Time Series Analysis

Lecture 2: Exploratory analysis and Time Series Regression

Tohid Ardeshiri

Linköping University
Division of Statistics and Machine Learning

September 4, 2019

Summary of Lecture 1

- Time series
  - White noise
  - Random walk
  - Moving average filter

- Autocovariance and autocorrelation functions:

$$\gamma(s,t) = \text{cov}(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)]$$

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$$

Autocovariance and ACF

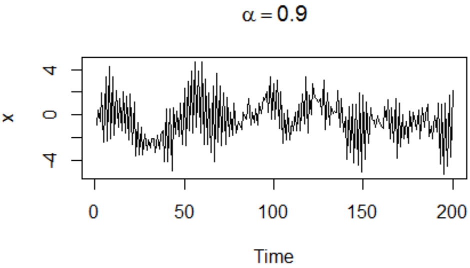
Examples: Autocovariance and ACF of **on whiteboard**

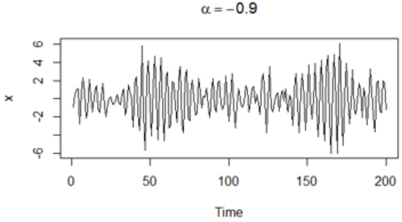
- White noise ✓
- Random walk $x_t = \delta t + \sum_{j=1}^{t} w_j$
- Moving average $x_t = 0.2w_{t-1} + 0.5w_t + 0.2w_{t+1}$

Autocovariance

- Intuition:

$$x_t = \phi x_{t-1} + w_t$$





- when $x_0 = 0$ and $w_t \sim wn(0, 1)$:

$$\text{cov}(x_t, x_{t-1}) = \phi$$

Autocovariance (read at home)

$$x_t = \phi x_{t-1} + w_t$$

Mean function:

$Ex_t = \phi Ex_{t-1} + Ew_t = \phi Ex_{t-1} = \phi(\phi Ex_{t-2}) = \cdots = \phi^t Ex_0$

for $Ex_0 = 0$, $Ex_t = 0$ for all t.

Variance $\mathrm{var}(x_t)$ when $Ex_0 = 0$ and w_t is uncorrelated with x_0 for all t:

$\mathrm{var}(x_t) = E\{(x_t - 0)^2\} = E\{\phi^2 x_{t-1}^2 + 2\phi x_{t-1} w_t + w_t^2\} =$

$\phi^2 \mathrm{var}(x_{t-1}) + 2\phi\, \mathrm{cov}(x_{t-1}, w_t) + \mathrm{var}(w_t) = \phi^2 \mathrm{var}(x_{t-1}) + \mathrm{var}(w_t) =$

$\phi^2 \mathrm{var}(x_{t-1}) + \sigma_w^2 = \phi^2(\phi^2 \mathrm{var}(x_{t-2}) + \sigma_w^2) + \sigma_w^2 =$

$\phi^{2t} \mathrm{var}(x_0) + \sigma_w^2 \sum_{k=0}^{t-1}(\phi^{2k}) = \phi^{2t} \mathrm{var}(x_0) + \frac{\sigma_w^2(1-\phi^{2t})}{1-\phi^2}$

When $\mathrm{var}(x_0) = \frac{\sigma_w^2}{1-\phi^2}$ then $\mathrm{var}(x_t) = \frac{\sigma_w^2}{1-\phi^2}$ and time independent.

Autocovariance (read at home)

$$x_t = \phi x_{t-1} + w_t$$

$$x_t = \phi(\phi x_{t-2} + w_{t-1}) + w_t = \cdots = \phi^h x_{t-h} + \sum_{j=0}^{h-1} \phi^j w_{t-j}$$

$$\gamma(x_t, x_{t-h}) = \text{cov}(x_t, x_{t-h}) = E(x_t x_{t-h}) =$$
$$E\{(\phi^h x_{t-h} + \sum_{j=0}^{h-1} \phi^j w_{t-j}) x_{t-h}\} = \phi^h \,\text{var}(x_{t-h}) = \frac{\phi^h \sigma_w^2}{1-\phi^2}$$

Hence,

$$\gamma(h) = \frac{\phi^h \sigma_w^2}{1 - \phi^2}$$

Also,

$$\rho(h) = \phi^h$$

Stationarity

Fact: sometimes $\rho(s,t)$ pends on lag $|s-t|$ only

Time series is strictly stationary if distributions of $\{x_{t1},...x_{tn}\}$ and $\{x_{t1+h},...x_{tn+h}\}$ are identical for any $\{t_1,...t_n\}$ and all lags $h = 0, \pm1, \pm2, ...$
$$P(x_{t1} \leq c_1, ...x_{tn} \leq c_n) = P(x_{t1+h} \leq c_1, ...x_{tn+h} \leq c_n)$$

Note: This means

- Mean function $\mu_t = Ex_t =$ const.
- Autocovariance $\gamma(t, t+h) =$ function only of lag h

Stationarity

Strict stationarity is often too strong!

- Time series x_t is weakly stationary (stationary) if
  - $Ex_t = const$
  - $\gamma(s,t) = \gamma(|s-t|)$
  - $\text{var}(x_t) < \infty$
- $\gamma(t, t+h) = \gamma(|t+h-t|) = \gamma(h)$
  - Autocovariance depends on lag only!
- Autocovariance for stationary process $\gamma(h) = \text{cov}(x_t, x_{t+h})$
- ACF for stationary process $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$

Stationarity

Properties of stationary process:

$$\gamma(h) = \gamma(-h) \qquad \rho(h) = \rho(-h)$$

$$|\gamma(h)| \leq \gamma(0) \qquad \rho(h) \leq 1, \rho(0) = 1$$

Reflect: Are these processes stationary?

- White noise
- Moving average, $x_t = 0.2w_{t-1} + 0.5w_t + 0.2w_{t+1}$
- Random walk, $x_t = \delta t + \sum_{j=1}^{t} w_j$

Sample autocovariance and ACF

Dependence measures for samples?

- Idea: replace mean and covariance with sample estimates

If x_t is stationary,

- Sample mean

$$Ex \approx \bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t$$

- Sample autocovariance function

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})$$

Sample autocovariance and ACF

Example: n=6, h=2

		X1	X2	X3	X4	X5	X6
X1	X2	X3	X4	X5	x6		

Sample autocorrelation function (sample ACF)

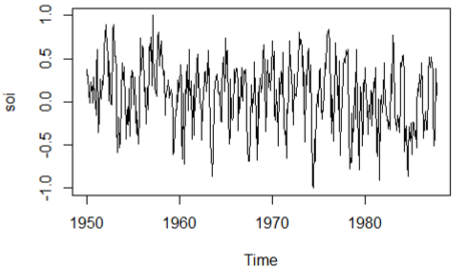
$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Sample ACF

In R: acf()

Example: southern oscillation index (SOI)

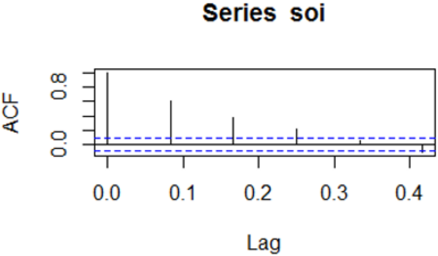
- rho=acf(soi, 5, type="correlation", plot=T)

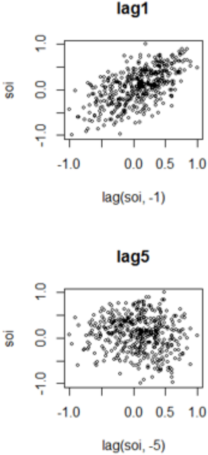



```
> print(rho)

Autocorrelations of series ‘soi’, by lag

0.0000 0.0833 0.1667 0.2500 0.3333 0.4167
 1.000  0.604  0.374  0.214  0.050 -0.107
```

Why is sample ACF '1' for h=0?

Sample ACF


Series soi
ACF
0.0
0.8
0.0
0.1
0.2
0.3
0.4
Lag


lag1
soi
-1.0
0.0
1.0
-1.0
0.0
0.5
1.0
lag(soi, -1)
lag5
soi
-1.0
0.0
1.0
-1.0
0.0
0.5
1.0
lag(soi, -5)

Sample ACF

What are these blue lines?

Theorem: Under weak conditions, if x_t is white noise and $n \to \infty$ then $\hat{\rho}(h)$ is approximately $N(0, \frac{1}{n})$

Consequence: If some $|\hat{\rho}(h)| > \frac{2}{\sqrt{n}}$ then the time series is not a white noise (with approximately 95 % confidence).

Typical modeling strategy:

- Fit a model
- Compute residuals
- Check ACF within $\pm\frac{2}{\sqrt{n}}$

Sample ACF vs theoretical

- Moving average $x_t = 0.2w_{t-1} + 0.5w_t + 0.2w_{t+1}$
  - $$ACF\gamma(h) = \begin{cases} 1 & h = 0 \\ 0.61 & h = 1 \\ 0.12 & h = 2 \\ 0 & other \end{cases}$$
  - n=10

```
Autocorrelations of series ‘y1’, by lag

     0      1      2      3      4      5 
 1.000  0.236 -0.399 -0.187 -0.008 -0.118 
```

  - n=1000

```
Autocorrelations of series ‘y1’, by lag

     0      1      2      3      4      5 
 1.000  0.609  0.129 -0.007  0.001  0.044 
```

Vector-valued time series

If $x_t = (x_{t1}, x_{t2}, \cdots, x_{tp})'$ is stationary,

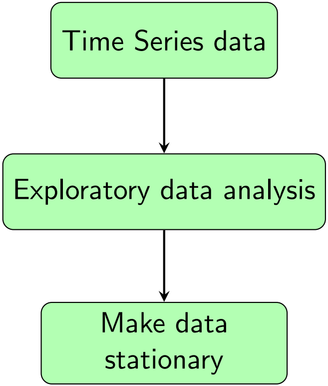
- mean vector is $\mu = E(x_t)$ and sample mean is its approximation

$$\mu = E(x_t) \approx \bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

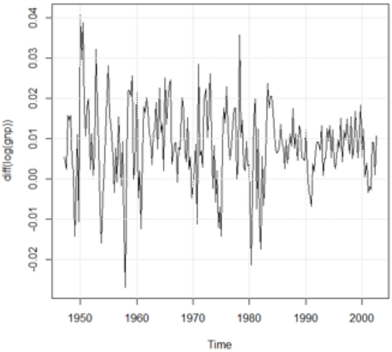
- Autocovariance function is $\Gamma(h) = E[(x_{t+h} - \mu)(x_t - \mu)']$ and sample autocovariance matrix

$$\hat{\Gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})'$$

Recap: time domain modeling



$$Y_t = \nabla(\log(X_t))$$



Stationarity

- Why do we need stationarity?
  - Sample ACF becomes consistent
  - ARIMA models require stationarity

- Tools
  - Detrending (trend removal)
  - Differencing
  - Transformations

whiteboard

- Introduce linear regression/least squares
- Trend removal, simple drift

Trend removal by regression

Regressing on covariates

Given x_t (dependent series) and $z_{t1}, ..., z_{t2}$ (independent series) we model

$$x_t = \beta_0 + \beta_1 z_{t1} + ... + \beta_q z_{tq} + w_t$$

where w_t is assumed white noise.

Note: w_t is seldom white noise in practice, used as a tool for detrending!

Trend removal by regression

Still a linear regression in β

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & z_{11} & \dots & z_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & \dots & z_{nq} \end{pmatrix}$$

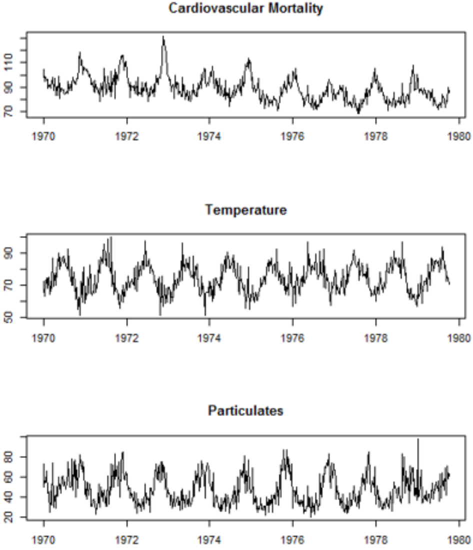
Least squares estimate is computed as

$$\hat{\beta} = (Z^T Z)^{-1} Z^T X$$

Trend removal

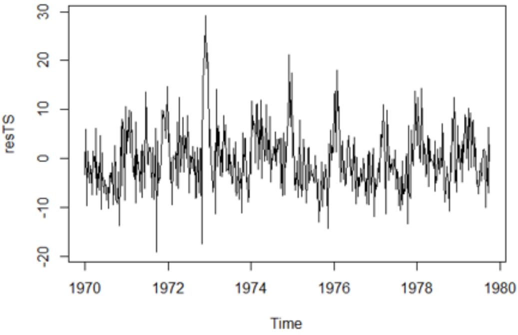
Example: Mortality

- x_t: Cardiovascular mortality
- z_{t1}: Temp (centered)
- z_{t2}: Temp (centered, squared)
- z_{t3}: Time
- z_{t4}: Levels of particles



Trend removal

- Residuals
  - Stationary?
  - Independent?
  - Some additional modeling of the residuals (ARIMA) can be done



Differencing

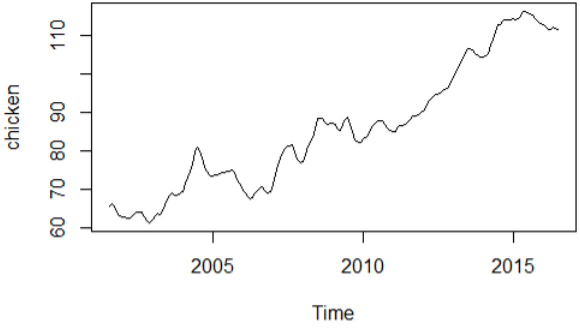
Assume $x_t = \mu_t + y_t$, y_t stationary

Differencing gives $z_t = \nabla x_t = x_t - x_{t-1}$

- Property 1: If $\mu_t = \alpha_0 + \alpha_1 t$ then z_t is stationary
- Property 2: If μ_t is random walk with a drift then z_t is stationary

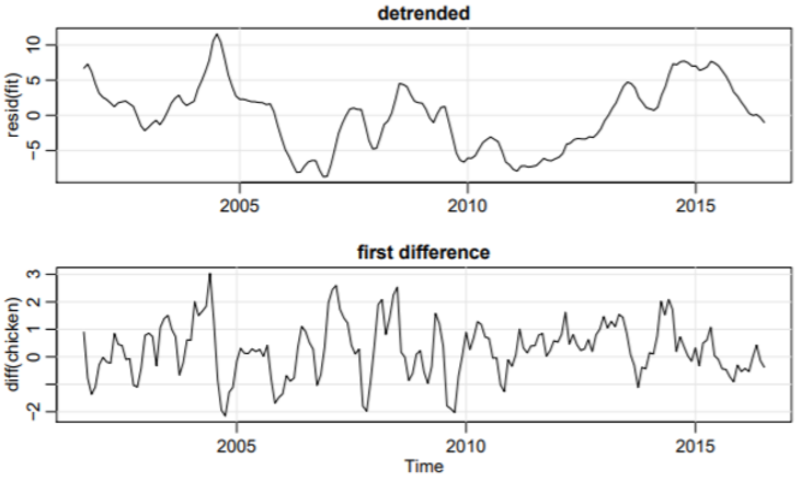
Example:

Chicken prices

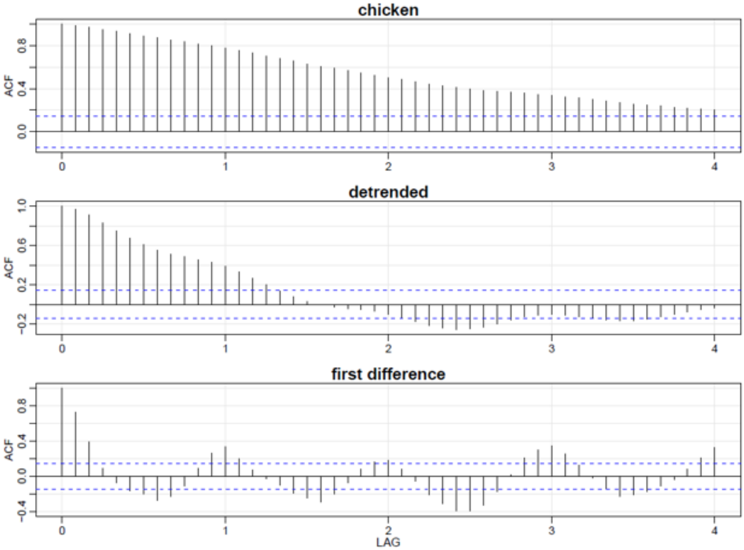


Differencing

Which looks **most random**? Other differences?



Differencing



Detrending vs differencing

- Differencing is more flexible than linear detrending
- Differencing does not require model estimation
- If trend is complex, detrending with a flexible (machine learning) model can be better
- Differencing does not give us the trend

Backshift operator

- Backshift operator $Bx_t = x_{t-1}$, Powers $B^k x_t = x_{t-k}$
- Forward-shift operator $B^{-1}x_t = x_{t+1}$
- Note $BB^{-1}x_t = x_t$ (i.e. $BB^{-1} = 1$)
- Differencing $\nabla x_t = (1 - B)x_t$
- Differences of order d: $\nabla^d = (1 - B)^d$
- Property: Operators can be manipulated as polynomials
- Example Check that $\nabla^2 x_t = x_t - 2x_{t-1} + x_{t-2}$
- Property: Differencing of order p can remove polynomial trend of order p

Transformations

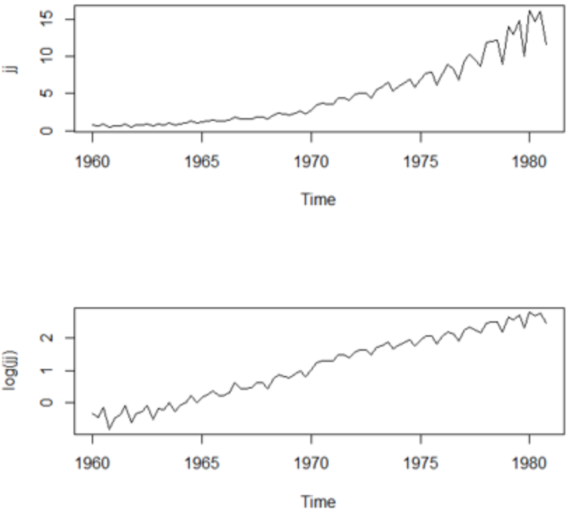
- Often used to stabilize variance
  - If for ex.var(x_t) $\neq$ var(x_s) then time series is non-stationary $\cdots$
- Sometimes makes data more similar to normal distr.

- Common transforms:
  - $z_t = \log(x_t)$
  - Power transformation

$$z_t = \begin{cases} \frac{(x_t^\lambda - 1)}{\lambda} & \lambda \neq 0 \\ \log(x_t) & \lambda = 0 \end{cases}$$

Transformations

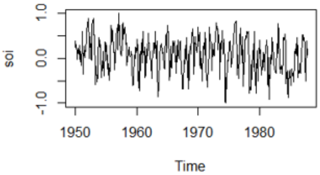
- Johnson & Johnson quarterly earnings



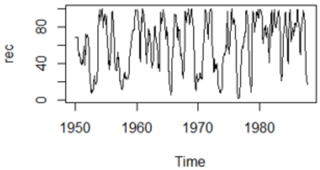
Scatterplots

- Plot x_t vs z_{t_i} or z_{t_i} vs z_{t_j}
- Exploratory tool: indicates which relationship to model

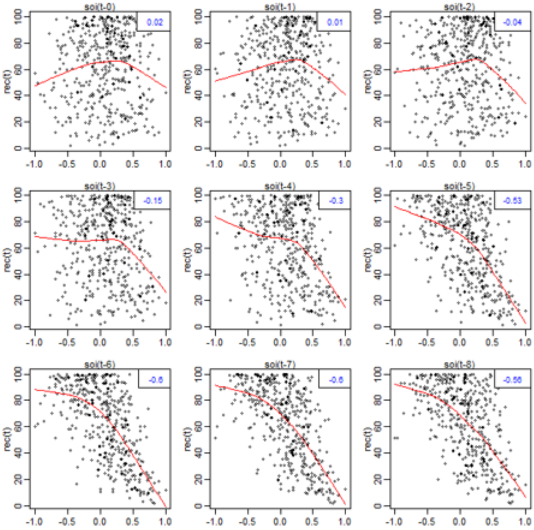
$$x_t = f(z_{t_1}, z_{t_2, \cdots, z_{t_q}}) + w_t$$



- Example: SOI and Recruitment



Scatterplots



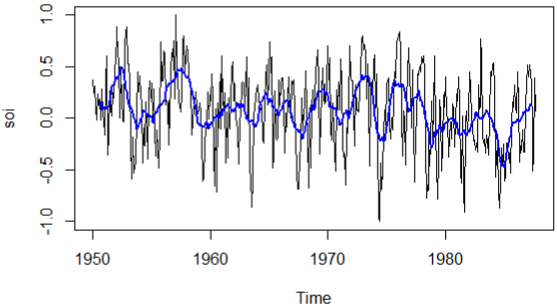
- Which relationships are nonlinear?
- Conclusion: include dummy variables $I(soi(t-j)>0)$ in the linear model

Smoothing

- Moving average smoother

$$m_t = \sum_{j=-k}^{j=k} a_j x_{t-j}$$

- Where $\sum_{j=-k}^{j=k} a_j = 1$ and $a_j = a_{-j} \geq 0$,
- Example: SOI data Disadvantage?



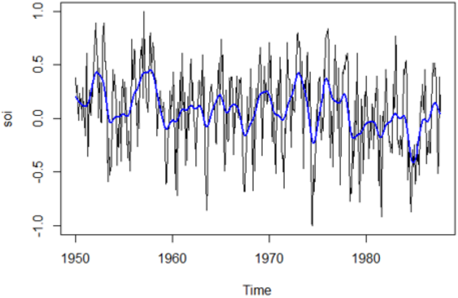
Smoothing

More flexible models?

- Splines
- Kernel smoothers
- Gaussian Process
- Neural networks
- ...

Welcome to ML courses!!

Example: kernel smoothers



Home reading

- Shumway and Stoffer, sections 1.4-1.6 and chapter 2
- TS functions: lag, ksmooth, lm, diff, lag1.plot, lag2.plot