Time series

UE de Master 2, AOS1
Fall 2022

S. Rousseau

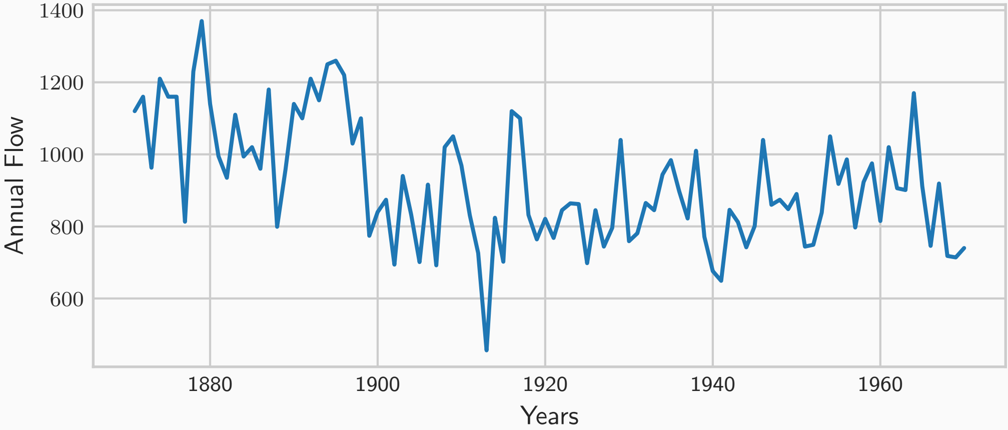
What is a time series

- A **time series** is a sequence of observations y_t recorded at a specific time t
- A **time series model** is a sequence of random variables Y_t where y_t is a realization of Y_t
- Also known as stochastic process $(Y_t)_{t \in \mathbb{Z}}$

- Observations are time dependent: assumption that observations are independent doesn't hold here
- Statistical tools that require iid samples don't apply here
- Need to develop specific methods summarized under **time series analysis**

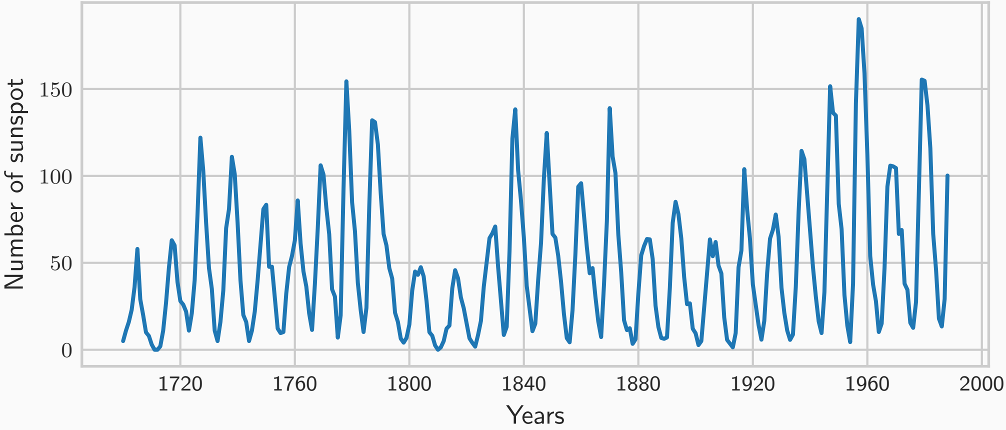
Example

Measurements of the annual flow of the river Nile at Aswan, 1871–1970, in $10^8 \cdot m^3$



Example

Yearly numbers of sunspots from 1700 to 1988



Nonstationarity

Trend

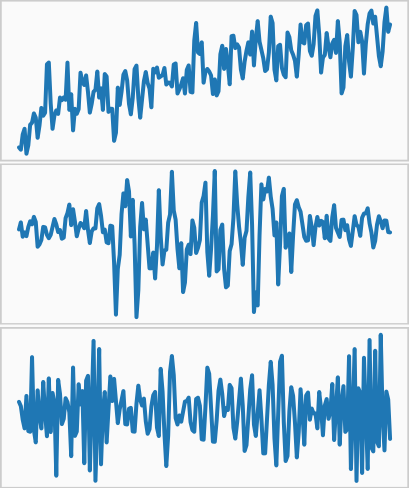
- Time series shows some linear trend
- Expectation is not constant over time

Heteroskedasticity

- Time series shows irregular changes
- Variance is not constant over time

Covariance

- Time series is unevenly spaced
- Covariance across different times is not constant



Definition of stationarity

We want the expectation, the variance and the covariance to be constant over time.

1. First a strong definition

Definition (strict stationarity)
A stochastic process $(Y_t)_{t \in \mathbb{Z}}$ is **strictly stationary** if for all T the joint distribution $(Y_{s+1}, \ldots, Y_{s+T})$ does not depend on s.

2. Then a weaker definition that is more realistic

Definition (weak stationarity)
A stochastic process $(Y_t)_{t \in \mathbb{Z}}$ is **(weakly) stationary** if

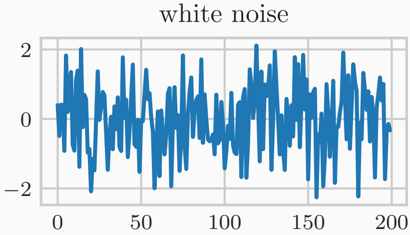
- The expectation is constant over time: $\mathbb{E}\left(Y_t\right) = \mu$
- The covariance only depends on time lag $|t-s|$: $\mathrm{Cov}(Y_t, Y_s) = \mathrm{Cov}(Y_{t+T}, Y_{s+T})$

White noise process

Definition (white noise process)
A process is a **white noise process** with mean μ and variance σ^2 if $\mathbb{E}\left(Y_t\right) = \mu$ for all $t \in \mathbb{Z}$ and

$$\operatorname{Cov}\left(Y_t, Y_s\right) = \begin{cases} 0 & \text{if } t \neq s \\ \sigma^2 & \text{if } t = s \end{cases}$$

- A white noise process is **stationary**

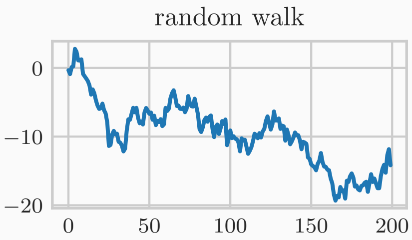


Random walk process

Definition (random walk process)
Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a white noise process. A **random walk** is defined by

$$Y_t = Y_{t-1} + \varepsilon_t$$

- A random walk process is **not stationary**



The autocovariance function

- When $(Y_t)_{t\in\mathbb{Z}}$ is **stationary**, to characterize the covariance between Y_t and the h-th lag Y_{t-h}, we define the **autocovariance function**

$$\begin{aligned}\gamma(h) &= \mathrm{Cov}\left(Y_{t-h}, Y_t\right) \\ &= \mathbb{E}\left((Y_{t-h}-\mu)(Y_t-\mu)\right)\end{aligned}$$

- $\gamma(0) \geqslant 0$ ($\gamma(0)$ is a variance)
- γ is symmetric: $\gamma(-h) = \gamma(h)$ (from stationarity)
- $|\gamma(h)| \leqslant \gamma(0)$ (from Cauchy–Schwarz)

The autocorrelation function (ACF)

The **autocorrelation function** (ACF) is just a rescaling of the autocovariance function so as to have $\rho(0) = 1$

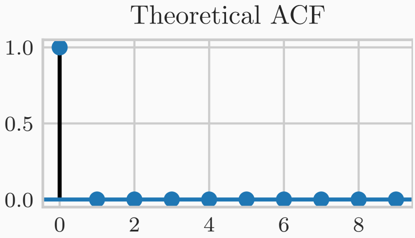
$$\begin{aligned} \rho(h) &= \frac{\gamma(h)}{\gamma(0)} \\ &= \text{Cor}\left(Y_{t-h}, Y_t\right) \qquad \text{(because } \gamma(0) = \text{Var}\left(Y_t\right)\text{)} \end{aligned}$$

- The ACF is also symmetric: $\rho(-h) = \rho(h)$
- Correlations are between -1 and 1 so: $-1 \leqslant \rho(h) \leqslant 1$

Example: ACF of white noise

- Suppose that $(Y_t)_{t\in\mathbb{Z}}$ is a white noise process:
  - The Y_t's are uncorrelated
  - $\mathbb{E}(Y_t) = \mu$ and $\mathrm{Var}(Y_t) = \sigma^2$
- The autocorrelation function is then

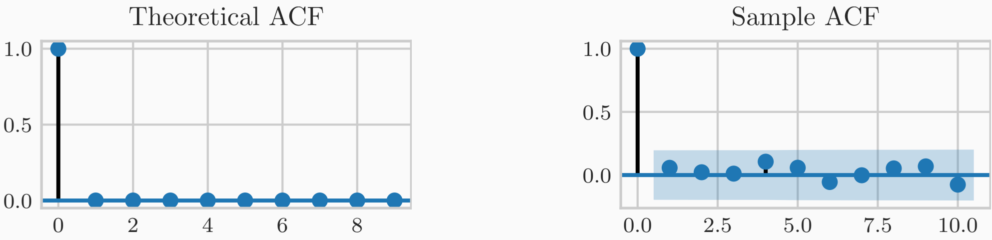
$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ 0 & \text{otherwise.} \end{cases}$$



Estimating the ACF: sample ACF

- ACF is theoretical; we need to estimate it
- Instead of correlations we use empirical correlations
- Let $y_1, y_2, \ldots, y_n$ be observations of $Y_1, Y_2, \ldots, Y_n$

$$r(h) = \frac{\sum_{i=1}^{n-h} (y_{i+h} - \overline{y})(y_i - \overline{y})}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

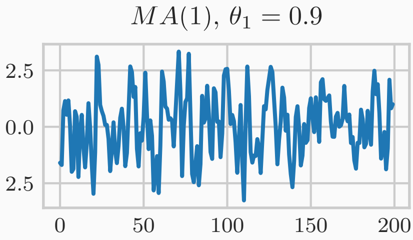


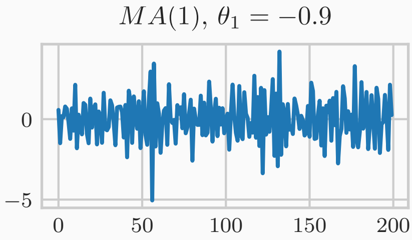
First order moving average process $MA(1)$

Idea: Current value Y_t is a linear combination of a previous error and a current error

Definition ($MA(1)$ model)
Let $(\varepsilon_i)_{t\in\mathbb{Z}}$ a (μ, σ^2)-white noise. The **first order moving average** process is defined by

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$





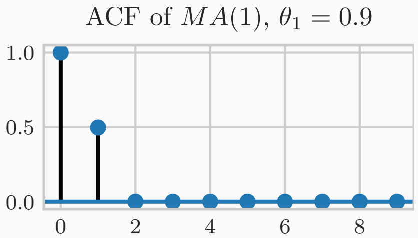
Properties of $MA(1)$

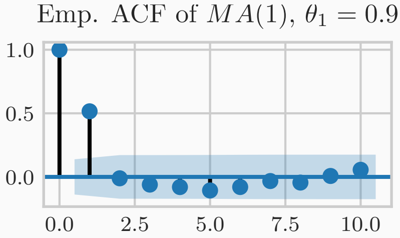
$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

- $MA(1)$ process is **stationary**
- $\mathbb{E}\left(Y_t\right) = (1 + \theta_1)\mu$
- $\mathrm{Var}\left(Y_t\right) = (1 + \theta_1^2)\sigma^2$
- ACF is

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ \frac{\theta_1}{1+\theta_1^2} & \text{if } h = 1 \\ 0 & \text{if } h > 1 \end{cases}$$

cuts off after lag 1





Idea: Current value Y_t is a linear combination of q past perturbations plus current perturbation

Definition ($MA(q)$ model)
Let $(\varepsilon_i)_{t\in\mathbb{Z}}$ a (μ,σ^2)-white noise. The **moving average** process of order q is defined by

$$Y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$$

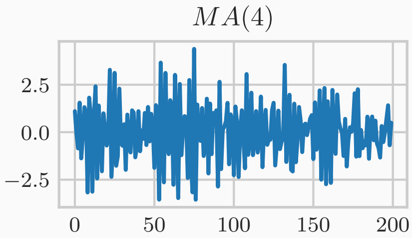
Some properties

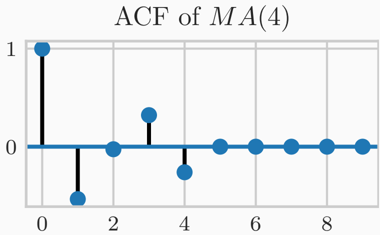
- $MA(q)$ is **stationary**
- $\mathbb{E}(Y_t) = \mu\sum_{i=0}^{q}\theta_i, \quad \theta_0 = 1$
- $\mathrm{Var}\,(Y_t) = \sigma^2\sum_{i=0}^{q}\theta_i^2 \quad \theta_0 = 1$

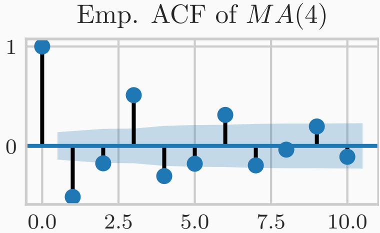
The ACF of $MA(q)$ processes

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ \frac{\sum_{i=0}^{q-h} \theta_i \theta_{i+h}}{\sum_{i=0}^{q} \theta_i^2} & \text{if } 1 \leqslant h \leqslant q \\ 0 & \text{if } h > q \end{cases}$$

Cuts off at time lag q







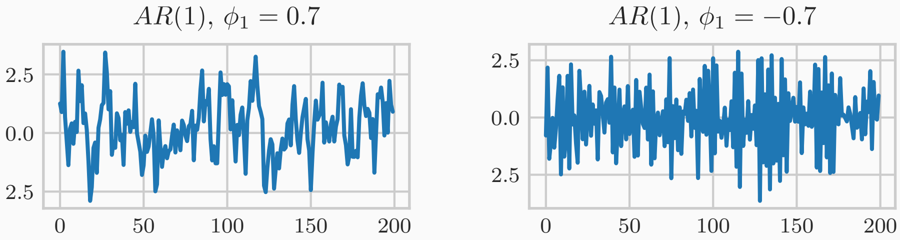
Autoregressive process

- Natural idea: Use the lagged values $Y_{t-1}, Y_{t-2}, \ldots$ to forecast Y_t
- An **autoregressive process** is a linear regression of Y_t against lagged values
- The number of regressors is called the order of the autoregressive process
- For example
    - In a **first order autoregressive process**, Y_t is regressed against Y_{t-1}
    - In a **autoregressive process of order** p, Y_t is regressed against $Y_{t-1}, Y_{t-2}, \ldots Y_{t-p}$

The first order autoregressive model $AR(1)$

Definition ($AR(1)$ model)
Let $(\varepsilon_i)_{t\in\mathbb{Z}}$ a centered white noise ($\mathbb{E}(\varepsilon_i) = 0$) and $(Y_t)_{t\in\mathbb{Z}}$ a random process such that $\mathbb{E}(Y_t) = 0$. It is a **first order autoregressive process** if we have

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

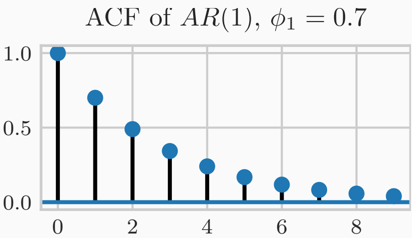


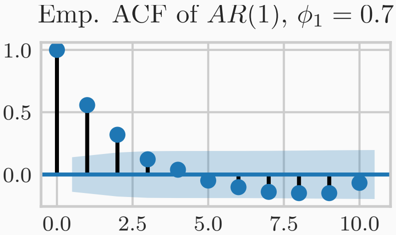
Properties of $AR(1)$

$AR(1)$ model

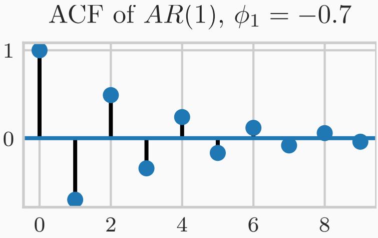
$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

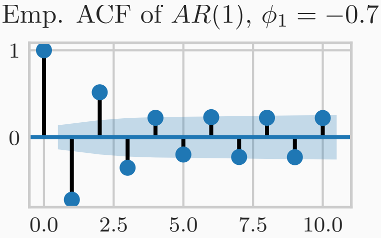
- Stationary iff $|\phi_1| < 1$
- ACF is $\rho(h) = \phi_1^h$





$AR(1)$ with $\phi_1 < 0$





The $AR(p)$ model

Definition
Let $(\varepsilon_i)_{t\in\mathbb{Z}}$ a white noise. The **autoregressive model of order** p is defined by

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

Some properties

- Not always a stationary process!
- Linear combination of p lagged values plus some noise

Backshift operator

- Let us introduce the backshift operator B

$$
\begin{aligned}
BY_t &= Y_{t-1} \\
B^k Y_t &= Y_{t-k}
\end{aligned}
$$

- $AR(p)$ can be rewritten

$$
\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \\
Y_t &= \phi_1 BY_t + \cdots + \phi_p B^p Y_t + \varepsilon_t \\
(1 - \phi_1 B - \cdots - \phi_p B^p) Y_t &= \varepsilon_t \\
\Phi(B) Y_t &= \varepsilon_t
\end{aligned}
$$

$AR(p)$ with backshift operator

Definition
Let $(\varepsilon_i)_{t\in\mathbb{Z}}$ a white noise. The autoregressive model of order p is defined by

$$\Phi(B)Y_t = \varepsilon_t$$

with Φ a **polynomial of degree** p

- Properties of $AR(p)$ process depend on the location of (complex) roots of polynomial Φ

Stationarity condition

- Unlike $MA(q)$ processes, $AR(p)$ processes are not automatically stationary
- For some $\phi_1, \phi_2, \ldots, \phi_p$ the corresponding $AR(p)$ process is not stationary
- $AR(p)$ is stationary if roots of Φ lies outside the **unit disc**

The $ARMA(p, q)$ model

- The $ARMA(p, q)$ model combines an $AR(p)$ and an $MA(q)$ model

Definition
Let $(\varepsilon_i)_{t\in\mathbb{Z}}$ a white noise. The autoregressive moving average model of order p and q is defined by

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- With the backshift operator B we have

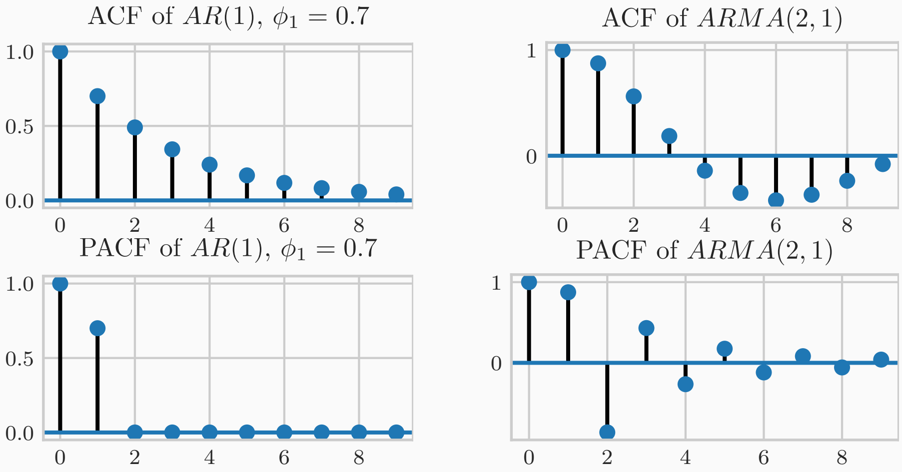
$$\Phi(B) Y_t = \Theta(B) \varepsilon_t$$

  with Φ a polynomial of order p and Θ a polynomial of order q
- $ARMA(p, q)$ is stationary if underlying $AR(p)$ is *i.e.* if roots of Φ lies outside the **unit disc**

Partial autocorrelation function (PACF)

- ACF is the correlation of Y_t and Y_{t-h}
  - ACF can make the difference between $MA(q)$ and $AR(p)/ARMA(p,q)$
  - but ACF looks the same for $AR(p)$ and $ARMA(q)$
- PACF is able to distinguish between $AR(p)$ and $ARMA(q)$
  - Roughly speaking partial correlation between Y_t and Y_{t-h} is the correlation when linear dependance from $Y_{t-1}, \ldots, Y_{t-h+1}$ is removed
  - Correlation of $Y_t - \widehat{Y}_t$ and $Y_{t-h} - \widehat{Y}_{t-h}$ where $\widehat{Y}_t$ and $\widehat{Y}_{t-h}$ are Y_t and Y_{t-h} linearly regressed over $Y_{t-1}, \ldots, Y_{t-h+1}$
- If $(Y_t)_{t\in\mathbb{Z}}$ is normally distributed partial correlation reduces to $\mathrm{Cor}\left(Y_t, Y_{t-h} \mid Y_{t-1}, \ldots, Y_{t-u+1}\right)$

ACF and PACF on $AR(p)$ and $ARMA(p, q)$



Shape of ACF and PACF

- Shape of ACF/PACF for MA/AR/ARMA models

	$MA(q)$	$AR(p)$	$ARMA(p,q)$
ACF	zero for $h > q$	decays	decays
PACF	decays	zero for $h > p$	decays

- $ARMA(p,q)$ always decays
- ACF is zero after q for $MA(q)$
- PACF is zero after p for $ARMA(q)$

Forecasting

What is the **best one-step ahead linear prediction** from n previous values we can make?

- Suppose $(Y_t)_{t\in\mathbb{Z}}$ is stationary and $\mathbb{E}(Y_0) = 0$.
- Denote Y_{n+1}^n the best linear prediction of Y_{n+1} w.r.t MSE from n previous values
- We have $Y_{n+1}^n = \phi_{n1} Y_n + \cdots + \phi_{nn} Y_1$
- One can show that the coefficients ϕ_{ni} verify $\Gamma_n \phi_n = \gamma_n$
  where $\gamma_n = (\gamma(1), \gamma(2), \ldots, \gamma(n))^T$, $\phi_n = (\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn})$ and

$$\Gamma_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(n-2) \\ \vdots & & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \ldots & \gamma(0) \end{pmatrix}$$

Forecasting an $AR(p)$

- Suppose $(Y_t)_{t\in\mathbb{Z}}$ is an $AR(p)$ process:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

- We can show that

$$Y_{n+1}^n = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p}, \qquad \text{for } n \geqslant p$$

- No need to solve the linear equations above
- Prediction Y_{n+1}^n is what we expect

How to make a time series stationary?

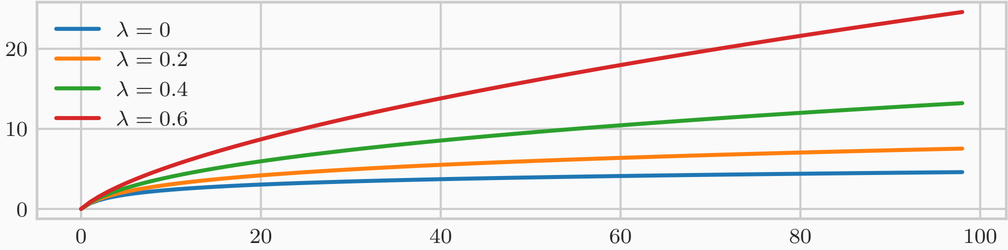
- We have well understood models for stationary time series: *AR*, *MA* and *ARMA*
- What if the time series is not stationary?

- Use transformations to equalize variability
- Use integrated models
- Decompose the time series

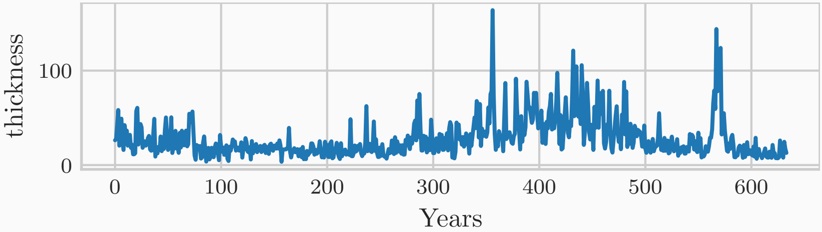
Box–Cox transformation

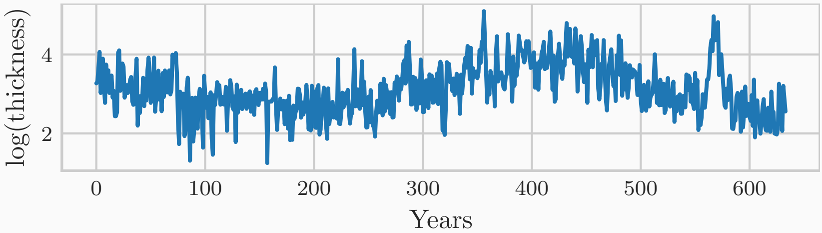
- A well known transformation is the Box–Cox transformation

$$Z_t = \begin{cases} (Y_t^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0 \\ \log(Y_t) & \text{if } \lambda = 0 \end{cases}$$



Examples: log





Integrated models

- Suppose that Y_t has a linear trend with Z_t stationary

$$Y_t = \beta_0 + \beta_1 t + Z_t$$

- Differencing at time t and $t-1$

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 t + Z_t \\ Y_{t-1} &= \beta_0 + \beta_1 (t-1) + Z_{t-1} \\ (Y_t - Y_{t-1}) &= \beta_1 + (Z_t - Z_{t-1}) \end{aligned}$$

- $(Y_t - Y_{t-1})$ is now **stationary**!

Integrated models

- Suppose that Y_t has a quadratic trend with Z_t stationary

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + Z_t$$

- Differencing at time t and $t-1$

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 t + \beta_2 t^2 + Z_t \\ Y_{t-1} &= \beta_0 + \beta_1 (t-1) + \beta_2 (t-1)^2 + Z_{t-1} \\ (Y_t - Y_{t-1}) &= \beta_1 - \beta_2^2 + 2\beta_2 t + (Z_t - Z_{t-1}) \end{aligned}$$

- $(Y_t - Y_{t-1})$ has now a **linear trend**!

Differencing operator

- Introducing the differencing operator ∇

$$\begin{aligned} \nabla Y_t &= Y_t - Y_{t-1} \\ &= (1-B)Y_t \end{aligned}$$

- The differencing operator can be composed

$$\nabla^d = (1-B)^d$$

- For example

$$\nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

ARIMA model

- An ARIMA model is just an ARMA on the d-difference

Definition (Integrated ARMA)
A process Y_t is an $ARIMA(p, d, q)$ process if

$$\nabla^d Y_t = (1 - B)^d Y_t$$

is $ARMA(p, q)$

- In short it can be written

$$\Phi(B)(1 - B)^d Y_t = \Theta(B)\varepsilon_t$$

Decompose the time series

- If Y_t is not stationary, we decompose it at follows

$$Y_t = m_t + s_t + Z_t$$

  where

  - m_t is a slowly changing function called the **trend**
  - s_t is a function with known period called the **seasonal component**
  - Z_t is a **stationary** time series
- Additive decomposition