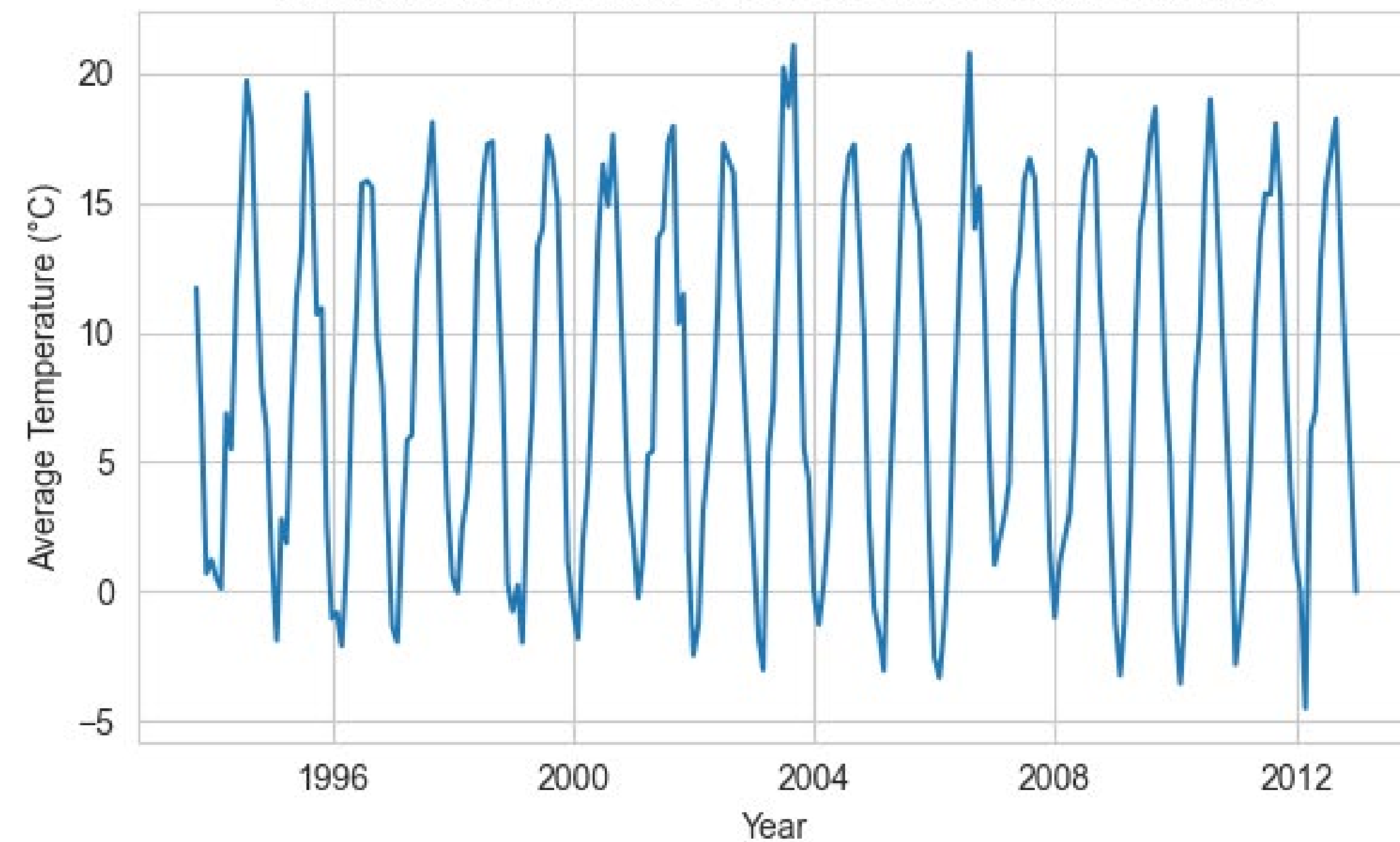
Apple Inc. (AAPL) Closing Prices - Last 5 Years



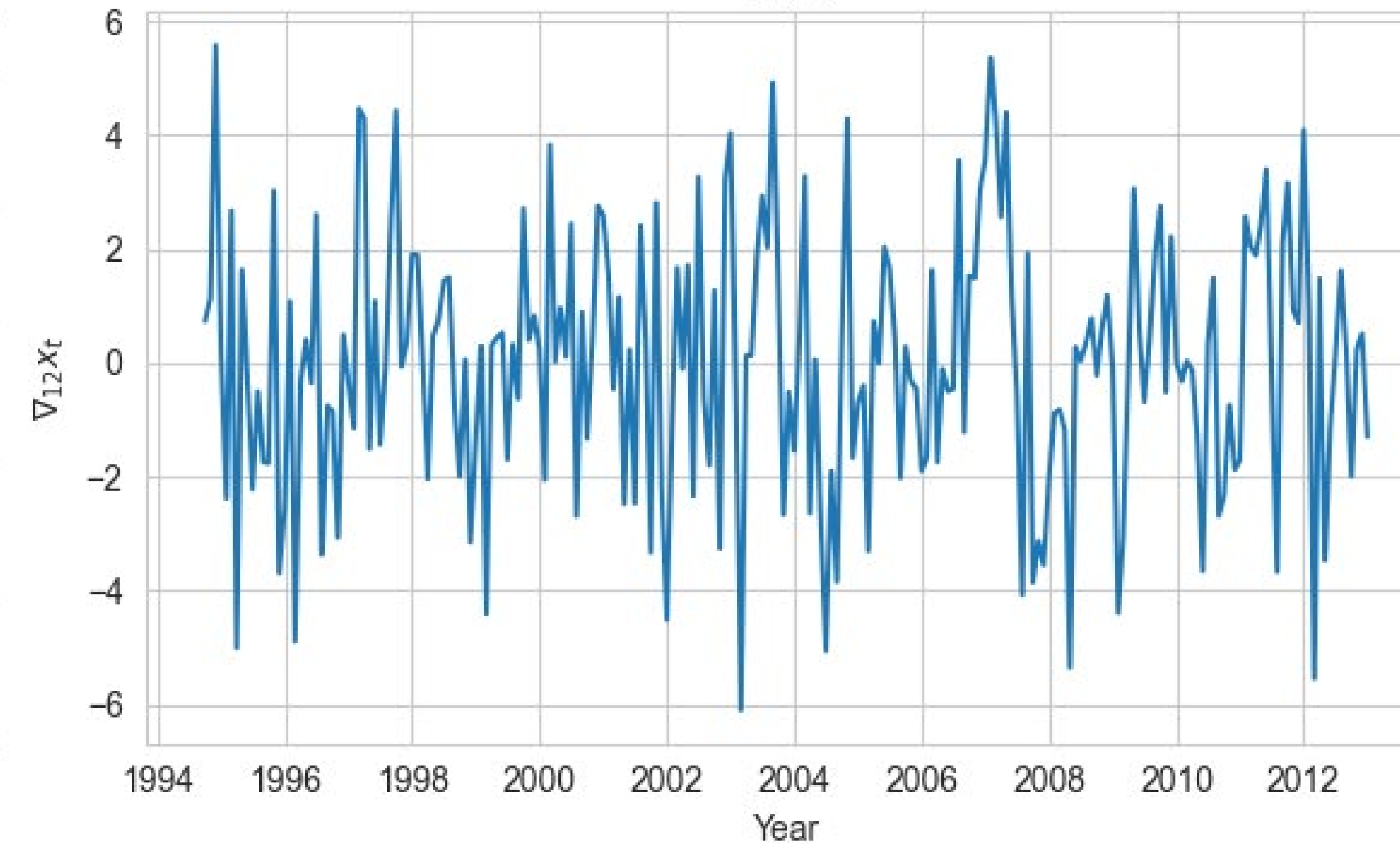
$\nabla X_t$ : Closing price change



Monthly Average Temperature in Switzerland (20 years)



$\nabla_{12} X_t$



## Higher-order differencing

Iterative differencing:  $\nabla^2 X_t = \nabla \nabla X_t = \nabla(X_t - X_{t-1}) = X_t - X_{t-1} - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}$

$B$  is called the **Backshift operator**:  $B^h X_t = X_{t-h}$  with  $B^1$  denoted as  $B$  and  $BX_t = X_{t-1}$

First-order differencing:  $\nabla X_t = X_t - X_{t-1} = X_t - BX_t = (1 - B)X_t$

Second-order differencing:  $\nabla^2 X_t = \nabla \nabla X_t = \nabla(1 - B)X_t = (1 - B)^2 X_t$

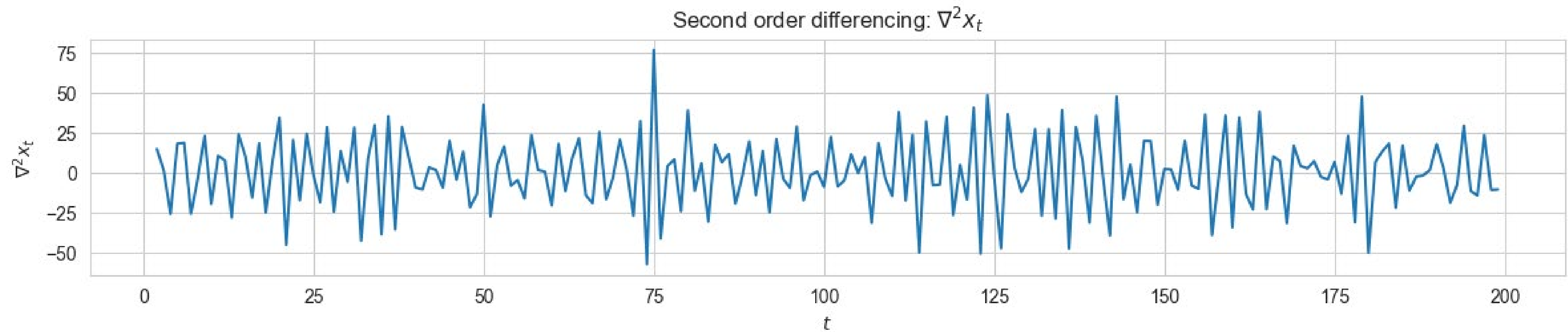
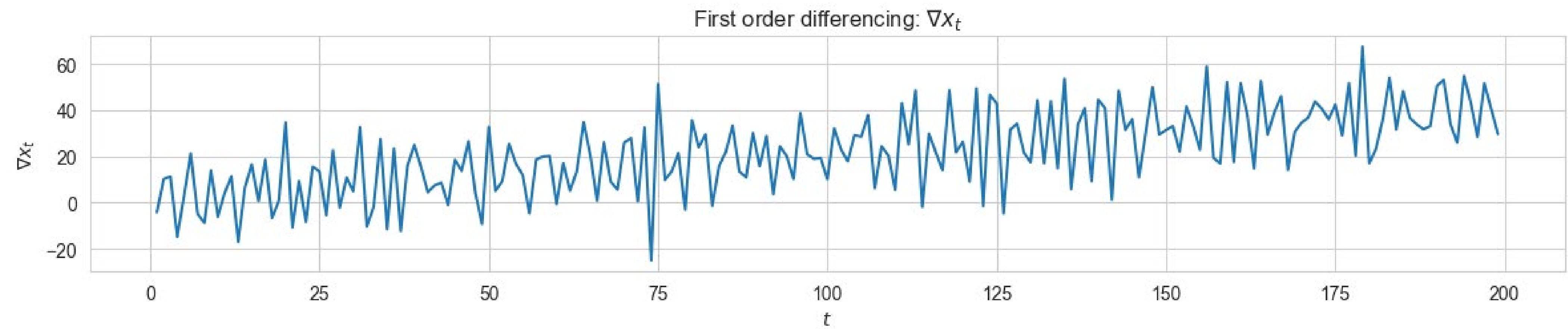
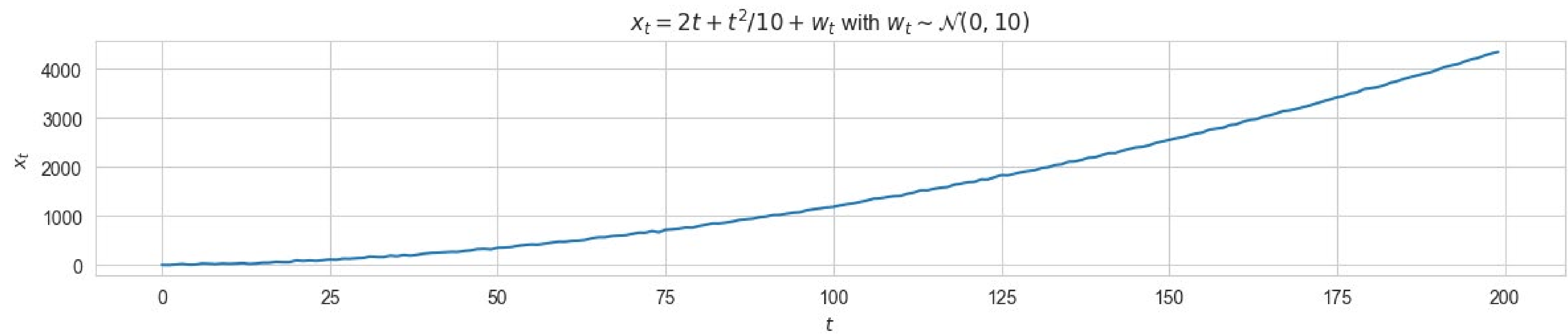
$k^{\text{th}}$ -order differencing:  $\nabla^k X_t = \nabla^{k-1} \nabla X_t = (1 - B)^k X_t$

Note that  $\nabla_h^k = (1 - B^h)^k$

**Start** with seasonal differencing, continue with differencing if some trend remain e.g.,  $\nabla^k \nabla_P X_t$

When trend is non-linear, apply non-linear transforms **before** any differencing.

Use higher-order differencing to remove **polynomial trend** e.g., second-order for quadratic trend.



# Exercise

Extend lecture 1 & 2 exercises with stationarity analysis: are the time series **stationary**?

- Visual inspection
- Statistical tests

Make the time series stationary

- Experiment with **non-linear transformations**.
- Experiment with **decomposition**.
- Experiment with **differencing**.

Explore **artificial auto-correlations** induced by differencing

- Start with a stationary time series.
- Apply differencing at various order.
- Compare **ACF**.