Statistical Theory and Modeling (ST2601)

Lecture 12 - Autocorrelation and autoregressive models for time series

Mattias Villani

Department of Statistics
Stockholm University

mattiasvillani.com @matvil @matvil mattiasvillani

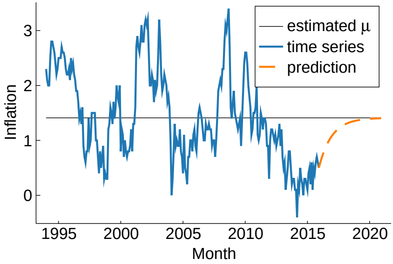
Overview

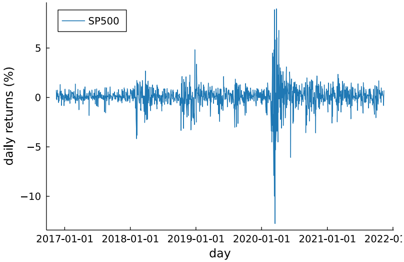
- **Time series**
- **Autocorrelation function**
- **Autoregressive models**

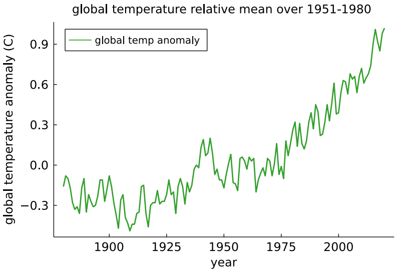
Time series data are special

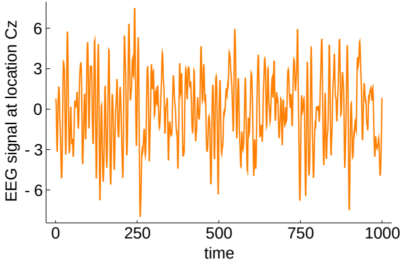
- **Time series**: data measured over time y_t, $t = 1, 2, ...$
- **Cross-sectional** data measured over time. **Time series regression**.
- Time series are special:
  - **Trend**, **seasonality**.
  - **Dependent observations** over time. Yesterday's value y_{t-1} can predict today's value y_t. **Autocorrelation**.
  - Sometimes the observations are **not equi-distant in time**.
- **Monte Carlo methods** like **MCMC** and **HMC** (see Bayes course!) give dependent simulated draws. Time series methods useful for measuring efficiency and diagnosing convergence problems.

Example time series

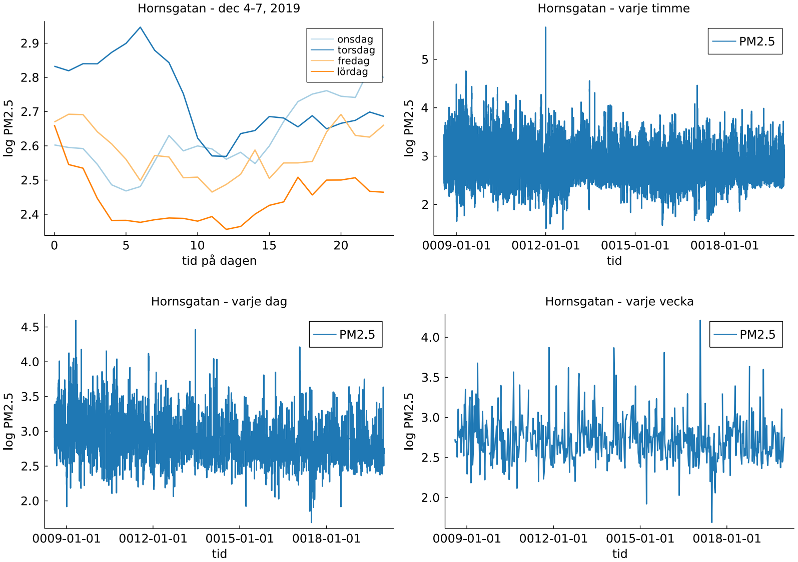








Particle matter (PM2.5) at the street Hornsgatan



Repetition - sample correlation

- **Covariance** between two variables

$$s_{xy} = \mathrm{cov}(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

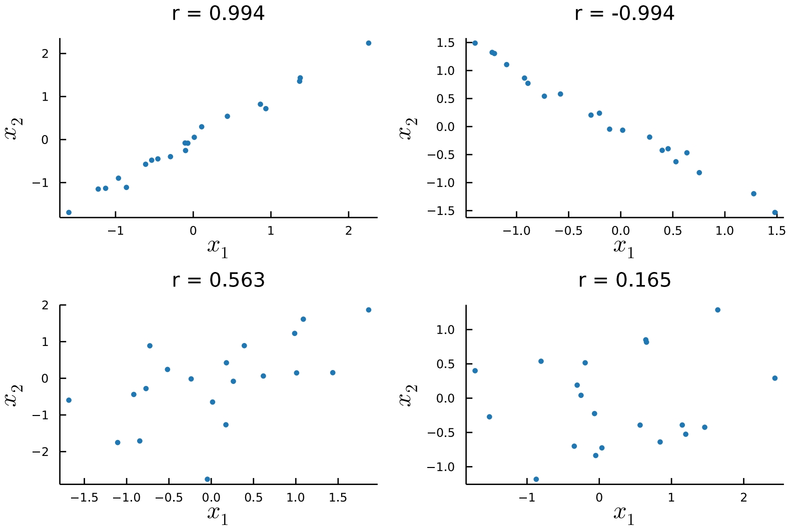
- **Correlation** between two variables:

$$r_{xy} = \mathrm{corr}(x, y) = \frac{s_{xy}}{s_x s_y}$$

where

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}$$

Repetition - sample correlation



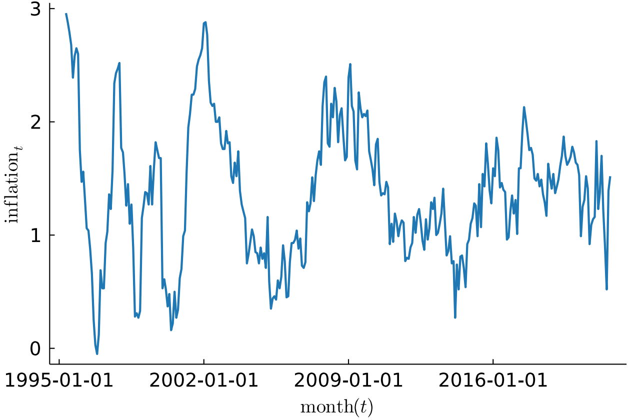
Autocorrelation of order 1

- Observations in a **time series** y_t are often dependent/**correlated**.
- **Autocorrelation** of **order 1**:

$$r_1 = \mathrm{corr}(y_t, y_{t-1})$$

- "Correlation between today's and yesterday's value."
- "Correlation between this month and the previous month."
- "First lag": y_{t-1}.

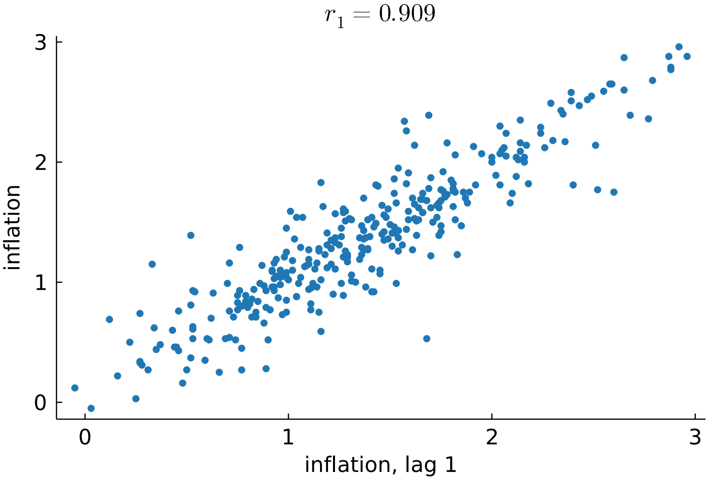
Inflation


3
2
1
0
inflation_t
1995-01-01
2002-01-01
2009-01-01
2016-01-01
$\text{month}(t)$

Lagged variables - inflation

	A	B	C	D	E	F
1	Månad	Inflation(t)	Inflation(t-1)	Inflation(t-2)	Inflation(t-3)	Inflation(t-4)
2	1995-05-01	2.96				
3	1995-06-01	2.88	2.96			
4	1995-07-01	2.79	2.88	2.96		
5	1995-08-01	2.68	2.79	2.88	2.96	
6	1995-09-01	2.39	2.68	2.79	2.88	2.96
7	1995-10-01	2.58	2.39	2.68	2.79	2.88
8	1995-11-01	2.65	2.58	2.39	2.68	2.79
9	1995-12-01	2.6	2.65	2.58	2.39	2.68
10	1996-01-01	1.75	2.6	2.65	2.58	2.39
11	1996-02-01	1.47	1.75	2.6	2.65	2.58
12	1996-03-01	1.56	1.47	1.75	2.6	2.65
13	1996-04-01	1.31	1.56	1.47	1.75	2.6
14	1996-05-01	1.06	1.31	1.56	1.47	1.75
15	1996-06-01	1.04	1.06	1.31	1.56	1.47
16	1996-07-01	0.88	1.04	1.06	1.31	1.56
17	1996-08-01	0.66	0.88	1.04	1.06	1.31
18	1996-09-01	0.25	0.66	0.88	1.04	1.06
19	1996-10-01	0.03	0.25	0.66	0.88	1.04
20	1996-11-01	-0.05	0.03	0.25	0.66	0.88
21	1996-12-01	0.12	-0.05	0.03	0.25	0.66
22	1997-01-01	0.69	0.12	-0.05	0.03	0.25

Inflation - autocorrelation lag 1



Autocorrelation of order 2

- **Autocorrelation** of **order 2**:

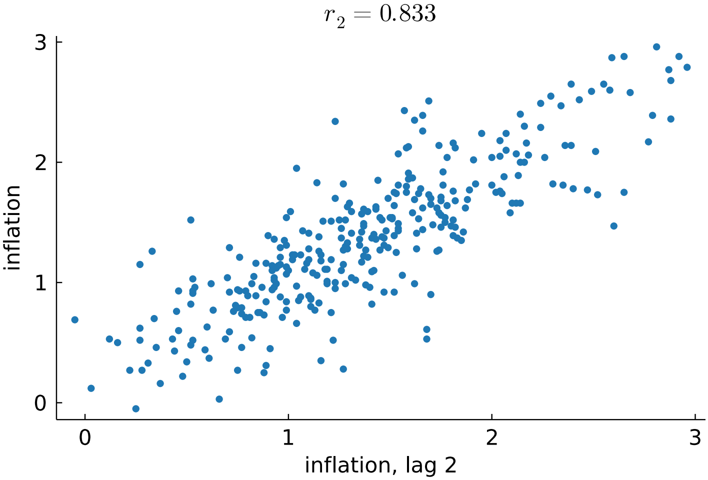
$$r_2 = \mathrm{corr}(y_t, y_{t-2})$$

- "Correlation between today's value and the value two days back."

- "Correlation between this month's value and the value two months back."

- "Second lag": y_{t-2}.

Autocorrelation lag 2



Autocorrelation function

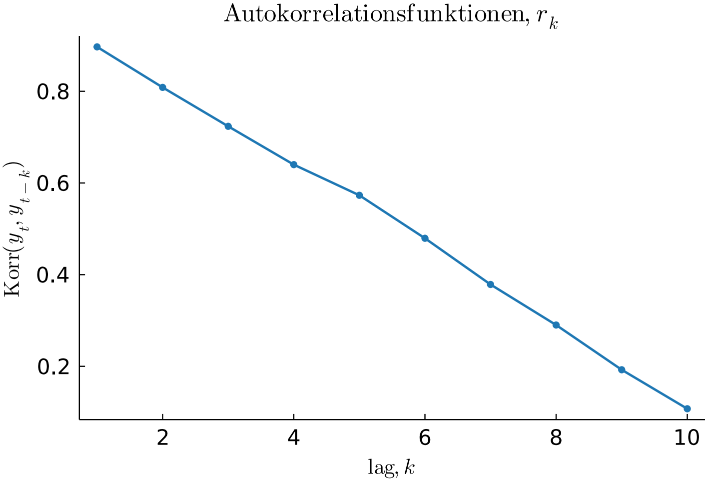
- **Autocorrelation** of **order** k

$$r_k = \operatorname{corr}(y_t, y_{t-k})$$

- "Correlation between this month's value and k months back in time.".

- **Autocorrelation function** (**ACF**) is r_k as a function of the time delay, k.

Inflation - autokorrelationsfunktion



Autoregressive models

- Autoregressive model of order 1 ($\mathrm{AR}(1)$)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

- $\mathrm{AR}(1)$ is a **regression with y_{t-1} as explanatory variable**!
- Fit with the **least squares** method

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$$

- Autoregressiv modell av ordning p ($\mathrm{AR}(p)$)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

- $\mathrm{AR}(p)$ is a **multiple regression** with the p explanatory variables $y_{t-1}, ..., y_{t-p}$.

AR(1) for inflation - R

```
> library(SUdatasets)
> arimafit = arima(swedinfl$KPIF, order = c(1,0,0))
> arima_coef_summary(arimafit)

Parameter estimates
--------------------------------------------------------------
     Estimate Std. Error z-ratio Pr(>|z|)   2.5 %  97.5 %
ar1   0.91801   0.022383 41.0135        0 0.87414 0.96188
mean  1.43624   0.165006  8.7042        0 1.11282 1.75965
>
> |
```

AR(4) for inflation - R

```
> library(SUdatasets)
> arimafit = arima(swedinfl$KPIF, order = c(4,0,0))
> arima_coef_summary(arimafit)

Parameter estimates
-------------------------------------------------------------
        Estimate Std. Error   z-ratio Pr(>|z|)     2.5 %   97.5 %
ar1    0.8900015   0.055640 15.995742  0.00000  0.780947 0.999056
ar2    0.0586250   0.075101  0.780619  0.43503 -0.088572 0.205822
ar3    0.0062025   0.076370  0.081216  0.93527 -0.143483 0.155888
ar4   -0.0405666   0.057249 -0.708605  0.47857 -0.152774 0.071641
mean   1.4334525   0.158225  9.059583  0.00000  1.123331 1.743573
```

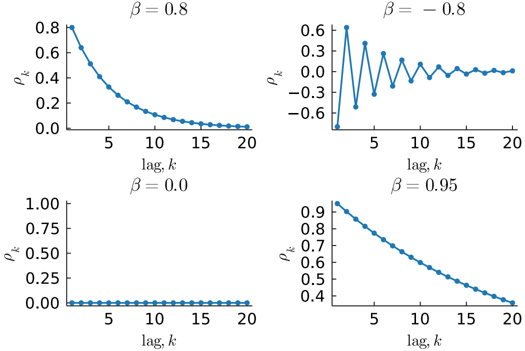
Autocorrelation function AR(1)

- $\mathrm{AR}(1)$

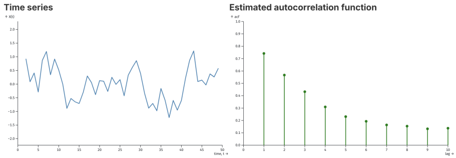
$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

- Population **autocorrelation function** (**ACF**) for $\mathrm{AR}(1)$

$$\rho_k = \beta^k, \text{ for } k = 1, 2, \ldots$$

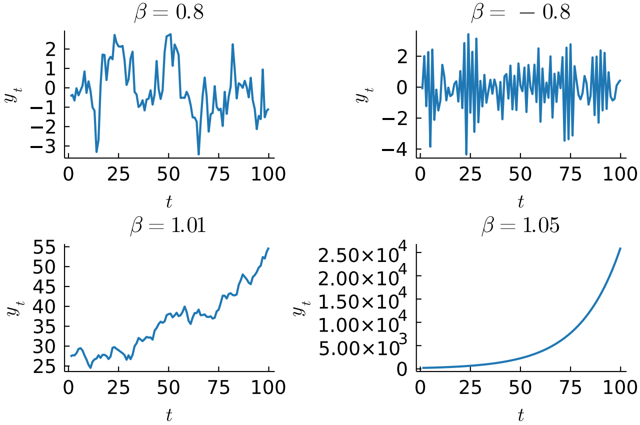


Simulate AR(p) and estimate ACF



Autoregressive models - stationarity

- $\mathrm{AR}(1)$ is a **stationary** (non-explosive) model if $-1 < \beta_1 < 1$.
- Some simulated $\mathrm{AR}(1)$ time series:



Prediction with an AR(1) model

- Fitted AR(1)-model

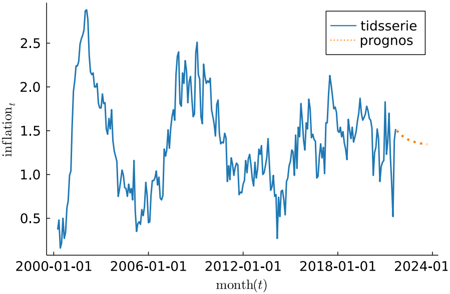
$$y_t = \hat{\beta}_0 + \hat{\beta}_1 \cdot y_{t-1}$$

- At time T, **prediction for next month $T+1$**

$$\hat{y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 \cdot y_T$$

- **Prediction for $T+2$**

$$\hat{y}_{T+2} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \hat{y}_{T+1}$$



Prediction with an AR(2) model

- Fitted AR(2)-model

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 \cdot y_{t-1} + \hat{\beta}_2 \cdot y_{t-2}$$

- At time T, **prediction for next month** $T+1$

$$\hat{y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 \cdot y_T + \hat{\beta}_2 \cdot y_{T-1}$$

- **Prediction for** $T+2$

$$\hat{y}_{T+2} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \hat{y}_{T+1} + \hat{\beta}_2 \cdot y_T$$

- **Prediction for** $T+3$

$$\hat{y}_{T+3} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \hat{y}_{T+2} + \hat{\beta}_2 \cdot \hat{y}_{T+1}$$