Analysis of Time Series, L1

Rolf Larsson

Uppsala University

24 mars 2025

Course overview

- Shumway and Stoffer:
  Time Series Analysis and its Applications, with R examples, 5th ed.
  Springer 2025.
- Chapters (from 4th ed.):
  1. Characteristics of Time Series (L1-2)
  2. Time Series Regression and Explanatory Data Analysis (L2)
  3. ARIMA Models (L3-9)
  4. Spectral Analysis and Filtering (L10-12)
  5. Additional Time Domain Topics (L13-16)
  6. State-Space Models (L17-18)

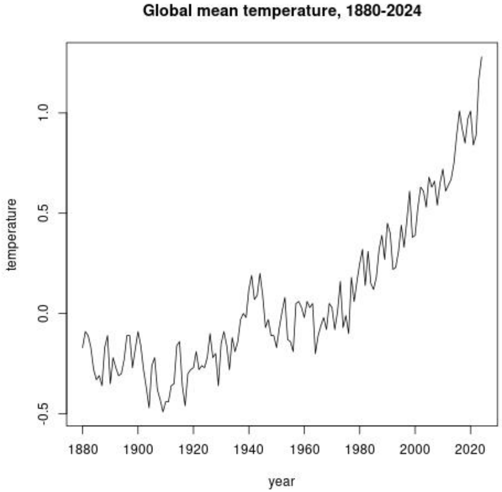
Course overview

- 25 meetings:
  - 18 theory lectures
  - 5 problem sessions
  - 2 project presentation sessions
- Written exam (with book and/or homemade notes).
- Two hand in assignments, not compulsory but give bonus points.
- Project: Analyse your own time series. (Compulsory.)
- Please check Studium for further information!
  (The file “kursinfo”, including a list of recommended exercises.)

Today

- 1.1-2: Introduction, examples
- 1.3: Statistical models
- 1.4: Measures of dependence
- Menti

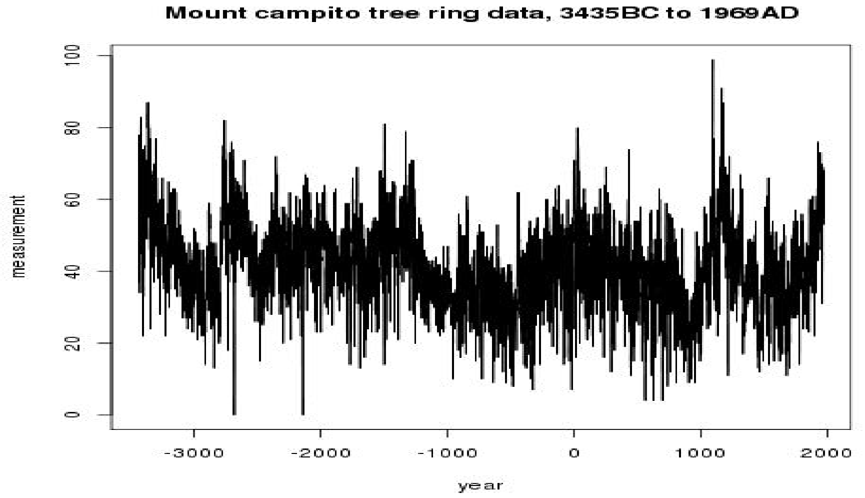
Introduction, examples



In R:

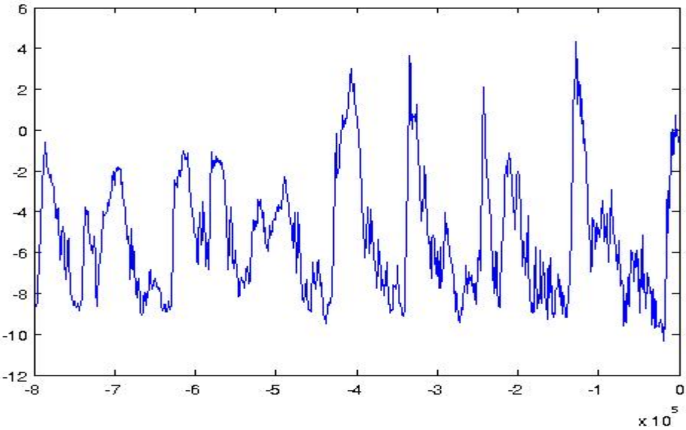
```
> year=read.table("temp.dat")$V1
> temp=read.table("temp.dat")$V2
> plot(year,temp,type='l',ylab='temperature',
main='Global mean temperature, 1880-2024')
```

Introduction, examples

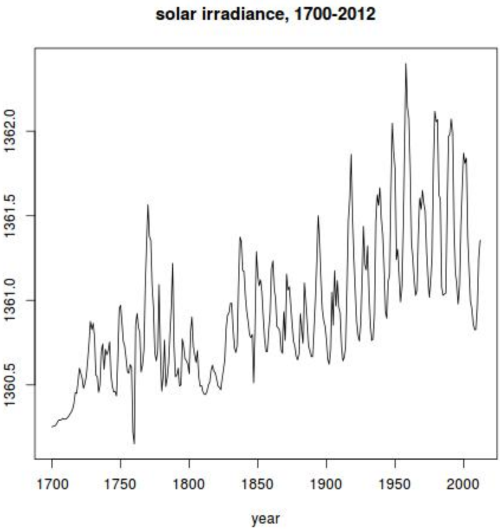


Introduction, examples

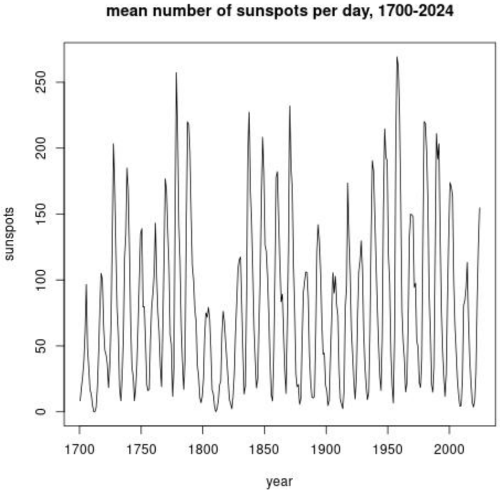
Antarctic ice core temperature proxies from about 800 000 BC to now.



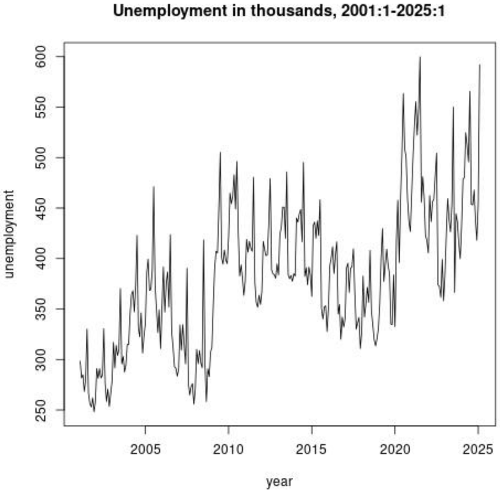
Introduction, examples



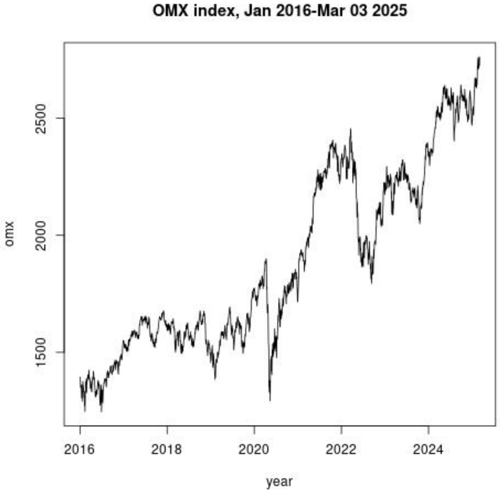
Introduction, examples



Introduction, examples

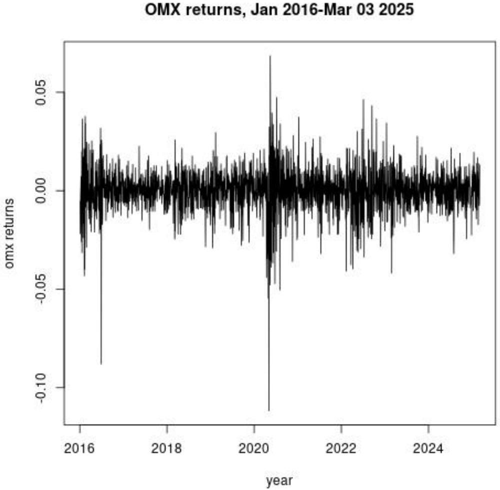


Introduction, examples



Introduction, examples

OMX index x_t, returns $r_t = \log x_t - \log x_{t-1}$.



Statistical models

Definition (Stochastic process)

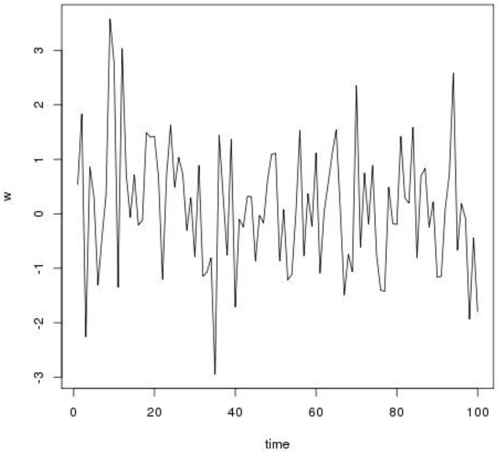
A collection of random variables $\{x_t\}$ where t ranges over a set of integers, is called a *stochastic process in discrete time* (time series).

Definition (Realization)

A collection of observed values of a stochastic process is called a *realization*.

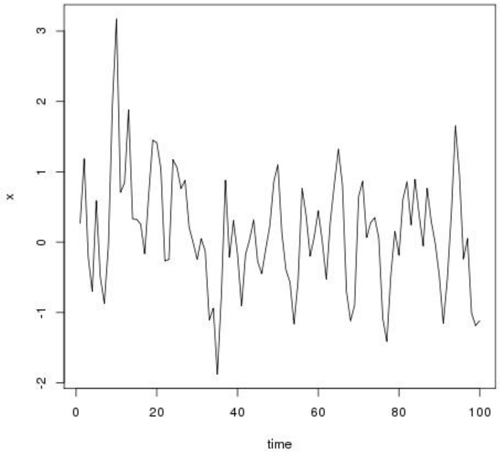
Statistical models

Example 1: White noise, $w_t \sim N(0, \sigma_w^2)$, independent



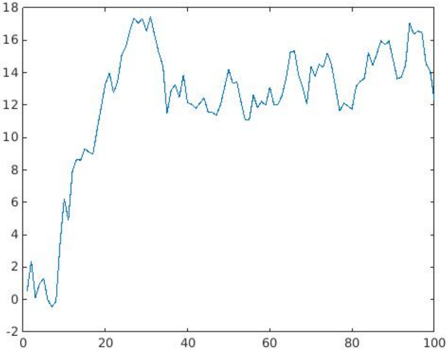
Statistical models

Example 2: Moving average, $x_t = \frac{1}{2}(w_t + w_{t-1})$



Statistical models

Example 3: Random walk, $x_t = x_{t-1} + w_t = x_0 + w_1 + \ldots + w_t$.



Measures of dependence

Definition (1.1)

The *mean function* of a stochastic process $\{x_t\}$ is defined as

$$\mu_t = E(x_t).$$

Definition (1.2)

The *autocovariance function* of a stochastic process $\{x_t\}$ is defined as

$$\gamma(s,t) = \mathrm{cov}(x_s, x_t).$$

Definition (1.3)

The *autocorrelation function* of a stochastic process $\{x_t\}$ is defined as

$$\rho(s,t) = \mathrm{corr}(x_s, x_t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}.$$

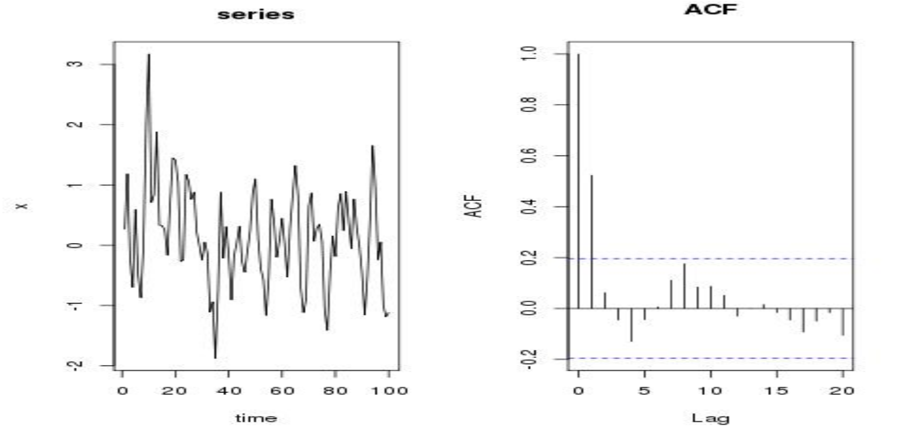
Measures of dependence

Calculate μ_t, $\gamma(s,t)$ and $\rho(s,t)$ for

1. the white noise process w_t.
2. the moving average process $x_t = \frac{1}{2}(w_t + w_{t-1})$.
3. the random walk process $x_t = x_{t-1} + w_t$ where $x_0 = 0$.
4. In general, is it true that $\gamma(s,t) = \gamma(t,s)$ and $\rho(s,t) = \rho(t,s)$?

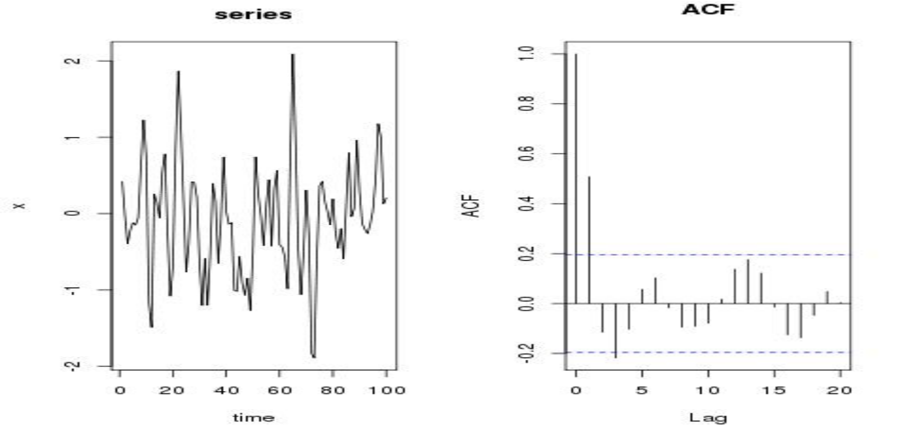
Measures of dependence

Simulation one of $x_t = \frac{1}{2}(w_t + w_{t-1})$



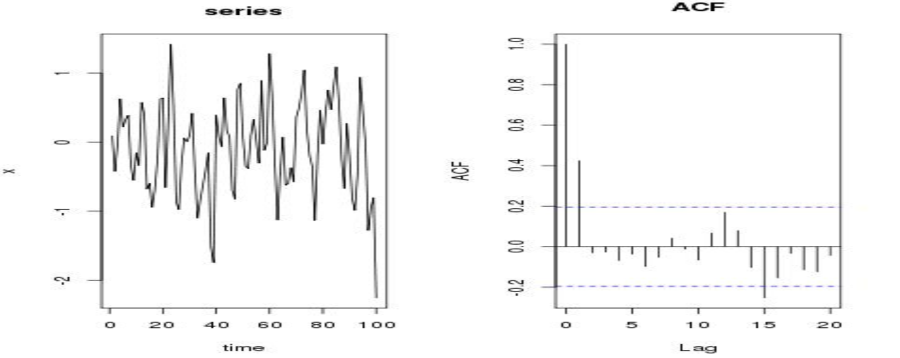
Measures of dependence

Simulation two of $x_t = \frac{1}{2}(w_t + w_{t-1})$



Measures of dependence

Simulation three of $x_t = \frac{1}{2}(w_t + w_{t-1})$



In all three simulations, the series look quite different but the ACFs are similar.

Measures of dependence

Definition (1.4)

The *cross-covariance function* between two series $\{x_t\}$ and $\{y_t\}$ is defined as

$$\gamma_{xy}(s, t) = \mathrm{cov}(x_s, y_t).$$

Definition (1.5)

The *cross-correlation function* between two series $\{x_t\}$ and $\{y_t\}$ is defined as

$$\rho_{xy}(s, t) = \mathrm{corr}(x_s, y_t) = \frac{\gamma_{xy}(s, t)}{\sqrt{\gamma_x(s, s)\gamma_y(t, t)}}.$$

Measures of dependence

1. Let $x_t = \frac{1}{2}(w_t + w_{t-1})$ and $y_t = w_t$.
    1. Calculate $\gamma_{xy}(s,t)$ and $\rho_{xy}(s,t)$.
    2. Calculate $\gamma_{yx}(s,t)$ and $\rho_{yx}(s,t)$.
2. Let $x_t = x_{t-1} + w_t$ where $x_0 = 0$, and $y_t = w_t$.
    1. Calculate $\gamma_{xy}(s,t)$ and $\rho_{xy}(s,t)$.
    2. Calculate $\gamma_{yx}(s,t)$ and $\rho_{yx}(s,t)$.
3. In general, is it true that $\gamma_{xy}(s,t) = \gamma_{xy}(t,s)$ and $\rho_{xy}(s,t) = \rho_{xy}(t,s)$?
4. In general, is it true that $\gamma_{xy}(s,t) = \gamma_{yx}(t,s)$ and $\rho_{xy}(s,t) = \rho_{yx}(t,s)$?

News of today

Definitions of

- a discrete time stochastic process (time series)
- the mean function
- the autocovariance function
- the autocorrelation function
- the cross-covariance function
- the cross-correlation function

Analysis of Time Series, L2

Rolf Larsson

Uppsala University

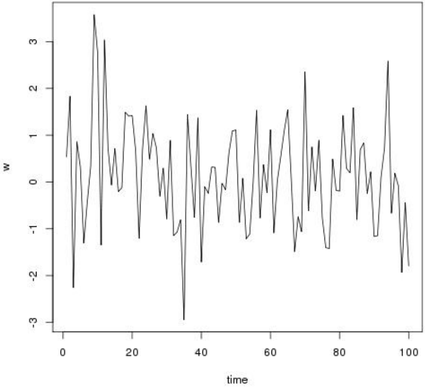
26 mars 2025

Today

- 1.5: Stationary time series
- 1.6: Estimation of correlation
- 2.3: Differencing
- Menti

Stationary time series

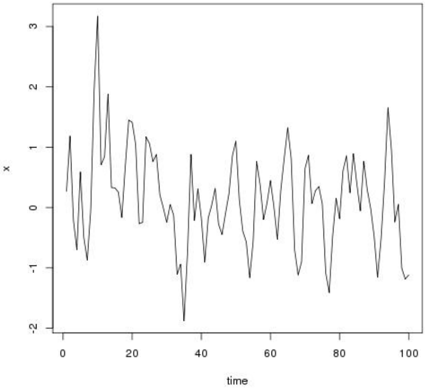
Example 1: White noise, $w_t \sim N(0, \sigma_w^2)$, independent



Erratic behaviour.

Stationary time series

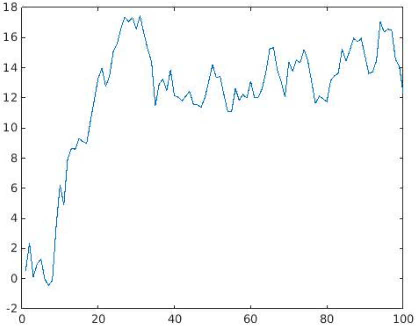
Example 2: Moving average, $x_t = \frac{1}{2}(w_t + w_{t-1})$



A little smoother.

Stationary time series

Example 3: Random walk, $x_t = x_{t-1} + w_t$



Recall: $E(x_t) = \mu_t = 0$,
$\mathrm{cov}(x_s, x_t) = \gamma(s,t) = \min(s,t)\sigma_w^2 \Rightarrow \mathrm{var}(x_t) = \gamma(t,t) = t\sigma_w^2$.

Stationary time series

Definition (1.6)

If

$$\{x_{t_1}, x_{t_2}, \ldots, x_{t_k}\} \quad \text{and} \quad \{x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_k+h}\}$$

have identical joint distributions for all choices of $(t_1, t_2, \ldots, t_k)$ and h, then $\{x_t\}$ is *strictly stationary*.

Stationary time series

Special cases:

- $k = 1$: x_t and x_{t+h} have identical distributions for all t and h.
- Implications:
  - The mean function μ_t is constant.
  - The variance function $\gamma(t, t)$ is constant.
- $k = 2$: (x_s, x_t) and (x_{s+h}, x_{t+h}) have identical joint distributions for all (s, t) and h.
- Implications:
  - The autocovariance function $\gamma(s, t)$ only depends on s and t through $|s - t|$.
  - The autocorrelation function $\rho(s, t)$ only depends on s and t through $|s - t|$.

Stationary time series

Definition (1.7)

If x_t has finite variance for all t,

1. the mean function μ_t is constant,
2. the autocovariance function $\gamma(s, t)$ only depends on s and t through $|s - t|$,

then $\{x_t\}$ is *weakly stationary*.

- $\mathrm{var}(x_t) < \infty$ and x_t strictly stationary $\Rightarrow x_t$ is weakly stationary.
- $\Leftarrow$ not true in general. (cf problem 1.16)

Stationary time series

Definition (1.13)

A process $\{x_t\}$ is said to be *Gaussian* if $\{x_{t_1}, x_{t_2}, ..., x_{t_k}\}$ has a multivariate normal distribution for all choices of $(t_1, t_2, ..., t_k)$.

For a Gaussian process, the concepts of strict and weak stationarity are equivalent.

Stationary time series

Definition (1.8)

The *autocovariance function* of a *stationary* stochastic process $\{x_t\}$ is defined as

$$\gamma(h) = \mathrm{cov}(x_{t+h}, x_t).$$

Definition (1.9)

The *autocorrelation function* of a *stationary* stochastic process $\{x_t\}$ is defined as

$$\rho(h) = \mathrm{corr}(x_{t+h}, x_t) = \frac{\gamma(h)}{\gamma(0)}.$$

Stationary time series

Calculate $\gamma(h)$ and $\rho(h)$ for

1. the white noise process w_t.
2. the moving average process $x_t = \frac{1}{2}(w_t + w_{t-1})$.

Stationary time series

Some properties:

1. $\mathrm{var}(x_t) = \gamma(0)$
2. $\gamma(h) = \gamma(-h)$ for all h
3. $\rho(h) = \rho(-h)$ for all h
4. $|\rho(h)| \leq 1$ for all h
5. $|\gamma(h)| \leq \gamma(0)$ for all h

Stationary time series

Recall:

Definition (1.4)

The *cross-covariance function* between two series $\{x_t\}$ and $\{y_t\}$ is defined as

$$\gamma_{xy}(s,t) = \mathrm{cov}(x_s, y_t).$$

Definition (1.5)

The *cross-correlation function* between two series $\{x_t\}$ and $\{y_t\}$ is defined as

$$\rho_{xy}(s,t) = \mathrm{corr}(x_s, y_t) = \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\gamma_y(t,t)}}.$$

Stationary time series

Definition (1.10)

Two stationary time series $\{x_t\}$ and $\{y_t\}$ are said to be *jointly (weakly) stationary* if the cross-covariance function

$$\gamma_{xy}(h) = \mathrm{cov}(x_{t+h}, y_t)$$

is a function only of h.

Definition (1.11)

The *cross-correlation function* between two jointly stationary time series $\{x_t\}$ and $\{y_t\}$ is defined as

$$\rho_{xy}(h) = \mathrm{corr}(x_{t+h}, y_t) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}}.$$

Stationary time series

- Let $x_t = \frac{1}{2}(w_t + w_{t-1})$ and $y_t = w_t$.
- Calculate $\gamma_{xy}(h)$ and $\rho_{xy}(h)$.

Stationary time series

Some properties:

1. $|\rho_{xy}(h)| \leq 1$ for all h
2. $\rho_{xy}(h) = \rho_{yx}(-h)$ for all h
3. $\rho_{xy}(h)$ and $\rho_{xy}(-h)$ are not equal in general!

Estimation of correlation

Definition (1.14)

The sample autocovariance function is defined as

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

for $h = 0, 1, ..., n-1$, $\hat{\gamma}(-h) = \hat{\gamma}(h)$.

Definition (1.15)

The sample autocorrelation function (ACF) is defined as

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

Estimation of correlation

Theorem (Property 1.2)

If the process $\{x_t\}$ is white noise with finite fourth moment, then for large n,

$$\hat{\rho}(h) \approx N\left(0, \frac{1}{n}\right).$$

Stricter formulation: $\sqrt{n}\hat{\rho}(h)$ converges to $N(0,1)$ as $n \to \infty$.

Proof: see theorem A.7.

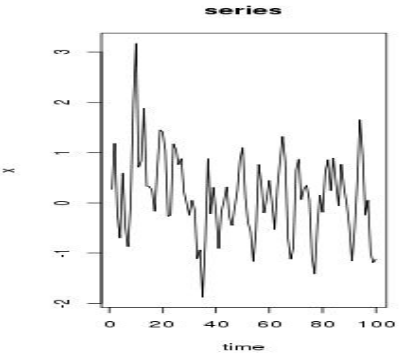
- For large n, H_0: $\rho(h) = 0$ is rejected vs H_1: $\rho(h) \neq 0$ at approximately the 5% level if

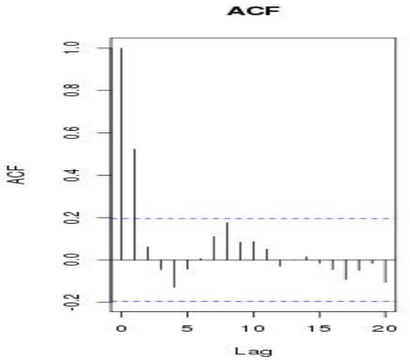
$$|\hat{\rho}(h)| > \frac{2}{\sqrt{n}}.$$

- In R: `acf(x)`

Estimation of correlation

Simulation of $x_t = \frac{1}{2}(w_t + w_{t-1})$

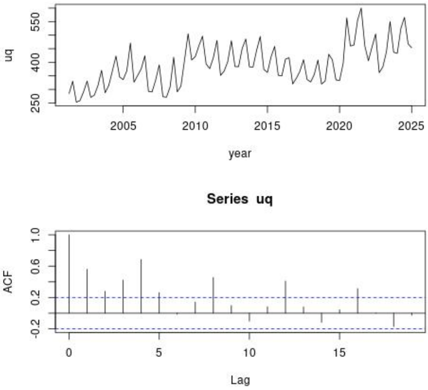




The ACF is significantly nonzero only for lag 1.

Estimation of correlation

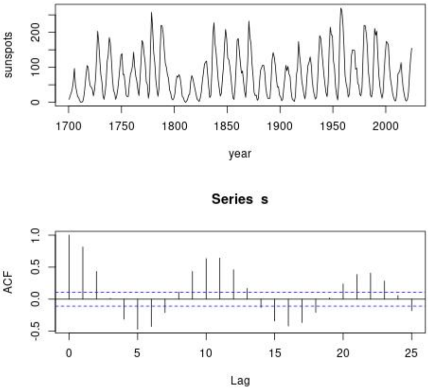
Unemployment, quarterly data.



The ACF is large for multiples of 4.

Estimation of correlation

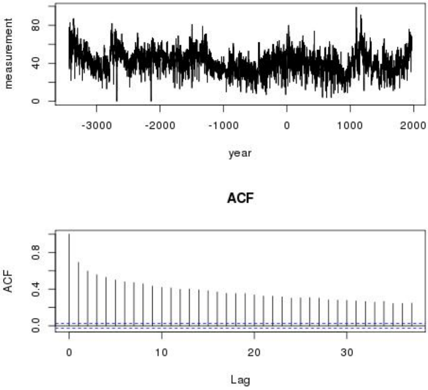
Average number of sunspots per year.



The ACF is large for multiples of around 10 or 11.

Estimation of correlation

Tree ring measurements, Mount Campito.



The ACF decays slowly with increasing lag. ('Long memory'.)

Estimation of correlation

Definition (1.16)

The sample cross-covariance function is defined as

$$\hat{\gamma}_{xy}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y}),$$

for $h = 0, 1, ..., n-1$, $\hat{\gamma}_{xy}(-h) = \hat{\gamma}_{yx}(h)$.

The sample cross-correlation function (CCF) is defined as

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}}.$$

Estimation of correlation

Theorem (Property 1.3)

If $\{x_t\}$ and $\{y_t\}$ are independent and at least one of $\{x_t\}$ or $\{y_t\}$ is independent white noise, then for n large,

$$\hat{\rho}_{xy}(h) \approx N\left(0, \frac{1}{n}\right).$$

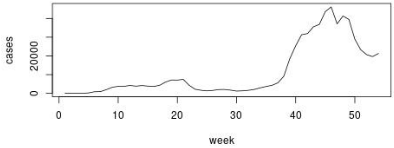
Stricter formulation: $\sqrt{n}\hat{\rho}_{xy}(h)$ converges to $N(0,1)$ as $n \to \infty$.

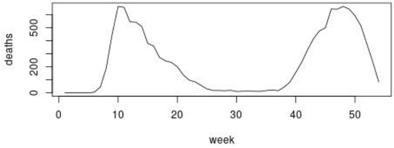
Proof: see theorem A.8.

In R: `ccf(x,y)`

Estimation of correlation

Covid-19, numbers of reported cases and deaths in Sweden, 2020-2021 week 6.

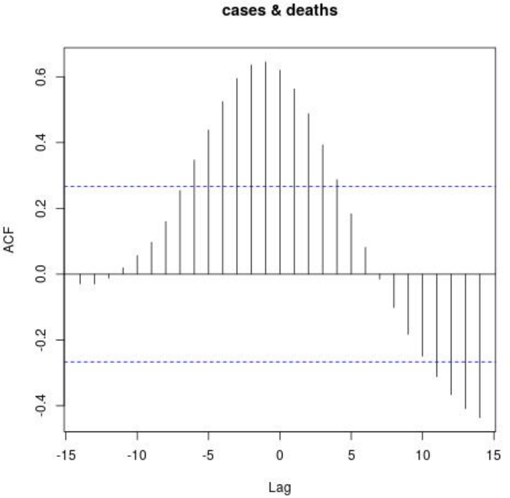




Not much testing in the beginning.

Estimation of correlation

Estimated cross correlation, cases and deaths. Largest at lag -1!



Differencing

Definition (2.4)

The *backshift operator* is defined by

$$Bx_t = x_{t-1}.$$

For $k = 1, 2, \ldots$, $B^k x_t = x_{t-k}$.

Definition (2.5)

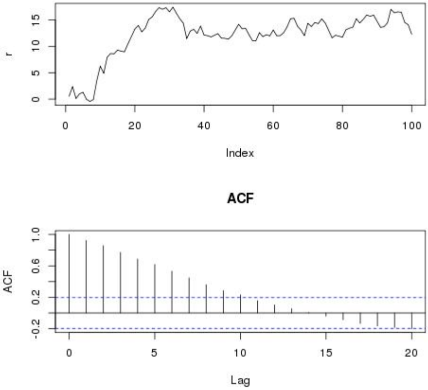
Differences of order d are defined by

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

- Special cases:
  - $\nabla^1 x_t = \nabla x_t = (1-B)x_t = x_t - x_{t-1}$ (often needed)
  - $\nabla^2 x_t = (1-B)^2 x_t = x_t - 2x_{t-1} + x_{t-2}$ (rarely needed)
- It is *very rare* that more than two differences are needed!
- In R: `diff(x)`

Differencing

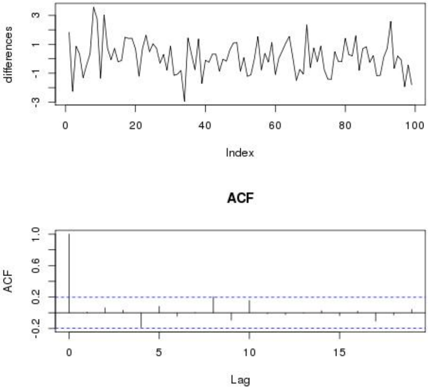
Simulated random walk.



The ACF decays slowly for increasing lags.

Differencing

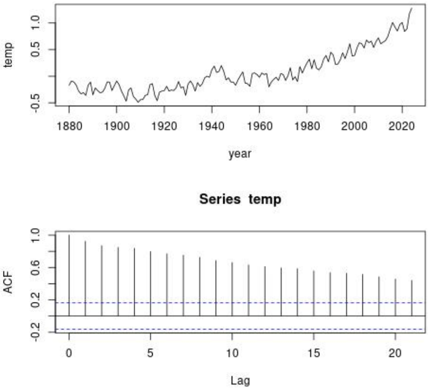
Simulated random walk, differences.



The ACF cuts off after lag zero. It is between the dashed lines for pos. lags.

Differencing

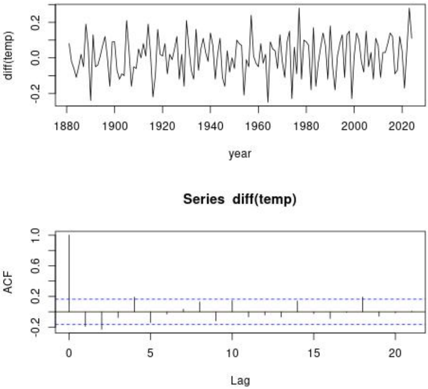
Global mean temperature.



Increasing trend. The ACF decays slowly for increasing lags.

Differencing

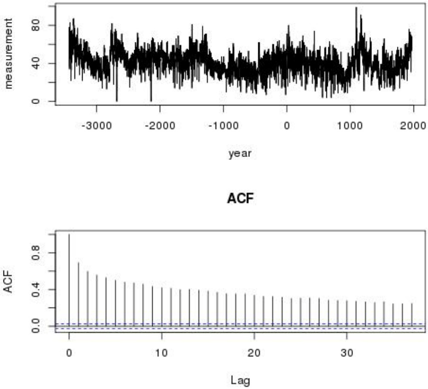
Global mean temperature, differences.



The ACF basically cuts off after lag zero.

Differencing

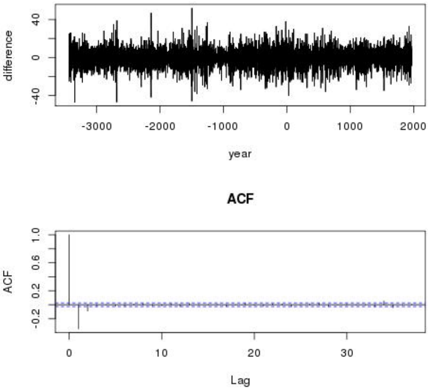
Tree ring measurements, Mount Campito.



The ACF decays slowly with increasing lag. ('Long memory'.)

Differencing

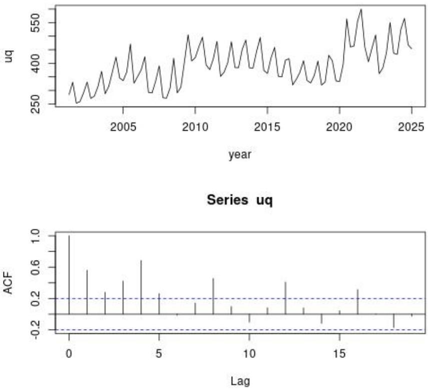
Tree ring measurements, Mount Campito, differences.



The ACF is markedly different from zero only for small lags.

Differencing

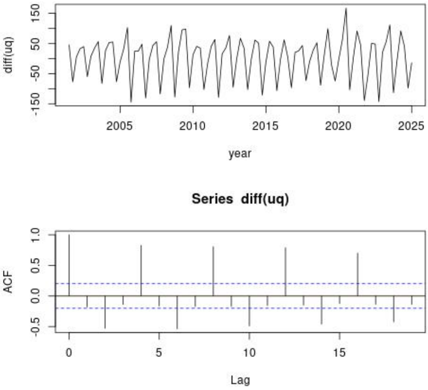
Unemployment (quarterly).



The ACF is large for multiples of 4.

Differencing

Unemployment, differences.



Still, the ACF is large for multiples of 4.

News of today

- strict stationarity
- weak stationarity
- ACF
- CCF
- differencing

Analysis of Time Series, L3

Rolf Larsson

Uppsala University

27 mars 2025

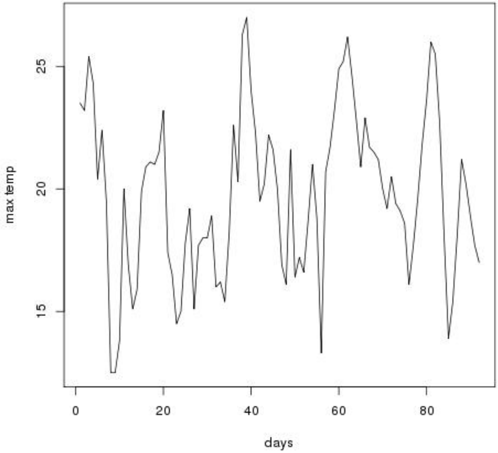
Today

3.1-3.2:

- Autoregressive (AR) models
- Moving average (MA) models
- Menti

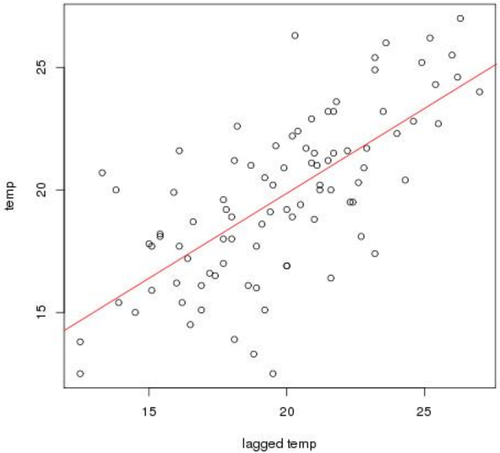
AR models

Daily temperatures in Uppsala, June-August 1984.



AR models

Regression on lagged temperatures: $x_t = 6.00 + 0.693x_{t-1} + w_t$



AR models

Autoregressive model of order 1, AR(1)

- without constant ($E(x_t) = 0$)

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

- with constant ($E(x_t) = \mu$)

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

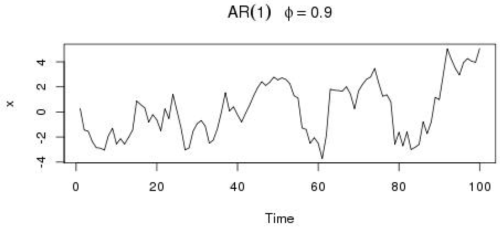
or equivalently

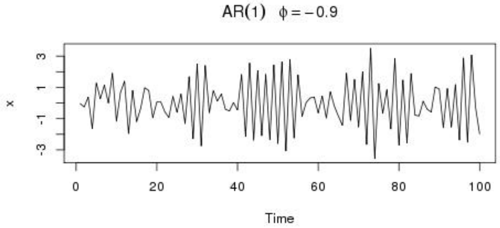
$$x_t = \mu + \phi(x_{t-1} - \mu) + w_t = \alpha + \phi x_{t-1} + w_t,$$

where $\alpha = \mu(1 - \phi)$.

AR models

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





AR models

Definition (3.1)

An *autoregressive model* of order p, AR(p), with mean zero is given by

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t,$$

where w_t is white noise with mean zero and variance σ^2.

An *autoregressive model* of order p, AR(p) with mean μ is given by

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \dots + \phi_p(x_{t-p} - \mu) + w_t,$$

or equivalently

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t,$$

where $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$.

AR models

Recall: $Bx_t = x_{t-1}$, $B^k x_t = x_{t-k}$.

Definition (3.2)

The *autoregressive operator* is defined as

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p.$$

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t$$

is equivalent to

$$\phi(B) x_t = w_t.$$

AR models

Definition (1.12)

A *linear process* x_t is given by

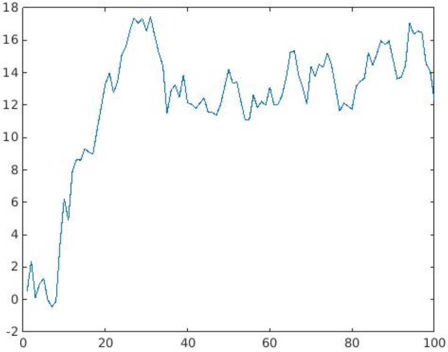
$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \qquad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

If x_t is AR(1) with mean zero, show that

1. $$x_t = \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j}.$$
2. If $|\phi| < 1$ and x_t is stationary, then x_t is a linear process.

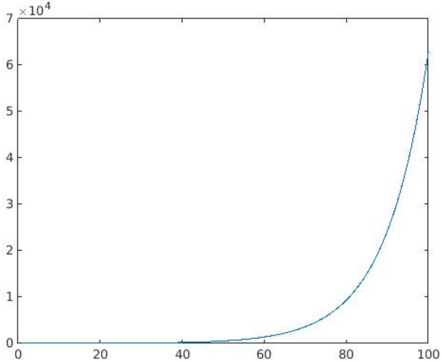
AR models

Simulated series: $x_t = x_{t-1} + w_t$



AR models

Simulated series: $x_t = 1.1x_{t-1} + w_t$ (observe the scale on the y axis)



By recursion: $x_t = 1.1^{t-1}x_1 + 1.1^{t-2}w_2 + ... + 1.1w_{t-1} + w_t$
The first term dominates!

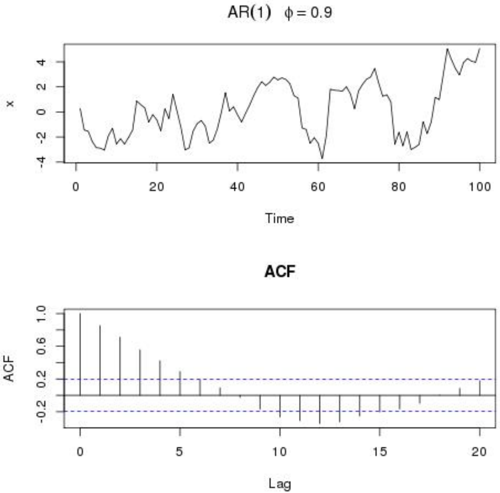
AR models

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

1. Derive the MA (moving average, linear process) representation using
   a) $\psi(B)\phi(B) = 1$ where $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots$
   b) $x_t = \phi^{-1}(B) w_t$
2. Prove that the autocorrelation function is $\rho(h) = \phi^{|h|}$.

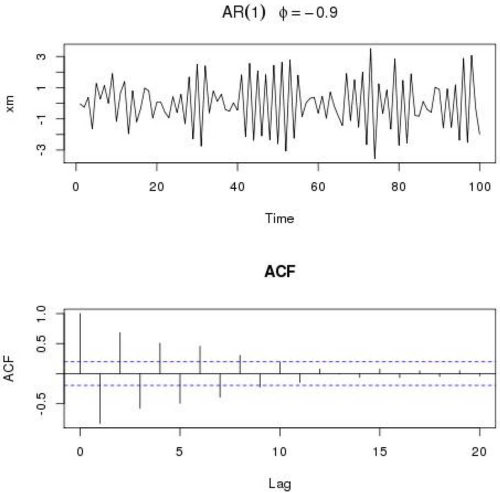
AR models

Simulated series: $x_t = 0.9x_{t-1} + w_t$ and estimated ACF



AR models

Simulated series: $x_t = -0.9x_{t-1} + w_t$ and estimated ACF



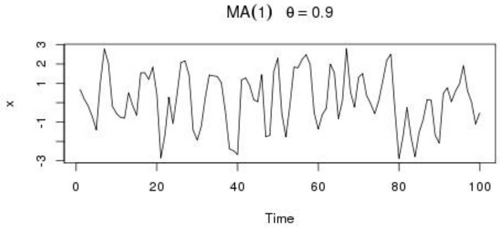
MA models

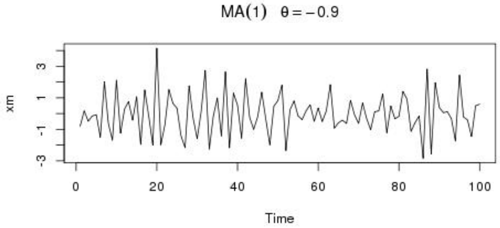
Moving average model of order 1, MA(1)

$$x_t = w_t + \theta w_{t-1}$$

MA models

Simulated series: $x_t = w_t + \theta w_{t-1}$





MA models

Definition (3.3)

A *moving average model* of order q, MA(q), is given by

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q},$$

where w_t is white noise with mean zero and variance σ^2.

MA models

Definition (3.4)

The *moving average operator* is defined as

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q.$$

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q}$$

is equivalent to

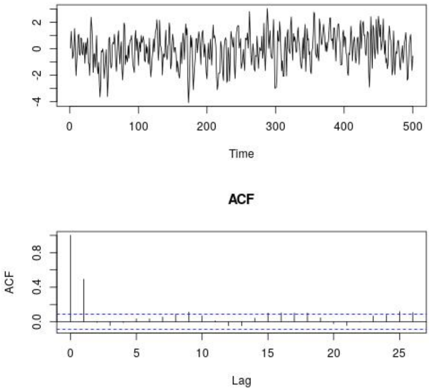
$$x_t = \theta(B) w_t.$$

MA models

1. Let $x_t = w_t + \theta w_{t-1}$.
   a) Derive the autocorrelation function $\rho(h)$, and show that it is the same if θ is replaced by $1/\theta$.
   b) Derive the AR representation.
2. Let x_t be an MA(q) process and show that $\rho(h) \neq 0$ only if $|h| \leq q$.

MA models

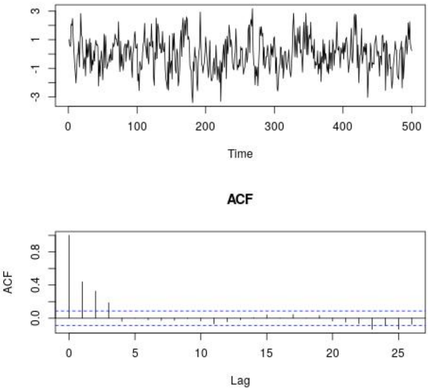
Simulation of $x_t = w_t + 0.8w_{t-1}$:



Theoretically, the ACF cuts off after lag 1.

MA models

Simulation of $x_t = w_t + 0.3w_{t-1} + 0.3w_{t-2} + 0.3w_{t-3}$:



Theoretically, the ACF cuts off after lag 3.

News of today

- AR processes
  - Definition
  - MA representation
  - Autocorrelation function
- MA processes
  - Definition
  - Autocorrelation function
  - AR representation

Analysis of Time Series, L4

Rolf Larsson

Uppsala University

28 mars 2025

Today

- 3.2: Autoregressive moving average (ARMA) models
- 3.3: Difference equations
- Menti

ARMA models

ARMA(1,1):

- Let
$$x_t = \phi x_{t-1} + w_t + \theta w_{t-1}$$
- On operator form:
$$\phi(B)x_t = \theta(B)w_t$$
where
$\phi(B) = 1 - \phi B$,
$\theta(B) = 1 + \theta B$.
- What happens if $\phi = -\theta$?

ARMA models

Definition (3.5)

An *autoregressive moving average model* of order p, q, ARMA(p, q), is given by

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q},$$

where w_t is white noise with mean zero and variance σ^2.

On operator form:

$$\phi(B)x_t = \theta(B)w_t$$

where
$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$,
$\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$.

ARMA models

Example 1:

Let

$$x_t = 0.7x_{t-1} - 0.1x_{t-2} + w_t - 0.9w_{t-1} + 0.2w_{t-2}.$$

Is this an ARMA(2,2) model?

ARMA models

Potential problems with ARMA models:

- Parameter redundancy
  (cf the previous example)
- Different MA models have the same ACF
  (cf $x_t = w_t + \theta w_{t-1}$, $x_t = w_t + \theta^{-1} w_{t-1}$).
- Stationary AR models that depend on the future
  (cf $x_t = \phi x_{t-1} + w_t$ with $\phi > 1$)

ARMA models

Definition (3.7)

An ARMA model is said to be *causal* if it can be written as a one-sided linear process, i.e.

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j},$$

where $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Consider the AR(1) process $x_t = \phi x_{t-1} + w_t$.
For which ϕ is it causal?

ARMA models

Definition (3.6)

The *AR and MA polynomials* are defined as

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p, \quad \phi_p \neq 0,$$

and

$$\theta(z) = 1 + \theta_1 z - \ldots + \theta_q z^q, \quad \theta_q \neq 0,$$

where z is a complex number.

Theorem (Property 3.1)

1. *An ARMA(p, q) model is causal if and only if $\phi(z) \neq 0$ for all $|z| \leq 1$.*
2. *The coefficients of the MA representation may be found through* $\sum_{j=0}^{\infty} \psi_j z^j = \psi(z) = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1.$

The proof (Appendix B) builds on a series expansion of $1/\phi(z)$.

ARMA models

Definition (3.8)

An ARMA(p, q) model is said to be *invertible* if it can be written on AR form as

$$\sum_{j=0}^{\infty} \pi_j x_{t-j} = w_t,$$

where $\sum_{j=0}^{\infty} |\pi_j| < \infty$.

Theorem (Property 3.2)

1. *An ARMA(p, q) model is invertible if and only if $\theta(z) \neq 0$ for all $|z| \leq 1$.*
2. *The coefficients of the AR representation may be found through*

$$\sum_{j=0}^{\infty} \pi_j z^j = \pi(z) = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1.$$

ARMA models

ARMA model $\phi(B)x_t = \theta(B)w_t$.
Note:

- MA processes are always causal, because $\phi(z) = 1 \neq 0$ for all z (in particular for all $|z| \leq 1$).
- AR processes are always invertible, because $\theta(z) = 1 \neq 0$ for all z (in particular for all $|z| \leq 1$).

ARMA models

Example 1:

Let

$$x_t = 0.7x_{t-1} - 0.1x_{t-2} + w_t - 0.9w_{t-1} + 0.2w_{t-2}.$$

1. Is the model causal?
2. Is the model invertible?
3. Find the MA representation.

ARMA models

Example 2: AR(2)

Let

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t.$$

For which choices of ϕ_1 and ϕ_2 is the model causal?

ARMA models

Example 2 cont.:

1. For which (ϕ_1, ϕ_2) do the solutions $z_{1,2}$ of

$$0 = \phi(z) = 1 - \phi_1 z - \phi_2 z^2$$

   fulfill $|z_{1,2}| > 1$?
   Assume $\phi_2 \neq 0$, $z_{1,2} \neq 0$.

2. Solutions (why?):

$$z_{1,2} = -\frac{\phi_1}{2\phi_2} \pm r, \quad r = \sqrt{\frac{\phi_1^2}{4\phi_2^2} + \frac{1}{\phi_2}}.$$

3. It follows that (why?)

$$\frac{1}{z_{1,2}} = \frac{\phi_1 \pm 2\phi_2 r}{2}.$$

ARMA models

Example 2 cont.:

4. This, together with $|z_{1,2}| > 1$, implies (why?)

$$\phi_2 - \phi_1 < 1, \quad \phi_2 + \phi_1 < 1.$$

5. From the factorization theorem,

$$z^2 + \frac{\phi_1}{\phi_2} z - \frac{1}{\phi_2} = (z - z_1)(z - z_2),$$

which implies (why?)

$$1 < |z_1||z_2| = |z_1 z_2| = \frac{1}{|\phi_2|}$$

i.e. $|\phi_2| < 1$.

6. Sketch the region in the (ϕ_1, ϕ_2) plane!

Difference equations

Order one:

- Solve $u_n - \alpha u_{n-1} = 0$, $u_0 = c$.
- Recursion yields $u_n = \alpha^n c$.
- Equivalently, $\alpha(B)u_n = (1 - \alpha B)u_n = 0$, $u_0 = c$
  is solved by $u_n = (z_0^{-1})^n c$
  where $z_0 = \alpha^{-1}$ is a root of $\alpha(z) = 1 - \alpha z$.

Difference equations

Order two:

- Solve
$$u_n - \alpha_1 u_{n-1} - \alpha_2 u_{n-2} = 0 \tag{1}$$
- Equivalently, write $\alpha(B)u_n = (1 - \alpha_1 B - \alpha_2 B^2)u_n = 0$.
- Denote the roots of $\alpha(z)$ by z_1 and z_2.
- If $z_1 \neq z_2$, the general solution to (1) is
$$u_n = c_1 z_1^{-n} + c_2 z_2^{-n}.$$
- If $z_1 = z_2$, the general solution to (1) is
$$u_n = z_1^{-n}(c_1 + c_2 n).$$

News of today

- ARMA processes
- Parameter redundancy
- Causality
- Invertibility
- Difference equations and their use

Analysis of Time Series, L5

Rolf Larsson

Uppsala University

31 mars 2025

Today

- 3.3: Difference equations (continued)
- 3.4:
  - ACF
  - PACF
  - Summary and examples
- Menti

Difference equations

Order one:

- Solve $u_n - \alpha u_{n-1} = 0$, $u_0 = c$.
- Recursion yields $u_n = \alpha^n c$.
- Equivalently, $\alpha(B)u_n = (1 - \alpha B)u_n = 0$, $u_0 = c$
  is solved by $u_n = (z_0^{-1})^n c$
  where $z_0 = \alpha^{-1}$ is a root of $\alpha(z) = 1 - \alpha z$.

Difference equations

Order two:

- Solve
$$u_n - \alpha_1 u_{n-1} - \alpha_2 u_{n-2} = 0 \tag{1}$$
- Equivalently, write $\alpha(B)u_n = (1 - \alpha_1 B - \alpha_2 B^2)u_n = 0$.
- Denote the roots of $\alpha(z)$ by z_1 and z_2.
- If $z_1 \neq z_2$, the general solution to (1) is
$$u_n = c_1 z_1^{-n} + c_2 z_2^{-n}.$$
- If $z_1 = z_2$, the general solution to (1) is
$$u_n = z_1^{-n}(c_1 + c_2 n).$$

Difference equations

Example 2: AR(2) (causal)

Let

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t,$$

i.e $\phi(B)x_t = w_t$, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$.

1. Show that the autocorrelation function satisfies, for $h \geq 1$,

$$\rho(h) - \phi_1 \rho(h-1) - \phi_2 \rho(h-2) = 0.$$

2. Denote the roots of $\phi(z)$ by z_1 and z_2. Find an expression for $\rho(h)$ when
   a) $z_1 \neq z_2$ and real
   b) $z_1 = z_2$
   c) $z_1 = \bar{z}_2$ is a complex conjugate pair

ACF

MA(q):

- Let

$$x_t = \theta_0 w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q}, \quad \theta_0 = 1.$$

- The ACF is given by (why?)

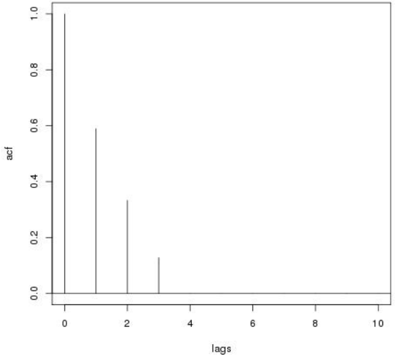
$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{\sum_{j=0}^{q} \theta_j^2}, & 1 \leq h \leq q, \\ 0, & h > q \end{cases}$$

  and $\rho(-h) = \rho(h)$.

- A good tool to identify q!

ACF

Theoretical ACF of $x_t = w_t + 0.6w_{t-1} + 0.4w_{t-2} + 0.2w_{t-3}$



´

ACF

AR(p):

- Let
$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t.$$
- The ACF satisfies (why?)
$$\rho(h) - \phi_1 \rho(h-1) - \phi_2 \rho(h-2) - ... - \phi_p \rho(h-p) = 0, \quad h \geq p.$$
- What is the general form of the solution?

ACF

Recall the difference equation of order two:

- Solve
$$u_n - \alpha_1 u_{n-1} - \alpha_2 u_{n-2} = 0 \tag{2}$$
- Equivalently, write $\alpha(B)u_n = (1 - \alpha_1 B - \alpha_2 B^2)u_n = 0$.
- Denote the roots of $\alpha(z)$ by z_1 and z_2.
- If $z_1 \neq z_2$, the general solution to (2) is
$$u_n = z_1^{-n} c_1 + z_2^{-n} c_2.$$
- If $z_1 = z_2$, the general solution to (2) is
$$u_n = z_1^{-n}(c_1 + c_2 n).$$

ACF

In general:

- Solve
$$u_n - \alpha_1 u_{n-1} - \alpha_2 u_{n-2} - \dots - \alpha_p u_{n-p} = 0. \tag{3}$$
- Equivalently, write
$$\alpha(B)u_n = (1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p)u_n = 0.$$
- Denote the roots of $\alpha(z)$ by z_1, z_2,..., z_r, each with multiplicity m_1, m_2,..., m_r.
- The general solution to (3) is of the form
$$u_n = z_1^{-n} P_1(n) + z_2^{-n} P_2(n) + \dots + z_r^{-n} P_r(n),$$
where $P_1(n), \dots, P_r(n)$ are polynomials of degrees $m_1 - 1, \dots, m_r - 1$.

ACF

AR(p):

- Solve

$$\rho(h) - \phi_1\rho(h-1) - \phi_2\rho(h-2) - \ldots - \phi_p\rho(h-p) = 0, \quad h \geq p. \quad (4)$$

- Equivalently, write

$$\phi(B)\rho(h) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)\rho(h) = 0.$$

- Denote the roots of $\phi(z)$ by z_1, z_2,..., z_r, each with multiplicity m_1, m_2,..., m_r.
- The general solution to (4) is of the form

$$\rho(h) = z_1^{-h}P_1(h) + z_2^{-h}P_2(h) + \ldots + z_r^{-h}P_r(h),$$

where $P_1(h), \ldots, P_r(h)$ are polynomials of degrees $m_1 - 1, \ldots, m_r - 1$.

ACF

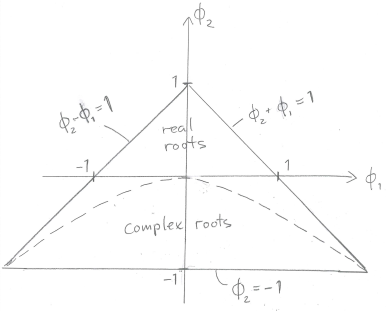
AR(p):

- $$\rho(h) = z_1^{-h} P_1(h) + z_2^{-h} P_2(h) + ... + z_r^{-h} P_r(h),$$

  where $P_1(h), ..., P_r(h)$ are polynomials of degrees $m_1 - 1, ..., m_r - 1$.
- Assume that the model is causal (all $|z_j| > 1$).
- If all roots are real, the ACF decays exponentially fast as $h \to \infty$.
- If some roots are complex, the ACF decays in a sinusoidal fashion.
- Not a good tool for identifying p!

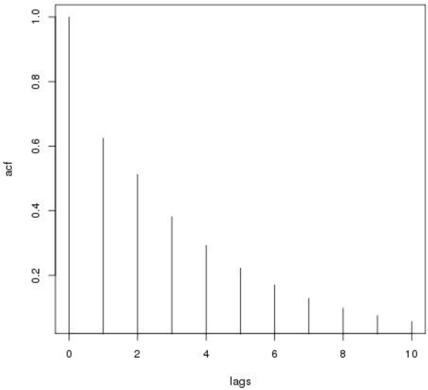
ACF

Causality region (inside the triangle) for AR(2):



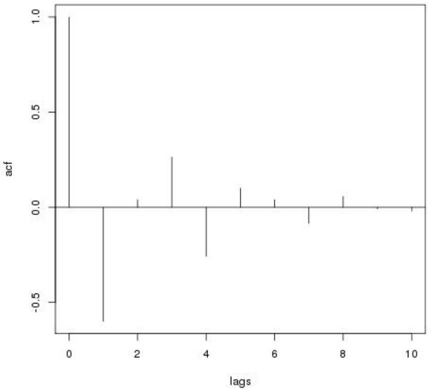
ACF

Theoretical ACF of $x_t = 0.5x_{t-1} + 0.2x_{t-2} + w_t$. (Real roots.)



ACF

Theoretical ACF of $x_t = -0.9x_{t-1} - 0.5x_{t-2} + w_t$. (Complex roots.)



PACF

Let $x_t = \phi x_{t-1} + w_t$ be causal.

1. Calculate $\mathrm{cov}(x_{t+2}, x_t)$. Is it zero?
2. a) Find a such that $E\{(x_{t+2} - ax_{t+1})^2\}$ is minimized.
   b) Find b such that $E\{(x_t - bx_{t+1})^2\}$ is minimized.
3. With such a, b, show that

$$\mathrm{cov}(x_{t+2} - ax_{t+1}, x_t - bx_{t+1}) = 0.$$

(uncorrelated 'projection errors')

PACF

- Let $\{x_t\}$ be a mean zero stationary process. Take any $h \geq 2$.
- Let
$$\hat{x}_{t+h} = a_1 x_{t+h-1} + a_2 x_{t+h-2} + \dots + a_{h-1} x_{t+1}$$
be such that $E\{(x_{t+h} - \hat{x}_{t+h})^2\}$ is minimized.
- Let
$$\hat{x}_t = b_1 x_{t+1} + b_2 x_{t+2} + \dots + b_{h-1} x_{t+h-1}$$
be such that $E\{(x_t - \hat{x}_t)^2\}$ is minimized.

Definition (3.9)

The *partial autocorrelation function* (PACF) is defined as

$$\phi_{hh} = \begin{cases} \mathrm{corr}(x_{t+1}, x_t), & h = 1, \\ \mathrm{corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t), & h \geq 2. \end{cases}$$

PACF

- $$\phi_{hh} = \begin{cases} \mathrm{corr}(x_{t+1}, x_t), & h = 1, \\ \mathrm{corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t), & h \geq 2. \end{cases}$$
- If $\{x_t\}$ is a normal (Gaussian) process,

$$\phi_{hh} = \mathrm{corr}(x_{t+h}, x_t | x_{t+1}, x_{t+2}, ..., x_{t+h-1}),$$

cf Appendix B.

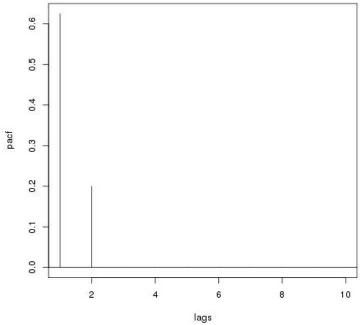
PACF

AR(p):

- Let
$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t.$$
- The PACF $\phi_{hh} = 0$ if $h > p$. (Why?)
- It is a good tool for identifying p!

PACF

Theoretical PACF of $x_t = 0.5x_{t-1} + 0.2x_{t-2} + w_t$.



Equals 0.2 at lag 2 and cuts off after this lag!

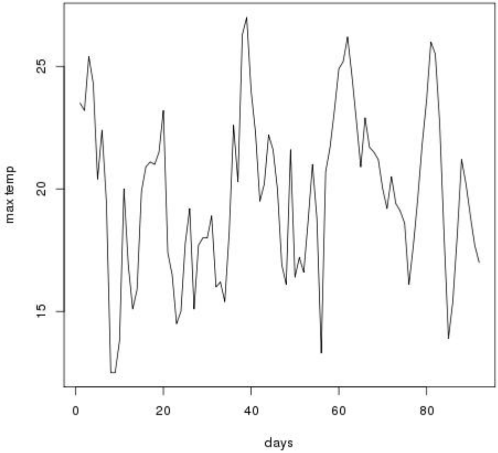
Summary and examples

Table 3.1:

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

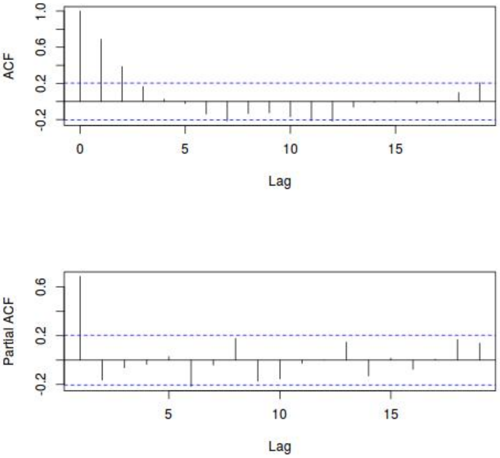
Summary and examples

Daily temperature, Uppsala, summer 1984.

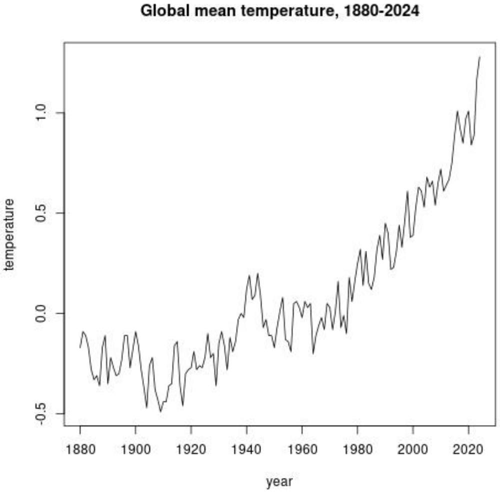


Summary and examples

Daily temperature, Uppsala, summer 1984, ACF and PACF. AR(1)?

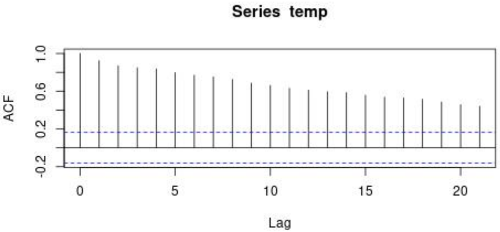


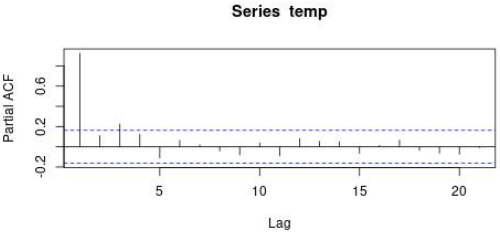
Summary and examples



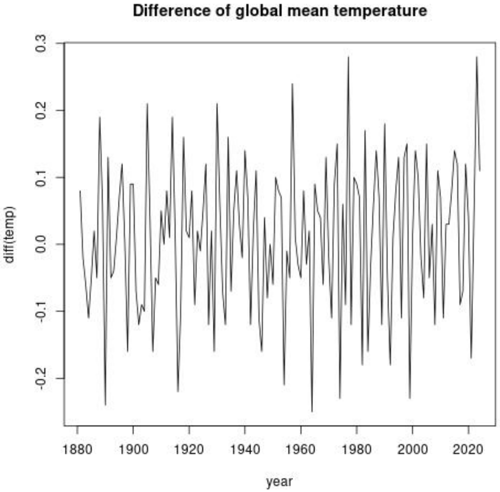
Summary and examples

Global mean temperature, ACF and PACF. (Indicates a trend.)



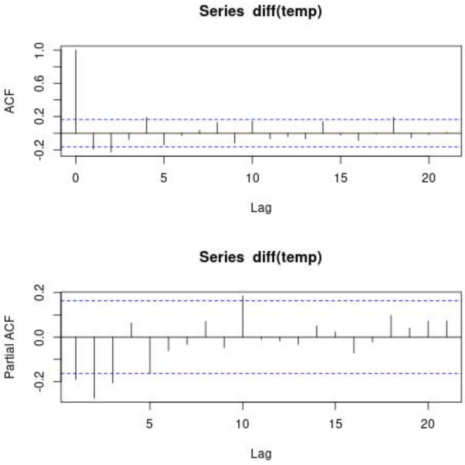


Summary and examples



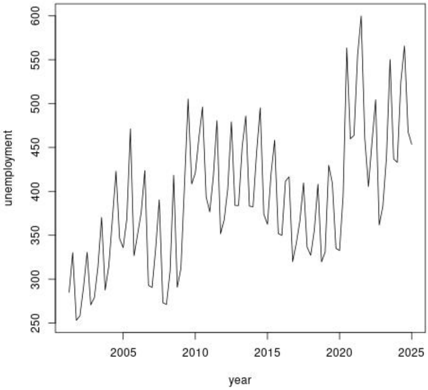
Summary and examples

Difference of global mean temperature, ACF and PACF.
MA(2), MA(4), AR(3) or AR(5)...?



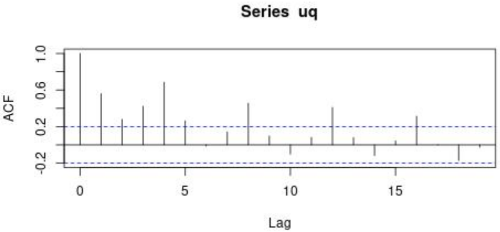
Summary and examples

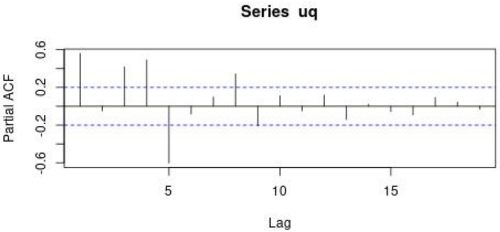
Quarterly unemployment:



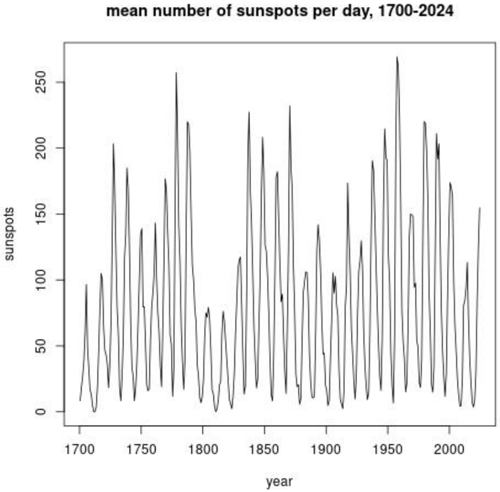
Summary and examples

Quarterly unemployment, ACF and PACF. MA(4) or AR(5)? Period of 4?



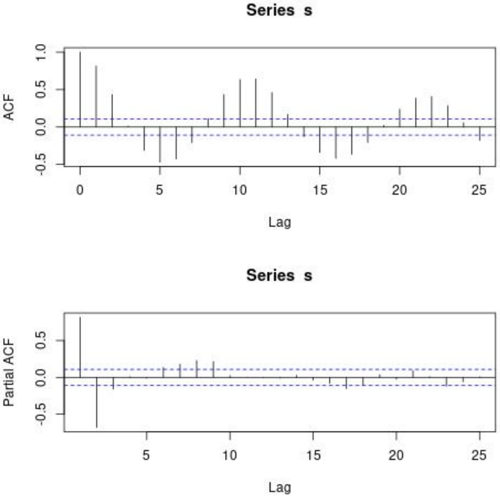


Summary and examples



Summary and examples

Sunspots, ACF and PACF. ARMA? Period of 10 or 11?



News of today

- Definition of the partial autocorrelation function (PACF).
- ACF and PACF for
  - MA processes
  - AR processes
  - ARMA processes
- Model identification, examples

Analysis of Time Series, L6

Rolf Larsson

Uppsala University

4 april 2025

Today

3.5: Forecasting

- The best linear prediction
- The Durbin-Levinson algorithm
- ARMA processes

The best linear prediction

AR(p):

- We observe $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t$ for $t = 1, 2, ..., n$.
- Assume that $\phi_1, \phi_2, ..., \phi_p$ are *known*.
- Predict x_{n+1}.
- A natural predictor is

$$\phi_1 x_n + \phi_2 x_{n-1} + ... + \phi_p x_{n-p+1}.$$

- Is it optimal in some sense?

The best linear prediction

- Assume that $\{x_t\}$ is stationary with known model parameters.
- We observe $x_1, x_2, ..., x_n$.
- Predict x_{n+m} by the linear predictor

$$\begin{aligned} x_{n+m}^n &= \alpha_0 + \alpha_n x_n + \alpha_{n-1} x_{n-1} + ... + \alpha_1 x_1 \\ &= \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k. \end{aligned}$$

- Mean square error $E\{(x_{n+m} - x_{n+m}^n)^2\}$.

Theorem (Property 3.3)

The linear predictor that minimizes the mean square error is found by solving

$$E\{(x_{n+m} - x_{n+m}^n)x_k\} = 0, \quad k = 0, 1, ..., n$$

where $x_0 = 1$, *for* $\alpha_0, \alpha_1, ..., \alpha_n$.

The best linear prediction

- Assume WLOG that $E(x_t) = \mu = 0$, i.e. $\alpha_0 = 0$.
- Find the optimal one-step predictor

$$x_{n+1}^{n} = \phi_{n1}x_n + \phi_{n2}x_{n-1} + \ldots + \phi_{nn}x_1.$$

- Solve (why?)

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \ldots & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{nn} \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{pmatrix}$$

- i.e. $\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n$.

The best linear prediction

- $\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n$.
- For ARMA models, it can be shown that Γ_n is positive definite.
- Hence, $\boldsymbol{\phi}_n = \Gamma_n^{-1} \boldsymbol{\gamma}_n$, i.e.

$$\begin{pmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{nn} \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(0) \end{pmatrix}^{-1} \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{pmatrix}$$

- Derive the one-step predictors for zero mean causal AR(1) and AR(2) processes with one and two observations, respectively.

The best linear prediction

- Problem 3.45: For a zero mean causal AR(p),

$$x_{n+1}^n = \phi_1 x_n + \phi_2 x_{n-1} + \ldots + \phi_p x_{n+1-p}.$$

- In general, for a zero mean causal process,

$$\begin{aligned} x_{n+1}^n &= \phi_{n1} x_n + \phi_{n2} x_{n-1} + \ldots + \phi_{nn} x_1 \\ &= \begin{pmatrix} \phi_{n1} & \phi_{n2} & \ldots & \phi_{nn} \end{pmatrix} \begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_1 \end{pmatrix} = \boldsymbol{\phi}_n' \boldsymbol{x}. \end{aligned}$$

- Mean square prediction error (see the book, p.102)

$$P_{n+1}^n = E\{(x_{n+1} - x_{n+1}^n)^2\} = \gamma(0) - \boldsymbol{\gamma}_n' \Gamma_n^{-1} \boldsymbol{\gamma}_n.$$

The Durbin-Levinson algorithm

- We have derived $x_{n+1}^n = \phi_n' \boldsymbol{x} = \gamma_n' \Gamma_n^{-1} \boldsymbol{x}$.
- Is it possible to calculate x_{n+1}^n without inverting Γ_n?

Theorem (Property 3.4, Problem 3.13)

ϕ_n and P_{n+1}^n may be found iteratively as follows:

$$\phi_{00} = 0, \quad P_1^0 = \gamma_0.$$

For $n \geq 1$,

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(k)}, \quad P_{n+1}^n = P_n^{n-1}(1 - \phi_{nn}^2),$$

where for $n \geq 2$,

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn}\phi_{n-1,n-k}, \quad k = 1, 2, ..., n-1.$$

The Durbin-Levinson algorithm

Theorem (Property 3.5, Problem 3.13)

The PACF of a stationary process $\{x_t\}$ is given by ϕ_{nn}.

Example 1:
Use the Durbin-Levinson algorithm to obtain the PACF of an AR(2) process!

The Durbin-Levinson algorithm

- Generalization to prediction more than one step is straightforward, see the book.
- For predicting MA processes it is easier to use *the innovations algorithm* (property 3.6).
- Read about it by yourself!

ARMA processes

- Recall: For a zero mean causal AR(p), the optimal one-step predictor is
$$x_{n+1}^n = \phi_1 x_n + \phi_2 x_{n-1} + \dots + \phi_p x_{n+1-p}.$$
- Idea: For causal ARMA processes (including MA), why not just use the AR representation $x_t + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \dots = w_t$, i.e.
$$x_t = w_t - \pi_1 x_{t-1} - \pi_2 x_{t-2} - \dots$$
for prediction?
- Problem: This requires an infinite past!
- *But* for large samples, by truncation it can be used as an approximation.
- In particular, it can also be used as an approximation for pure MA processes.

ARMA processes

One step ahead, forecast:

- Replace the minimum mean square error predictor

$$x_{n+1}^n = E(x_{n+1}|x_n, ..., x_1)$$

  by

$$\tilde{x}_{n+1} = E(x_{n+1}|x_n, ..., x_1, x_0, x_{-1}, ...).$$

- The AR form

$$x_{n+1} = w_{n+1} - \pi_1 x_n - \pi_2 x_{n-1} - ... = w_{n+1} - \sum_{j=1}^{\infty} \pi_j x_{n+1-j}$$

- implies (take $E(\cdot|x_n, ..., x_1, x_0, x_{-1}, ...)$)

$$\tilde{x}_{n+1} = -\pi_1 x_n - \pi_2 x_{n-1} - ... = -\sum_{j=1}^{\infty} \pi_j x_{n+1-j}.$$

ARMA processes

One step ahead, prediction error:

- For invertible processes, the MA representation

$$x_{n+1} = w_{n+1} + \psi_1 w_n + \psi_2 w_{n-1} + \ldots = w_{n+1} + \sum_{j=1}^{\infty} \psi_j w_{n+1-j}$$

- implies (take $E(\cdot|x_n, \ldots, x_1, x_0, x_{-1}, \ldots)$)

$$\tilde{x}_{n+1} = \psi_1 w_n + \psi_2 w_{n-1} + \ldots = \sum_{j=1}^{\infty} \psi_j w_{n+1-j}.$$

- Hence, $x_{n+1} - \tilde{x}_{n+1} = w_{n+1}$, and the mean square prediction error is

$$P_{n+1}^n = E\{(x_{n+1} - \tilde{x}_{n+1})^2\} = E(w_{n+1}^2) = \sigma_w^2.$$

ARMA processes

$m \geq 2$ steps ahead, forecast:

- Calculate
$$\tilde{x}_{n+m} = E(x_{n+m}|x_n, ..., x_1, x_0, x_{-1}, ...).$$
- The AR form
$$\begin{aligned} x_{n+m} &= w_{n+m} - \pi_1 x_{n+m-1} - \pi_2 x_{n+m-2} - \dots \\ &= w_{n+m} - \sum_{j=1}^{m-1} \pi_j x_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j} \end{aligned}$$
- implies (take $E(\cdot|x_n, ..., x_1, x_0, x_{-1}, ...)$)
$$\tilde{x}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j}.$$

ARMA processes

$m \geq 2$ steps ahead, prediction error:

- For invertible processes, the MA representation ($\psi_0 = 1$)

$$\begin{aligned} x_{n+m} &= w_{n+m} + \psi_1 w_{n+m-1} + \psi_2 w_{n+m-2} + \dots \\ &= \sum_{j=0}^{m-1} \psi_j w_{n+m-j} + \sum_{j=m}^{\infty} \psi_j w_{n+m-j} \end{aligned}$$

- $E(w_{n+m-j}|x_n, \dots, x_1, x_0, x_{-1}, \dots) = 0$ if $m > j$ and w_{n+m-j} if $m \leq j$
- implies $\tilde{x}_{n+m} = \sum_{j=m}^{\infty} \psi_j w_{n+m-j}$.
- Hence, the mean square prediction error (MSPE) is

$$\begin{aligned} P_{n+m}^n &= E\{(x_{n+m} - \tilde{x}_{n+m})^2\} \\ &= E\left\{\left(\sum_{j=0}^{m-1} \psi_j w_{n+m-j}\right)^2\right\} = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2. \end{aligned}$$

ARMA processes

To summarize:

- Forecast

$$\tilde{x}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j}$$

- MSPE

$$P_{n+m}^n = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2$$

- Prediction interval

$$\tilde{x}_{n+m} \pm c_{\alpha/2} \sqrt{P_{n+m}^n}$$

- For normal (Gaussian) processes with $\alpha = 0.05$, $c_{\alpha/2} = 1.96$.

ARMA processes

Example 2: Let

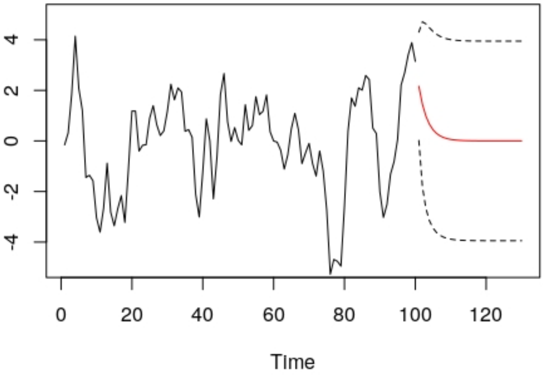
$$x_t = 0.8x_{t-1} + w_t + 0.4w_{t-1}.$$

We observe $x_1, x_2, ..., x_n$, where $n = 100$, $x_{95} = 0.06$, $x_{96} = 2.21$, $x_{97} = 2.70$, $x_{98} = 3.40$, $x_{99} = 3.89$, $x_{100} = 3.15$.

1. Calculate the forecast $\tilde{x}_{n+m}$ for $m = 1, 2, 3$.
2. Calculate the MSPE.
3. Calculate the $1 - \alpha = 0.95$ prediction interval assuming that $\{w_t\}$ is a normal process with $\sigma_w^2 = 1$.

ARMA processes

Example 2: Plot of x_t with forecast horizon 30 and pointwise two standard deviations prediction intervals (obtained from estimated parameters).



In R:

```
> x=arima.sim(list(order=c(1,0,1),ar=0.8,ma=0.4),100)
> m=arima(x,order=c(1,0,1),include.mean=FALSE)
> m

Call:
arima(x = x, order = c(1, 0, 1), include.mean = FALSE)

Coefficients:
         ar1     ma1
      0.6721  0.4730
s.e.  0.0838  0.0969

sigma^2 estimated as 1.148:  log likelihood = -149.51,  aic = 305.02

> fore=predict(m,n.ahead=30)
> ts.plot(x,fore$pred,col=1:2,ylim=c(-5,5))
> lines(fore$pred+2*fore$se,lty='dashed')
> lines(fore$pred-2*fore$se,lty='dashed')
```

News of today

- Optimal prediction of AR processes
  - by matrix inversion
  - by iteration
- Optimal prediction of ARMA processes with infinite past (useful as approximations even if not infinite past).
- Note that all of this required *known* parameters.

Analysis of Time Series, L7

Rolf Larsson

Uppsala University

7 april 2025

Today

3.6: Estimation

- Method of moments
- Maximum likelihood

Method of moments

AR(p):

- We observe $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t$ for $t = 1, 2, \ldots, n$.
- Estimate $\phi_1, \phi_2, \ldots, \phi_p$ and $\sigma_w^2 = \mathrm{var}(w_t)$.

Definition (3.10)

The *Yule-Walker equations* are given by

$$\begin{aligned} \gamma(h) &= \phi_1 \gamma(h-1) + \ldots + \phi_p \gamma(h-p), \quad h = 1, 2, \ldots, p, \\ \sigma_w^2 &= \gamma(0) - \phi_1 \gamma(1) - \ldots - \phi_p \gamma(p). \end{aligned}$$

Method of moments

- The system

$$\gamma(h) = \phi_1\gamma(h-1) + ... + \phi_p\gamma(h-p), \quad h = 1, 2, ..., p$$

  is equivalent to

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix}$$

  i.e. $\Gamma_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p$.

- Method of moments estimators $\hat{\boldsymbol{\phi}} = \hat{\Gamma}_p^{-1}\hat{\boldsymbol{\gamma}}_p$.

Method of moments

Calculate the moment estimators of the AR/MA parameters for

1. The AR(1) process $x_t = \phi x_{t-1} + w_t$.
2. The AR(2) process $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$.
3. The MA(1) process $x_t = w_t + \theta w_{t-1}$. (Two solutions!)

Method of moments

Theorem (Property 3.8)

Let $\hat{\phi}$ be the vector of moment estimators for a causal AR process. Then, as $n \to \infty$,

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(\mathbf{0}, \sigma_w^2 \Gamma_p^{-1}).$$

Hence, $\hat{\phi} \approx N(\phi, n^{-1}\sigma_w^2 \Gamma_p^{-1})$ for large n. (Proof in Appendix B.)

For causal AR processes, the moment estimators are *asymptotically efficient* in the sense that they attain the minimal asymptotic variance.

Theorem (Property 3.9)

For a causal AR(p) process, the PACF fulfills, as $n \to \infty$,

$$\sqrt{n}\hat{\phi}_{hh} \xrightarrow{d} N(0,1), \quad \textit{for } h > p.$$

Method of moments

Calculate the asymptotic variances (and covariances) of the AR parameter moment estimators for

1. The AR(1) process $x_t = \phi x_{t-1} + w_t$.
2. The AR(2) process $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$.

Maximum likelihood

Causal AR(1) without constant: $x_t = \phi x_{t-1} + w_t$, $t = 1, 2, ..., n$, $w_t \sim N(0, \sigma_w^2)$, independent.

- Log likelihood (why?)

$$l(\phi, \sigma_w^2) = -\frac{n}{2}\log(2\pi\sigma_w^2) + \frac{1}{2}\log(1-\phi^2) - \frac{S(\phi)}{2\sigma_w^2},$$

  where

$$S(\phi) = (1-\phi^2)x_1^2 + \sum_{t=2}^{n}(x_t - \phi x_{t-1})^2.$$

- No “simple” explicit expression for the MLE of ϕ.

Maximum likelihood

Causal AR(1) without constant: $x_t = \phi x_{t-1} + w_t$, $t = 1, 2, ..., n$, $w_t \sim N(0, \sigma_w^2)$, independent.

- *Conditional* log likelihood (the x_1 density 'disappears')

$$l(\phi, \sigma_w^2 | x_1) = -\frac{n-1}{2} \log(2\pi\sigma_w^2) - \frac{S_c(\phi)}{2\sigma_w^2},$$

  where

$$S_c(\phi) = \sum_{t=2}^{n} (x_t - \phi x_{t-1})^2.$$

- Conditional MLEs (give zero partial derivatives, cf linear regression)

$$\hat{\sigma}_w^2 = \frac{S_c(\hat{\phi})}{n-1}, \quad \hat{\phi} = \frac{\sum_{t=2}^{n} x_t x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2}.$$

Maximum likelihood

Causal AR(p) without constant:

$$x_t = \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + w_t, \quad t = 1, 2, ..., n,$$

$w_t \sim N(0, \sigma_w^2)$, independent.

- Conditional log likelihood

$$l(\phi_1, ..., \phi_p, \sigma_w^2 | x_1, ..., x_p) = -\frac{n-p}{2} \log(2\pi\sigma_w^2) - \frac{S_c}{2\sigma_w^2},$$

where

$$S_c = \sum_{t=p+1}^{n} (x_t - \phi_1 x_{t-1} - ... - \phi_p x_{t-p})^2.$$

- Find the conditional MLEs!

Maximum likelihood

Causal and invertible ARMA(p, q) without constant:

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}, \quad t = 1, 2, \ldots, n,$$

$w_t \sim N(0, \sigma_w^2)$, independent.

- Not possible to find explicit conditional (or unconditional) MLEs (see p.118-119).
- Use numerical methods! (p.119-122).

Maximum likelihood

Causal and invertible ARMA(p, q) with constant, $t = 1, 2, ..., n$:

$x_t - \mu = \phi_1(x_{t-1} - \mu) + ... + \phi_p(x_{t-p} - \mu) + w_t + \theta_1 w_{t-1} + ... + \theta_q w_{t-q}$,

i.e. $\phi(B)(x_t - \mu) = \theta(B) w_t$.

Let $\boldsymbol{\beta} = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q)'$. Let $\hat{\boldsymbol{\beta}}$ be the vector of MLEs.

Theorem (Property 3.10)

1. *Under appropriate conditions, as $n \to \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \sigma_w^2 \Gamma_{p,q}^{-1}),$$

where

$$\Gamma_{p,q} = \begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta} \end{pmatrix}.$$

2. *$\hat{\boldsymbol{\beta}}$ is asymptotically efficient.*

Maximum likelihood

$$\Gamma_{p,q} = \begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta} \end{pmatrix},$$

where

$$\Gamma_{\phi\phi} = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{pmatrix},$$

where $\gamma(h)$ is the autocovariance function of $\phi(B)x_t = w_t$.

$\Gamma_{\theta\theta}$ is similarly constructed from $\theta(B)y_t = w_t$.

$\Gamma_{\phi\theta}$ is similarly constructed from the cross covariance function between $\phi(B)x_t = w_t$ and $\theta(B)y_t = w_t$, and $\Gamma_{\theta\phi}$ is analogous.

Maximum likelihood

Calculate the asymptotic variances (and covariances) of the ML parameter estimators for

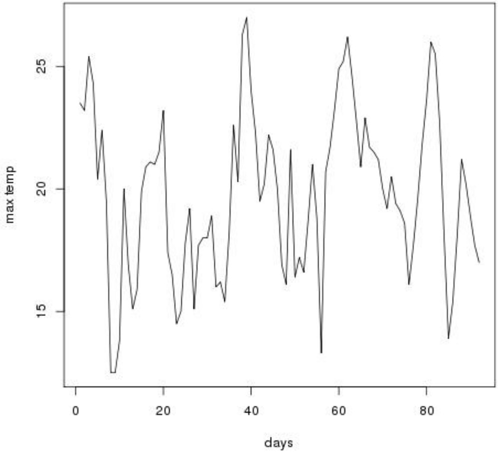
1. The MA(1) process $x_t = w_t + \theta w_{t-1}$.
2. The ARMA(1,1) process $x_t = \phi x_{t-1} + w_t + \theta w_{t-1}$.

Maximum likelihood

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

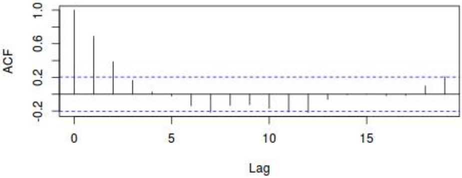
Maximum likelihood

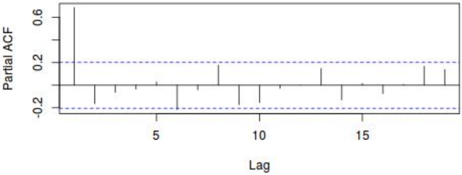
Daily temperature, Uppsala, summer 1984.



Maximum likelihood

Daily temperature, Uppsala, summer 1984, ACF (tails off) and PACF (cuts off after lag 1). Try AR(1)!

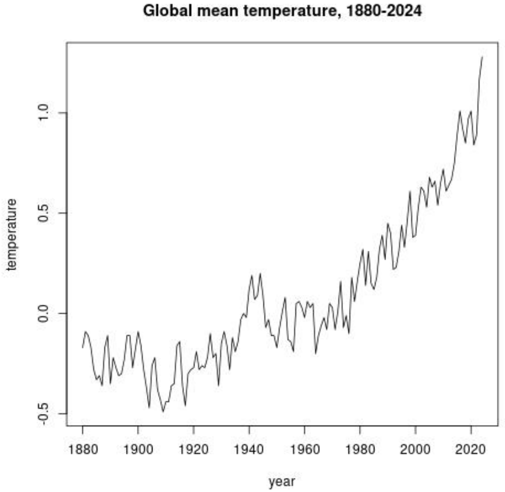




In R: (note: $\hat{\phi} \approx 0.69$ is outside the $\pm 2\cdot$ s.e. bound.)

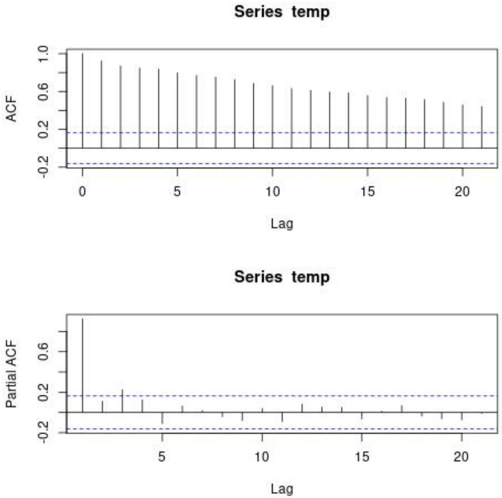
```
> x=read.table("tempUasom84.txt")$V1
> plot(x,type='l',xlab='days',ylab='max temp')
> par(mfrow=c(2,1))
> acf(x,main='')
> pacf(x,main='')
> arima(x,order=c(1,0,0))

Call:
arima(x = x, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.6946    19.7769
s.e.  0.0745     0.8053

sigma^2 estimated as 5.839:  log likelihood = -212.04,
aic = 430.08
```

Maximum likelihood

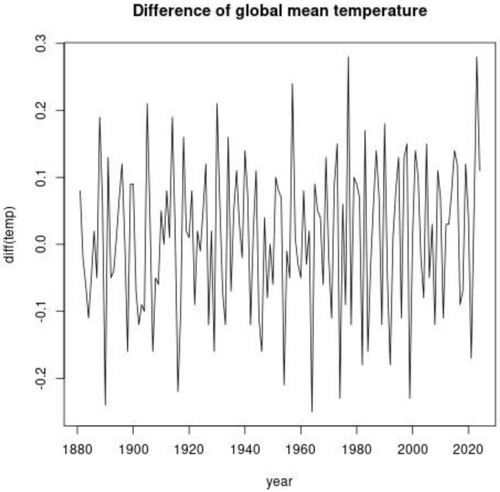


Maximum likelihood

Global mean temperature, ACF and PACF. (Typical signs of a trend.)

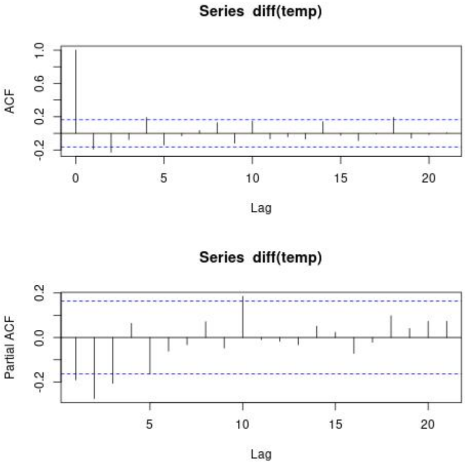


Maximum likelihood



Maximum likelihood

Difference of global mean temperature, ACF (cuts off after lag 2?) and PACF (tails off?). Try MA(2)!



Estimation in R with ML:

```
> arima(temp,order=c(0,1,2))

Call:
arima(x = temp, order = c(0, 1, 2))

Coefficients:
          ma1      ma2
      -0.3010  -0.2118
s.e.   0.0783   0.0702

sigma^2 estimated as 0.01106:  log likelihood = 119.84,
aic = -233.68
```

Estimation in R with CSS (similar results):

```
> arima(temp,order=c(0,1,2),method = "CSS")

Call:
arima(x = temp, order = c(0, 1, 2), method = "CSS")

Coefficients:
          ma1      ma2
      -0.3034  -0.2139
s.e.   0.0782   0.0703

sigma^2 estimated as 0.01106:  part log likelihood = 119.96
```

News of today

- Method of moments, Yule Walker equations
- Maximum likelihood/Least squares
  - Conditional on initial values
  - Unconditional

Analysis of Time Series, L8

Rolf Larsson

Uppsala University

8 april 2025

Today

- 3.7: Integrated models
- 3.8: Building ARIMA models

Integrated models

Recall:

Definition (2.4)

The *backshift operator* is defined by

$$Bx_t = x_{t-1}.$$

For $k = 1, 2, \ldots,$ $B^k x_t = x_{t-k}$.

Definition (2.5)

Differences of order d are defined by

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

Special cases:

- $\nabla^1 x_t = \nabla x_t = (1 - B)x_t = x_t - x_{t-1}$
- $\nabla^2 x_t = (1 - B)^2 x_t = x_t - 2x_{t-1} + x_{t-2}$

Integrated models

- Two types of trends:
  1. Deterministic (example: $x_t = \beta_0 + \beta_1 t + w_t$)
  2. Stochastic (example: $x_t = x_{t-1} + w_t$)
- May be removed by differencing.
- Watch out for overdifferencing!
- Example: $x_t = x_{t-1} + w_t \Rightarrow \nabla x_t = w_t \Rightarrow \nabla^2 x_t = \nabla w_t = w_t - w_{t-1}$. $\mathrm{corr}(\nabla x_{t+1}, \nabla x_t) = 0$, $\mathrm{corr}(\nabla^2 x_{t+1}, \nabla^2 x_t) = -1/2 \neq 0$.
- Too much differencing may introduce extra autocorrelations and non invertibility!

Integrated models

Example:

- Let

$$x_t = \beta_0 + \beta_1 t + \beta_2 t^2 + y_t,$$

where $\beta_2 \neq 0$ and y_t is stationary.

- How many differences are required to make x_t stationary?

Integrated models

Definition (3.11)

A process $\{x_t\}$ is said to be ARIMA(p, d, q) if

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

is ARMA(p, q). We may write the model as

$$\phi(B)(1 - B)^d x_t = \theta(B) w_t.$$

If $E(\nabla^d x_t) = \mu$, we write the model as

$$\phi(B)(1 - B)^d x_t = \delta + \theta(B) w_t,$$

where

$$\delta = \mu(1 - \phi_1 - \dots - \phi_p).$$

Why?

Integrated models

Prediction (based on infinite past):

- Write the process on AR form to obtain the forecast.
- Write the process on MA form to obtain the prediction error.
- Example: Forecasting IMA(1,1)

$$\nabla x_t = w_t - \lambda w_{t-1}.$$

Leads to the Holt and Winter method:

$$\tilde{x}_{n+1} = (1-\lambda)x_n + \lambda \tilde{x}_n.$$

Why?

Building ARIMA models

1. Check if the series is stationary. If not:
    - Remove trends by differencing.
    - Make the variance constant by transformations.
2. Identify an ARMA model by looking at
    - ACF
    - PACF
    - Information criteria (AIC, BIC...)
3. Estimate the model.
4. Check the model fit by performing residual diagnostics.
5. If the model does not fit well, start over from 1.

Building ARIMA models

- Make the variance constant!
- Box-Cox transformation (chap.2, p.59)

$$y_t = \begin{cases} \frac{x_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log x_t, & \lambda = 0. \end{cases}$$

- Modified (not in book):

$$\tilde{y}_t = \mathrm{gm}(x)^{1-\lambda} y_t,$$

where $\mathrm{gm}(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$.

- By Taylor expansion (why?),

$$\frac{x_t^\lambda - 1}{\lambda} = \log x_t + O(\lambda).$$

Building ARIMA models

Model identification via ACF and PACF:

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Building ARIMA models

Let $\hat{\sigma}_k^2 = SSE_k/n$, where SSE_k is the residual sum of squares and k is the number of parameters in the model.

Definition (Akaike Information Criterion)

$$\mathrm{AIC} = \log(\hat{\sigma}_k^2) + \frac{2k}{n}.$$

In R: $\mathrm{AIC} = -2\log(L) + 2k$.

Definition (Bayesian Information Criterion)

$$\mathrm{BIC} = \log(\hat{\sigma}_k^2) + \frac{k\log(n)}{n}.$$

- BIC is recommended for large samples, AIC for small samples.
- It is *not recommended* to compare AIC, BIC between ARIMA models with different d, or for different transformations.

Building ARIMA models

- Test H_0: The residuals are white noise vs H_1: $\neg H_0$.
- Ljung-Box-Pierce Q statistic (p.139)

$$Q = n(n+2)\sum_{h=1}^{H} \frac{\hat{\rho}_e^2(h)}{n-h},$$

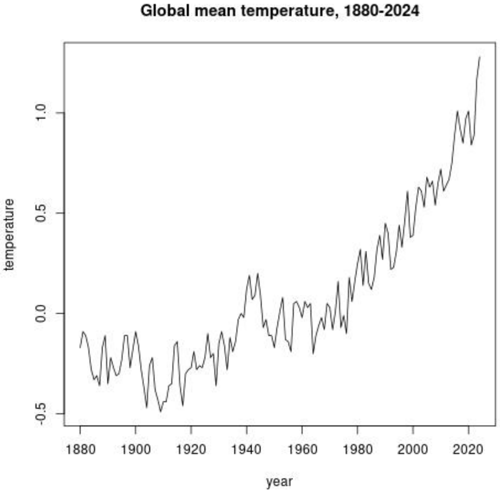
where $\hat{\rho}_e(h)$ are the estimated autocorrelations of the residuals.

- Asymptotically, $Q \sim \chi^2_{H-p-q}$.

Building ARIMA models

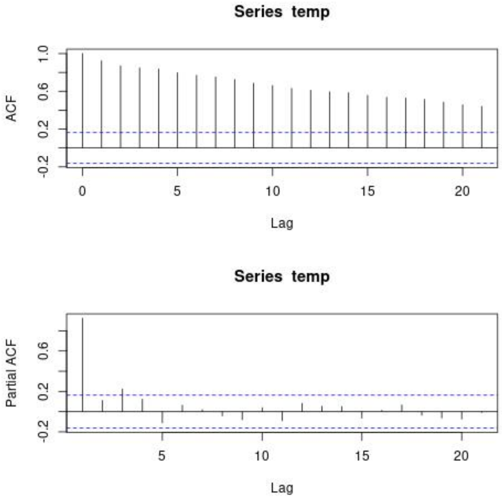
1. Check if the series is stationary. If not:
    - Remove trends by differencing.
    - Make the variance constant by transformations.
2. Identify an ARMA model by looking at
    - ACF
    - PACF
    - Information criteria (AIC, BIC...)
3. Estimate the model.
4. Check the model fit by performing residual diagnostics.
5. If the model does not fit well, start over from 1.

Building ARIMA models

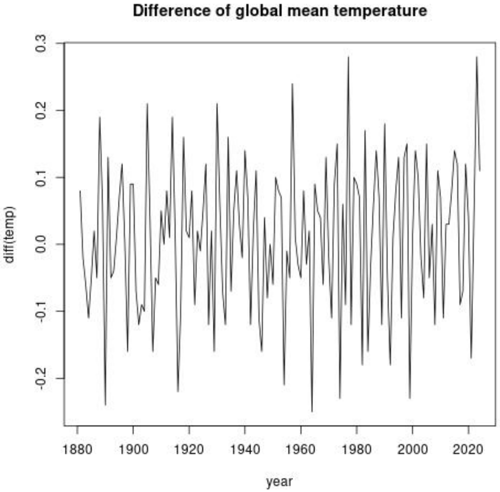


Building ARIMA models

Global mean temperature, ACF and PACF (typical signs of a trend):

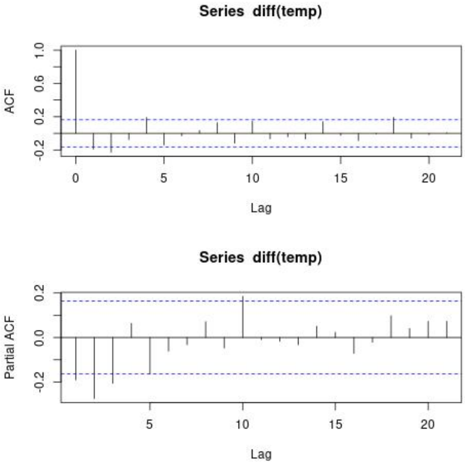


Building ARIMA models



Building ARIMA models

Difference of global mean temperature, ACF (cuts off after lag 2) and PACF (tails off?):



In R, try MA(2) for differences (ARIMA(0,1,2)):

```
> arima(temp,order=c(0,1,2))

Call:
arima(x = temp, order = c(0, 1, 2))

Coefficients:
          ma1      ma2
      -0.3010  -0.2118
s.e.   0.0783   0.0702

sigma^2 estimated as 0.01106:  log likelihood = 119.84,
aic = -233.68
```

Among many models, ARIMA(3,1,1) gave the smallest AIC.

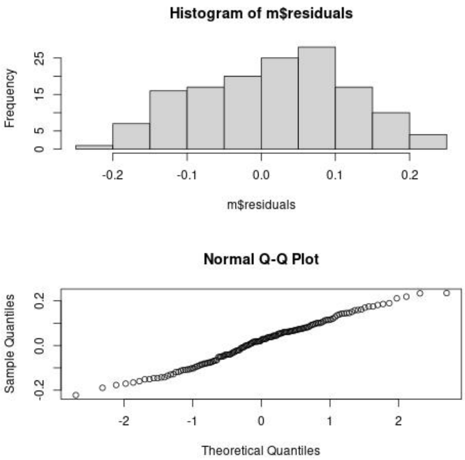
```
> arima(temp,order=c(3,1,1))

Call:
arima(x = temp, order = c(3, 1, 1))

Coefficients:
          ar1      ar2      ar3     ma1
      -1.0247  -0.5030  -0.3750  0.7904
s.e.   0.1566   0.1125   0.0795  0.1607

sigma^2 estimated as 0.01053:  log likelihood = 123.27,
aic = -236.55
```

Check: all coefficients are significant, i.e. outside the two times s.e. bound.

Residual diagnostics, ARIMA(3,1,1):

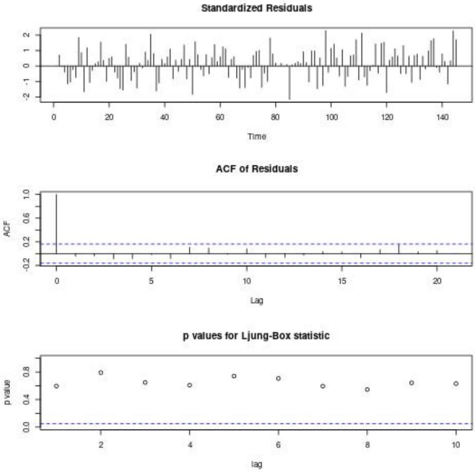
```
> m=arima(temp,order=c(3,1,1))
> par(mfrow=c(2,1))
> hist(m$residuals)
> qqnorm(m$residuals)
> dev.off()
> tsdiag(m)
```

Building ARIMA models



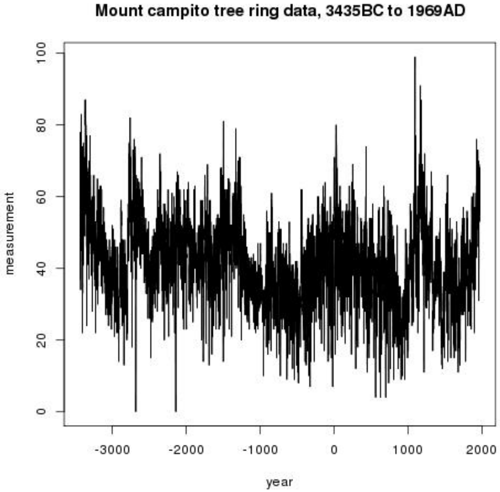
Normal distribution ok (?)

Building ARIMA models



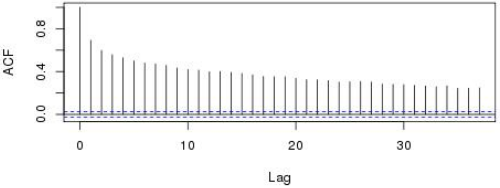
No sign of autocorrelations in the residuals!

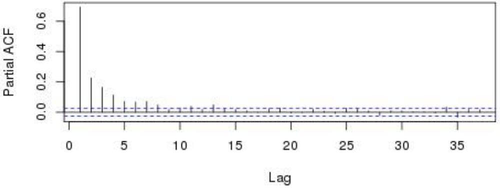
Building ARIMA models



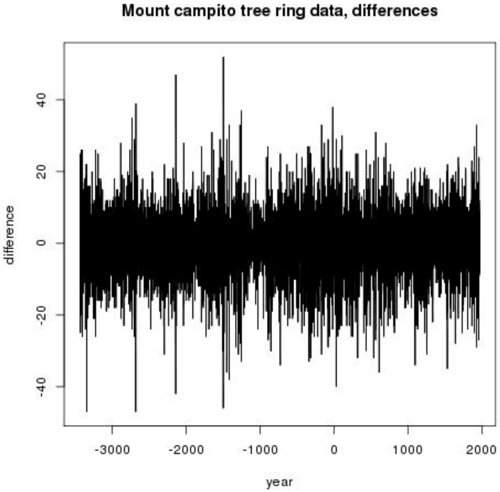
Building ARIMA models

Mount Campito, ACF and PACF. Maybe not stationary?



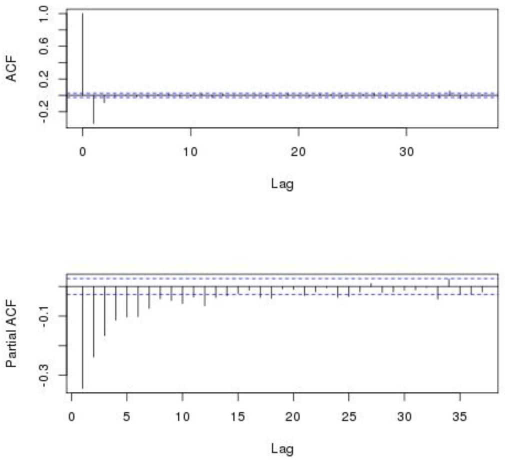


Building ARIMA models



Building ARIMA models

Mount Campito differences, ACF and PACF. MA(2)?



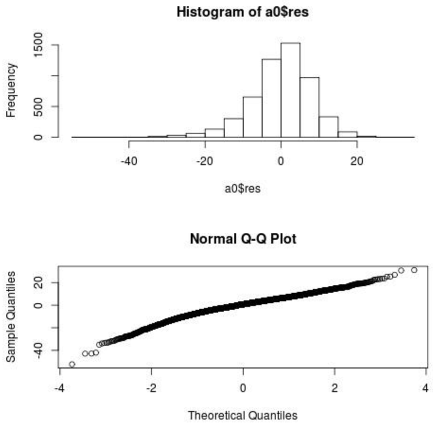
MA(2) for differences, estimation in R (intercept not significant):

```
> dy=diff(y)
> a=arima(dy,order=c(0,0,2));a

Call:
arima(x = dy, order = c(0, 0, 2))

Coefficients:
          ma1      ma2  intercept
      -0.5449  -0.1921     0.0009
s.e.   0.0130   0.0140     0.0289

sigma^2 estimated as 65.34:  log likelihood = -18961.64,
aic = 37931.27
```

MA(2) without constant (note: AIC two units lower):

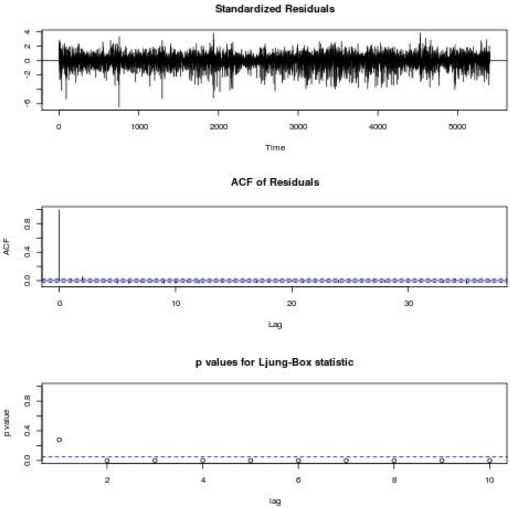
```
> a0=arima(dy,order=c(0,0,2),include.mean=FALSE);a0

Call:
arima(x = dy, order = c(0, 0, 2), include.mean = FALSE)

Coefficients:
          ma1      ma2
      -0.5449  -0.1921
s.e.   0.0130   0.0140

sigma^2 estimated as 65.34:  log likelihood = -18961.64,
aic = 37929.27
```

```
> par(mfrow=c(2,1))
> hist(a0$res)
> qqnorm(a0$res)
```



Skew distribution!?

```
> tsdiag(a0)
```



Small p values, i.e. significant autocorrelations.

Building ARIMA models

MountCampito: Find the ARMA(p, q) model for differences without constant with the smallest AIC:

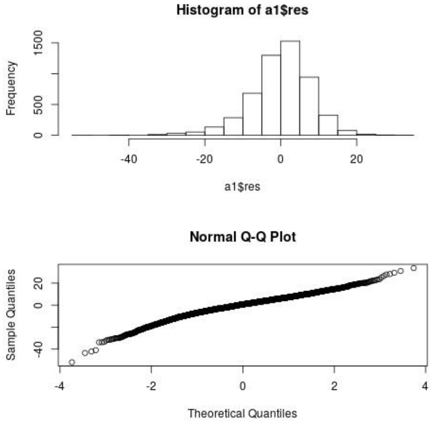
p	q	AIC
0	0	39339.2
0	1	38107.1
1	0	38657.1
0	2	37929.3
1	1	37885.3
2	0	38341.7
0	3	37901.2
1	2	37854.4
2	1	37872.9
3	0	38191.7
0	4	37887.8
1	3	37818.9
2	2	**37815.9**
3	1	37851.8
4	0	38123.2
2	3	37826.7
3	2	37818.0
3	3	37819.8

Try ARMA$(2,2)$:

```
> a1=arima(dy,order=c(2,0,2),include.mean=FALSE);a1

Call:
arima(x = dy, order = c(2, 0, 2), include.mean = FALSE)

Coefficients:
         ar1      ar2      ma1     ma2
      1.1512  -0.2216  -1.7007  0.7059
s.e.  0.0367   0.0260   0.0322  0.0313

sigma^2 estimated as 63.93:  log likelihood = -18902.95,
aic = 37815.9
```

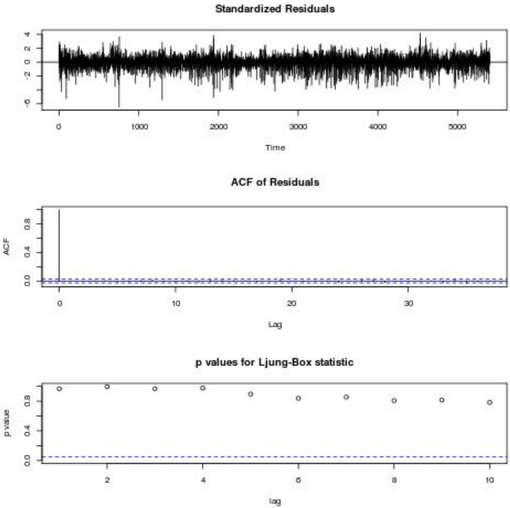
Observe: The invertibility condition $\theta_2 - \theta_1 < 1$ is not satisfied!

```
> par(mfrow=c(2,1))
> hist(a1$res)
> qqnorm(a1$res)
```



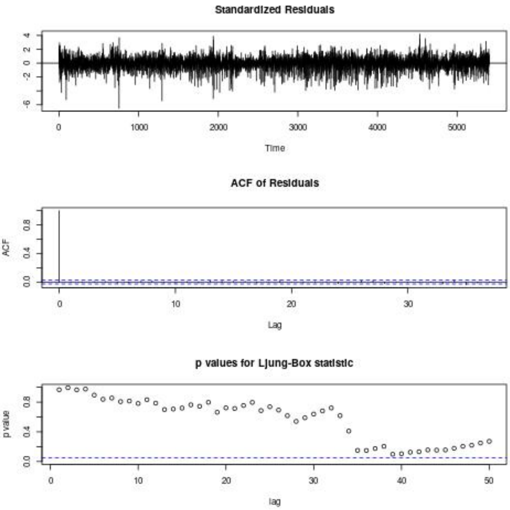
Still a bit skew!

```
> tsdiag(a1)
```



Better on autocorrelations, but try more lags for Ljung-Box!

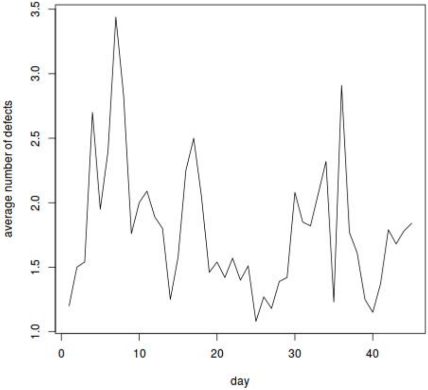
```
> tsdiag(a1,50)
```



Ljung-Box a bit suspicious here (p values for high lags close to 0.05)...

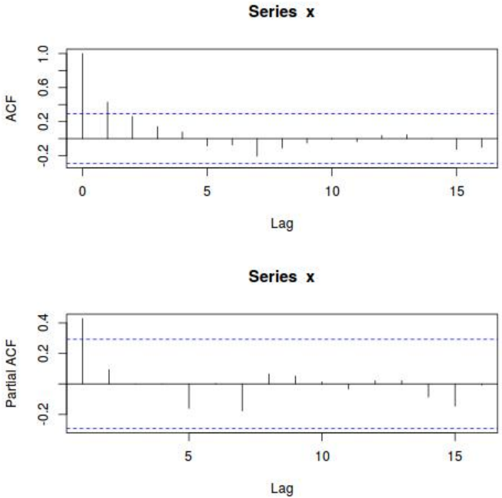
Building ARIMA models

Daily average number of defects per truck found in the final inspection at the end of the assembly line of a truck manufacturing plant.
(Looks stationary.)



Building ARIMA models

Trucks, ACF (tails off) and PACF (cuts off after lag 1). AR(1)?

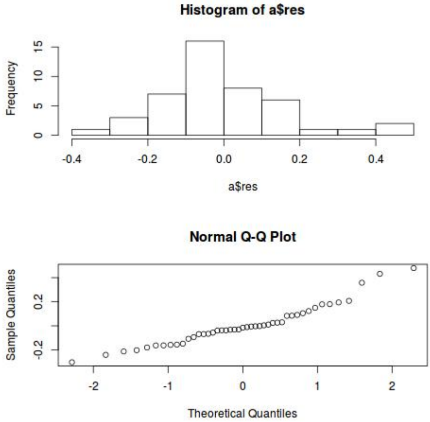


Try AR(1):

```
> a=arima(x,order=c(1,0,0));a

Call:
arima(x = x, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.4322     1.7799
s.e.  0.1340     0.1189

sigma^2 estimated as 0.2118:  log likelihood = -29.04,
aic = 64.07
```

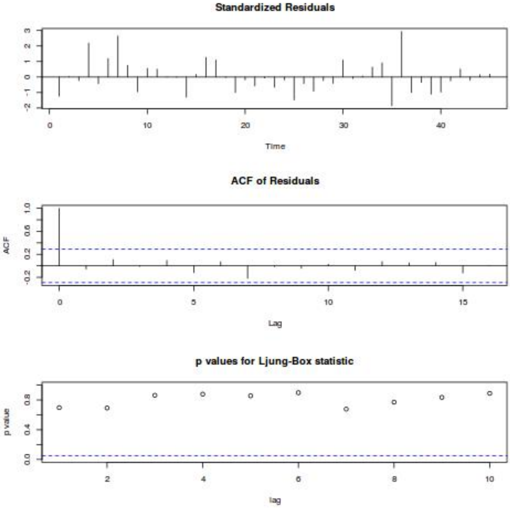
Significant coefficients!

```
> par(mfrow=c(2,1))
> hist(a$res)
> qqnorm(a$res)
```



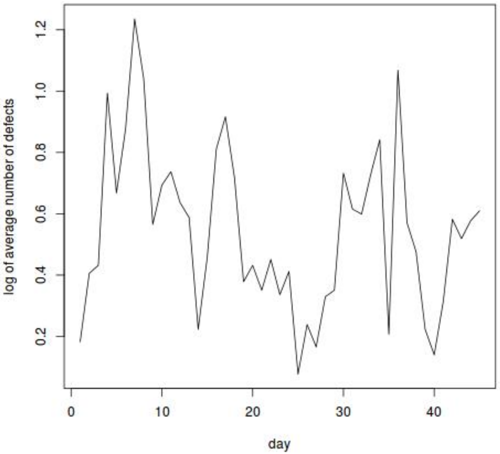
Maybe some outliers to the right?

```
> tsdiag(a)
```


Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

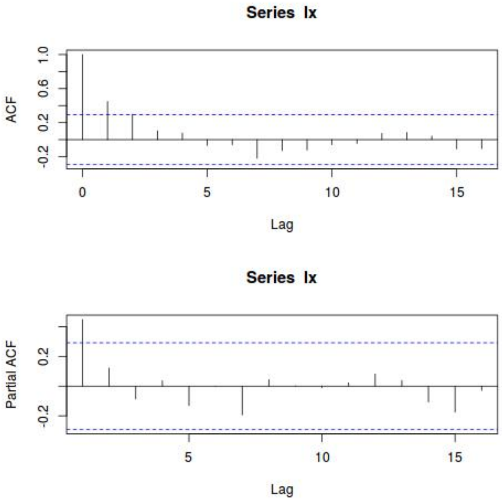
Building ARIMA models

Take logarithms (to get a smaller effect of outliers):



Building ARIMA models

Trucks in logs, ACF and PACF

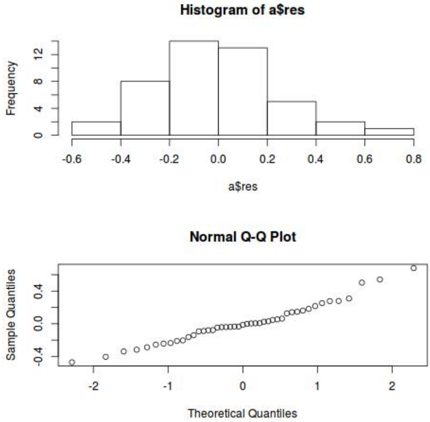


Try AR(1):

```
> lx=log(x)
> a=arima(lx,order=c(1,0,0));a

Call:
arima(x = lx, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.4582     0.5391
s.e.  0.1330     0.0641

sigma^2 estimated as 0.05625:  log likelihood = 0.78,
aic = 4.43
```

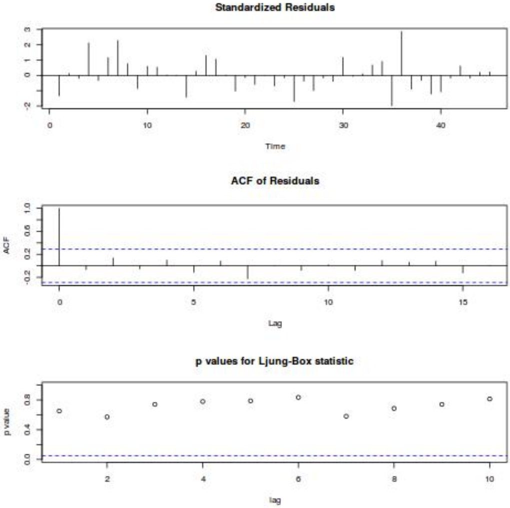
Do not compare AIC to the previous value!

```
> par(mfrow=c(2,1))
> hist(a$res)
> qqnorm(a$res)
```



Less pronounced outliers now.

```
> tsdiag(a)
```


Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

News of today

- Removing trends by differencing
- Model building:
    - Transformation
    - Identification
    - Estimation
    - Diagnostics

Analysis of Time Series, L9

Rolf Larsson

Uppsala University

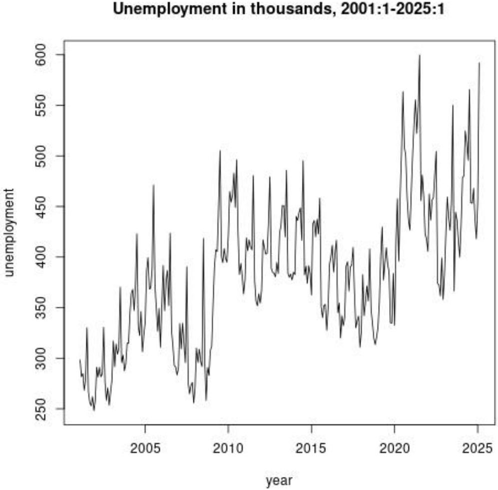
9 april 2025

Today

3.9: Multiplicative Seasonal ARIMA (SARIMA) Models

- SARIMA models
- Empirical examples
- Menti

SARIMA models



SARIMA models

- Recall: ARMA(p, q)

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

- or equivalently $\phi(B)x_t = \theta(B)w_t$ where

$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p,$$
$$\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q.$$

- *Pure seasonal ARMA model*, ARMA$(P, Q)_s$

$$x_t = \Phi_1 x_{t-s} + \ldots + \Phi_P x_{t-Ps} + w_t + \Theta_1 w_{t-s} + \ldots + \Theta_Q w_{t-Qs}$$

- or equivalently $\Phi_P(B^s)x_t = \Theta_Q(B^s)w_t$ where

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \ldots - \Phi_p B^{Ps},$$
$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \ldots + \Theta_Q B^{Qs}.$$

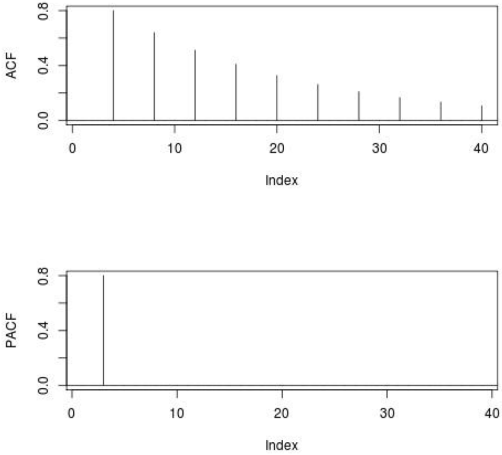
SARIMA models

Calculate the ACF of

1. $\text{ARMA}(0, 1)_4$
2. $\text{ARMA}(1, 0)_4$

SARIMA models

ACF and PACF for $x_t = 0.8x_{t-4} + w_t$



SARIMA models

- Pure seasonal ARMA$(P, Q)_s$ model

$$\Phi_P(B^s)x_t = \Theta_Q(B^s)w_t$$

where

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \ldots - \Phi_p B^{Ps},$$
$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \ldots + \Theta_Q B^{Qs}.$$

- Multiplicative (mixed) ARMA$(p, q) \times (P, Q)_s$ model

$$\Phi_P(B^s)\phi(B)x_t = \Theta_Q(B^s)\theta(B)w_t$$

where in addition

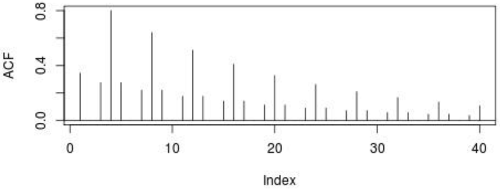
$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p,$$
$$\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q.$$

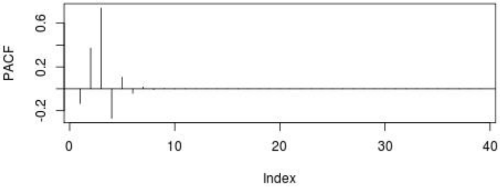
SARIMA models

1. Write down the ARMA$(0,1) \times (1,0)_4$ model explicitly.
2. Derive its ACF.

SARIMA models

ACF and PACF for $x_t = 0.8x_{t-4} + w_t + 0.4w_{t-1}$





SARIMA models

Seasonal difference

$$\nabla_s x_t = (1 - B^s)x_t = x_t - x_{t-s}$$

Definition (3.12)

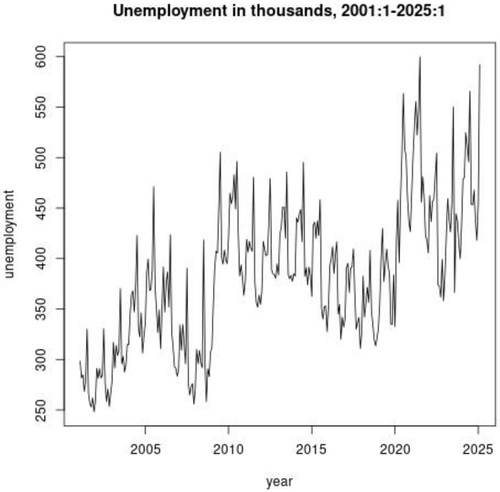
The multiplicative *seasonal autoregressive integrated moving average* (SARIMA) model is given by

$$\Phi_P(B^s)\phi(B)\nabla_s^D \nabla^d x_t = \delta + \Theta_Q(B^s)\theta(B)w_t,$$

where w_t is white noise.
It is denoted by ARIMA$(p, d, q) \times (P, D, Q)_s$.

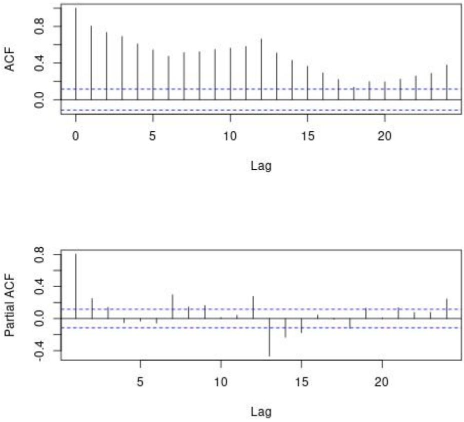
Write down the ARIMA$(0, 1, 1) \times (1, 1, 0)_4$ model without constant explicitly.

Empirical examples



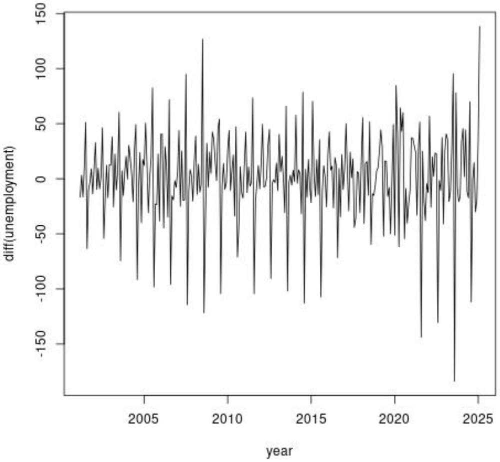
Empirical examples

Unemployment, ACF and PACF. Slowly decaying ACF.
Try to take a difference!



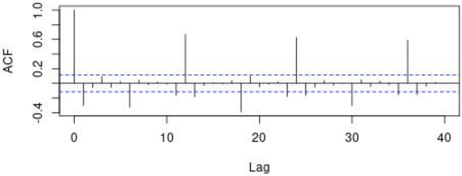
Empirical examples

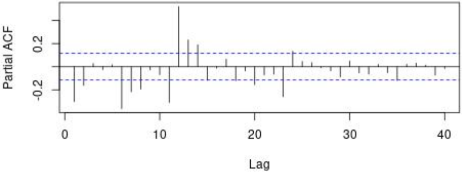
Plot of the differenced series:



Empirical examples

Differenced series, ACF and PACF. Slowly decaying ACF at lags 12, 24, 36,... (R: `acf(diff(u),40)` etc.) Try a seasonal difference.



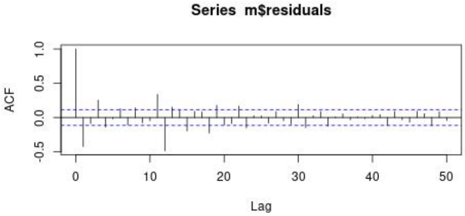


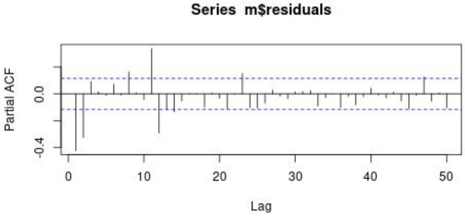
"Empty" seasonally differenced model:

```
> m=arima(u,order=c(0,1,0),seasonal=list(order=c(0,1,0),period=12));m

Call:
arima(x = u, order = c(0, 1, 0), seasonal = list(order = c(0, 1, 0), period = 12))


sigma^2 estimated as 1054:  log likelihood = -1352.2,  aic = 2706.39
```

ACF and PACF for residuals: Seasonal MA(1) behavior?

```
> acf(m$residuals,50)
> pacf(m$residuals,50)
```

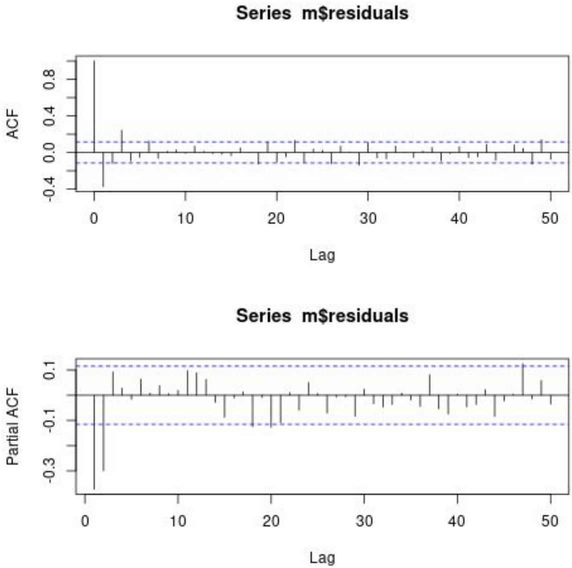




Try ARIMA$(0,1,0) \times (0,1,1)_{12}$ (smallest AIC of purely seasonal models).

```
> m=arima(u,order=c(0,1,0),seasonal=list(order=c(0,1,1),period=12));m

Call:
arima(x = u, order = c(0, 1, 0), seasonal = list(order = c(0, 1, 1), period = 12))

Coefficients:
         sma1
      -0.8687
s.e.   0.0562

sigma^2 estimated as 626.1:  log likelihood = -1288.7,  aic = 2581.4
```

ACF and PACF for residuals: AR(2) behavior?

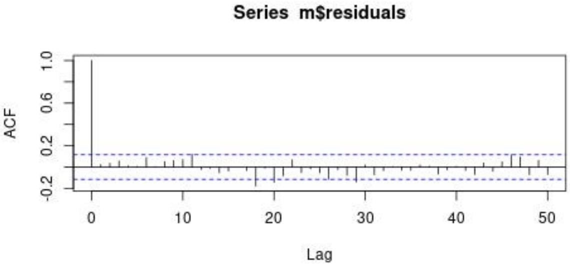


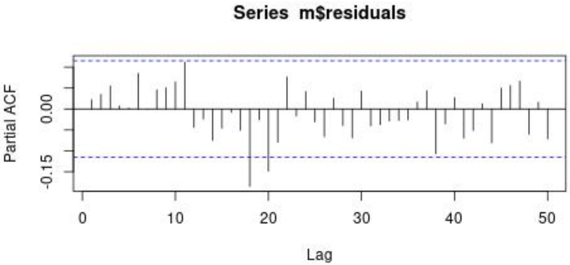
Try ARIMA$(2,1,0) \times (1,1,1)_{12}$. (Smaller AIC than the 'closest' models.)

```
> m=arima(u,order=c(2,1,0),seasonal=list(order=c(1,1,1),period=12));m

Call:
arima(x = u, order = c(2, 1, 0), seasonal = list(order = c(1, 1, 1), period = 12))

Coefficients:
          ar1      ar2    sar1     sma1
      -0.5384  -0.3427  0.1075  -0.8991
s.e.   0.0592   0.0587  0.0748   0.0574

sigma^2 estimated as 471.3:  log likelihood = -1250.03,  aic = 2510.07
```

Observe: the `sar1` coefficient is not significant.

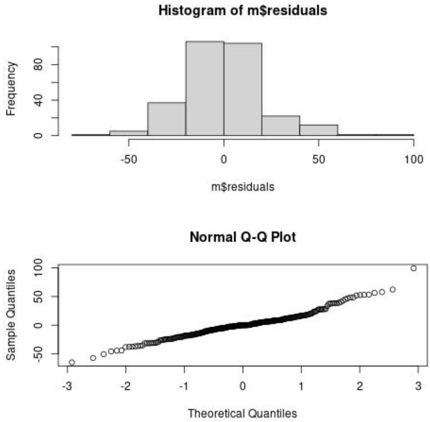
ACF and PACF for residuals: White noise behavior?





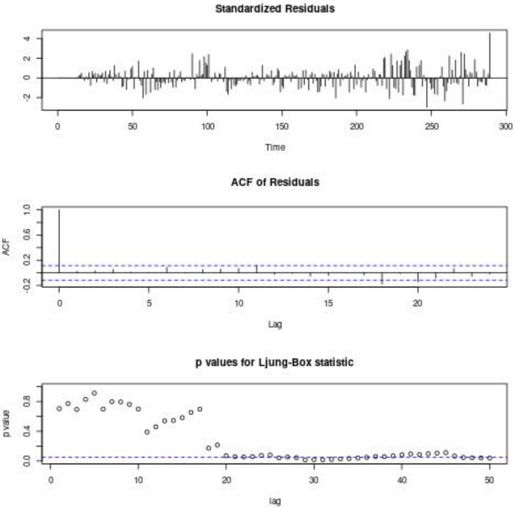
Histogram and qqplot for residuals:

```
> par(mfrow=c(2,1))
> hist(m$residuals)
> qqnorm(m$residuals)
```



Diagnostics for residuals: Too small P values!

```
> tsdiag(m,50)
```

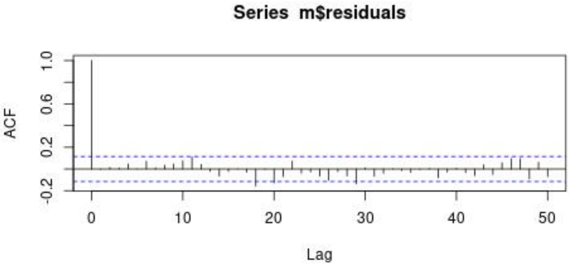


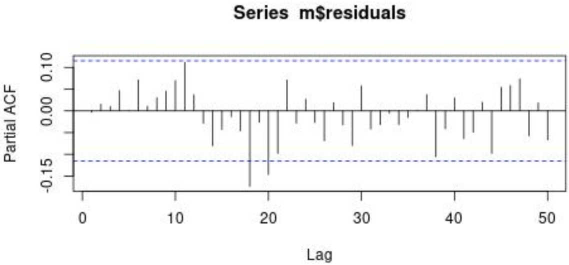
Instead, try ARIMA$(3,1,0) \times (0,1,1)_{12}$. (A little bit higher AIC, but maybe better diagnostics?)

```
> m=arima(u,order=c(3,1,0),seasonal=list(order=c(0,1,1),period=12));m

Call:
arima(x = u, order = c(3, 1, 0), seasonal = list(order = c(0, 1, 1), period = 12))

Coefficients:
          ar1      ar2     ar3     sma1
      -0.5065  -0.3016  0.0704  -0.8534
s.e.   0.0625   0.0679  0.0624   0.0532

sigma^2 estimated as 475.8:  log likelihood = -1250.42,  aic = 2510.85
```

Observe that the `ar3` coefficient is not significant.
(But removing it gave worse Ljung-Box P values than for this model.)

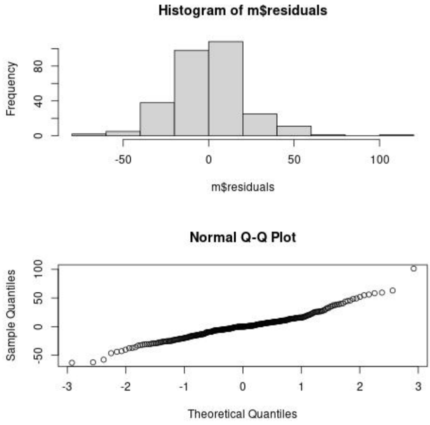
ACF and PACF for residuals: White noise behavior?





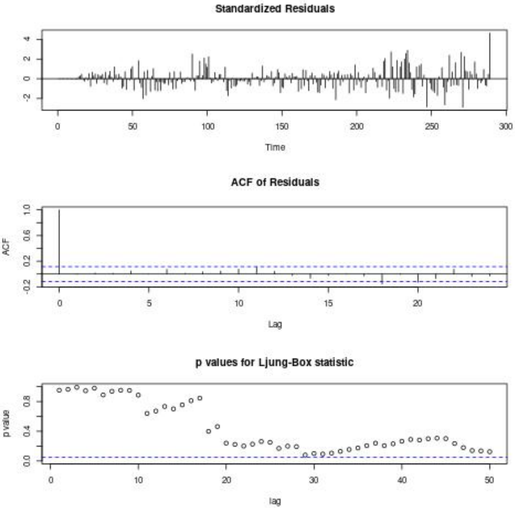
Histogram and qqplot for residuals:

```
> par(mfrow=c(2,1))
> hist(m$residuals)
> qqnorm(m$residuals)
```

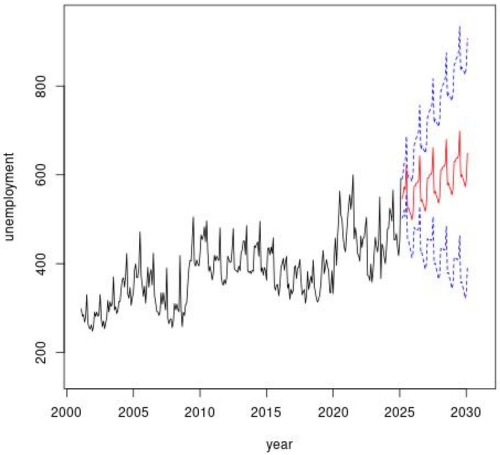


Diagnostics for residuals: Better P values!

```
> tsdiag(m,50)
```



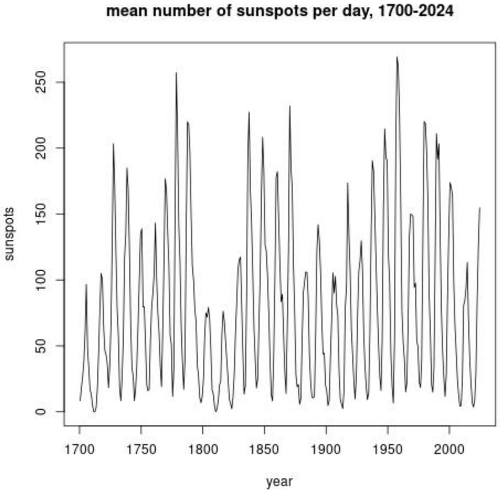
Plot of forecasts (5 years ahead):



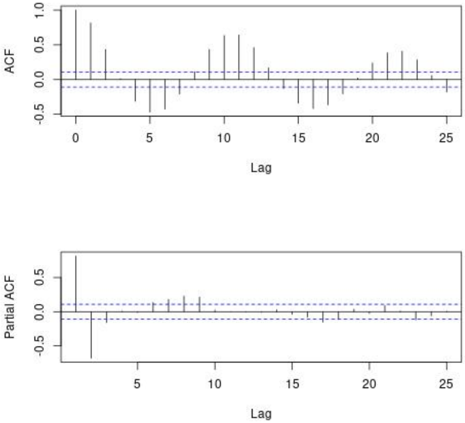
R code for this plot:

```
> fore=predict(m,n.ahead=60)
> plot(seq(2001+1/12,2025+1/12,1/12),u,type='l',xlab='year',
xlim=c(2001,2031),ylim=c(150,950),ylab='unemployment')
> lines(seq(2025+2/12,2030+1/12,1/12),fore$pred,col='red')
> lines(seq(2025+2/12,2030+1/12,1/12),fore$pred+2*fore$se,col='blue',
lty='dashed')
> lines(seq(2025+2/12,2030+1/12,1/12),fore$pred-2*fore$se,col='blue',
lty='dashed')
```

Empirical examples



Empirical examples

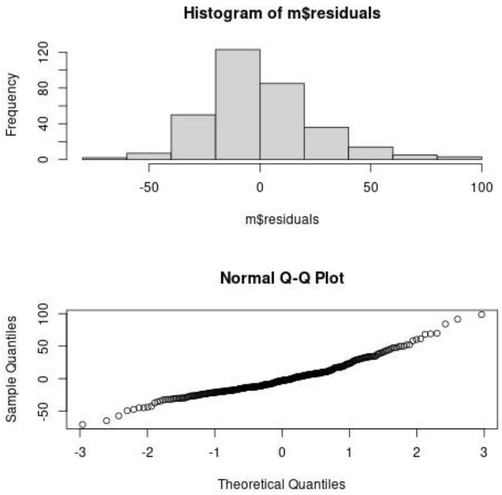


10-11 year season. AR(3) structure of PACF?

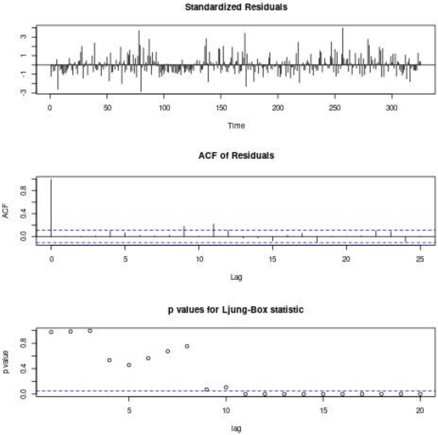
Try ARIMA$(3,0,0) \times (0,0,1)_{10}$ (smallest AIC among similar models):

```
> m=arima(s,order=c(3,0,0),seasonal=list(order=c(0,0,1),period=10));m

Call:
arima(x = s, order = c(3, 0, 0), seasonal = list(order = c(0, 0, 1), period = 10))

Coefficients:
         ar1      ar2      ar3    sma1  intercept
      1.2052  -0.4028  -0.1670  0.2309    78.5312
s.e.  0.0587   0.0866   0.0548  0.0581     4.6094

sigma^2 estimated as 611.7:  log likelihood = -1505.21,  aic = 3022.41
```

Empirical examples



Empirical examples



Small P values for Ljung-Box! Problem: the season is not exactly 10 years.

News of today

- Seasonal differencing
- SARIMA models

Analysis of Time Series, L10

Rolf Larsson

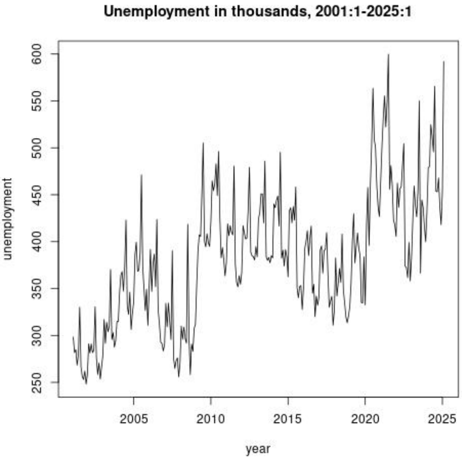
Uppsala University

22 april 2025

Today

- 4.1: Cyclical behaviour and periodicity
- 4.2: The spectral density

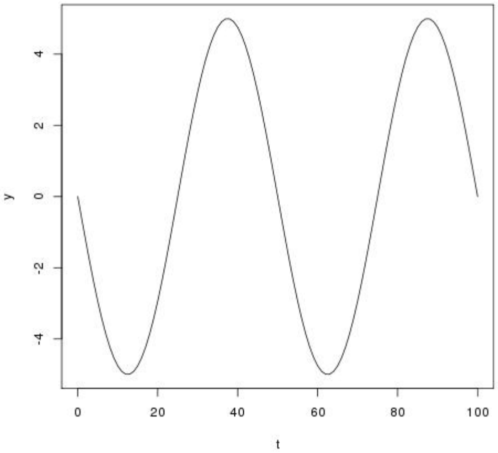
Cyclical behaviour and periodicity



Season length 12 months. Business cycle 10 years?

Cyclical behaviour and periodicity

Plot of the function $y = 5\cos(2\pi t/50 + 0.5\pi)$:



Cyclical behaviour and periodicity

Fourier analysis:

- Let x_t be deterministic, $t = 1, ..., n$, $n = 2m$ where m is an integer.
- We may write

$$x_t = \sum_{k=0}^{m} \left\{ u_{k1} \cos(2\pi t\omega_k) + u_{k2} \sin(2\pi t\omega_k) \right\},$$

where $\omega_k = k/n$, $u_{02} = u_{m2} = 0$ and Fourier coefficients

$$u_{k1} = \begin{cases} \frac{1}{n} \sum_{t=1}^{n} x_t \cos(2\pi t\omega_k), & \text{if } k = 0 \text{ or } k = m, \\ \frac{2}{n} \sum_{t=1}^{n} x_t \cos(2\pi t\omega_k), & k = 1, 2, ..., m-1, \end{cases}$$

$$u_{k2} = \frac{2}{n} \sum_{t=1}^{n} x_t \sin(2\pi t\omega_k), \quad k = 1, 2, ..., m-1.$$

Cyclical behaviour and periodicity

- Let x_t be a mixture of periodic series

$$x_t = \sum_{k=1}^{q} \left\{ U_{k1} \cos(2\pi\omega_k t) + U_{k2} \sin(2\pi\omega_k t) \right\},$$

  where U_{k1}, U_{k2} for $k = 1, 2, ..., q$ are independent with zero mean and variances σ_k^2.
- Autocovariance function (why?)

$$\gamma(h) = \mathrm{cov}(x_{t+h}, x_t) = \sum_{k=1}^{q} \sigma_k^2 \cos(2\pi\omega_k h).$$

- In particular,

$$\gamma(0) = \mathrm{var}(x_t) = \sum_{k=1}^{q} \sigma_k^2.$$

The spectral density

Example 4.4:

- Let

$$x_t = U_1 \cos(2\pi\omega_0 t) + U_2 \sin(2\pi\omega_0 t),$$

where U_1 and U_2 are independent random variables with mean zero and variance σ^2.

- The autocovariance function satisfies (why?)

$$\begin{aligned}\gamma(h) = \sigma^2 \cos(2\pi\omega_0 h) &= \frac{\sigma^2}{2} e^{-2\pi i \omega_0 h} + \frac{\sigma^2}{2} e^{2\pi i \omega_0 h} \\ &= \int_{-1/2}^{1/2} e^{2\pi i \omega h} dF(\omega),\end{aligned}$$

where

$$F(\omega) = \begin{cases} 0 & \omega < -\omega_0, \\ \sigma^2/2 & -\omega_0 \le \omega < \omega_0, \\ \sigma^2, & \omega_0 \le \omega. \end{cases}$$

The spectral density

- By theorem C.1, for *any* autocovariance function for a stationary process, $\gamma(h)$, there is a non decreasing function F with $F(-1/2) = 0$, $F(1/2) = \gamma(0)$, such that

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i\omega h} dF(\omega).$$

- $F(\omega)$ is called *the spectral distribution function*.
- If $F(\omega)$ is differentiable with derivative $f(\omega)$, then

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i\omega h} f(\omega) d\omega,$$

where $f(\omega)$ is called *the spectral density function*.

The spectral density

Theorem (Property 4.2)

If the autocovariance function $\gamma(h)$ *for a stationary process satisfies*

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty,$$

then for $h = 0, \pm 1, \pm 2, \ldots,$

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i \omega h} f(\omega) d\omega$$

where the spectral density f *has the representation*

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h}, \quad -\frac{1}{2} \le \omega \le \frac{1}{2}.$$

The spectral density

Some properties:

1. For all ω, $f(\omega) \geq 0$
2. $f(\omega) = f(-\omega)$
3. $f(\omega) = f(1-\omega)$
4. $\gamma(0) = \int_{-1/2}^{1/2} f(\omega) d\omega$

Calculate the spectral density function of a white noise process!

The spectral density

Recall: If x_t is ARMA(p, q),

$$\phi(B)x_t = \theta(B)w_t,$$

where $\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$, $\theta(B) = 1 + \theta_1 B + ... + \theta_q B^q$.

Theorem (Property 4.4)

If x_t is ARMA(p, q), its spectral density is given by

$$f(\omega) = \sigma_w^2 \frac{|\theta(e^{-2\pi i\omega})|^2}{|\phi(e^{-2\pi i\omega})|^2}.$$

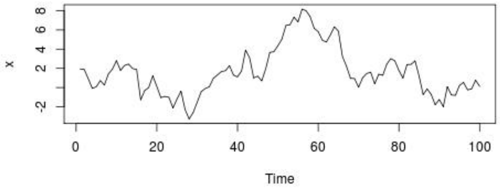
The spectral density

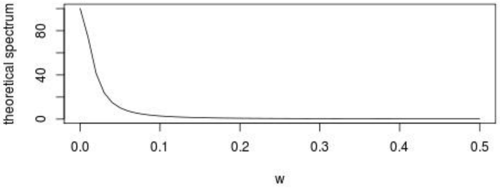
Calculate the spectral density of

1. An MA(1) process
2. An AR(1) process
3. An ARMA(1,1) process
4. A SARMA$(1,0) \times (1,0)_4$ process

The spectral density

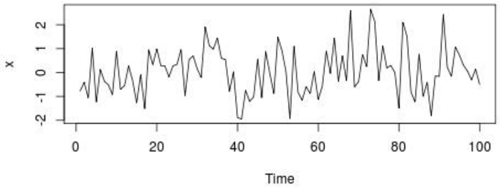
$x_t = 0.9x_{t-1} + w_t$ (smooth, high weight on low frequencies)

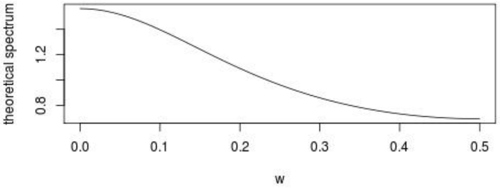




The spectral density

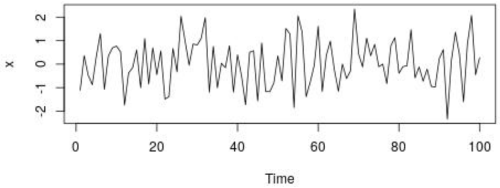
$x_t = 0.2x_{t-1} + w_t$ (less smooth)

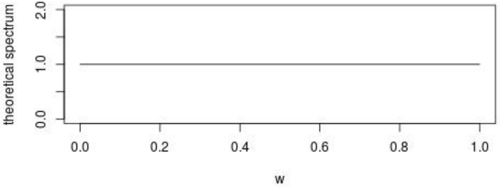




The spectral density

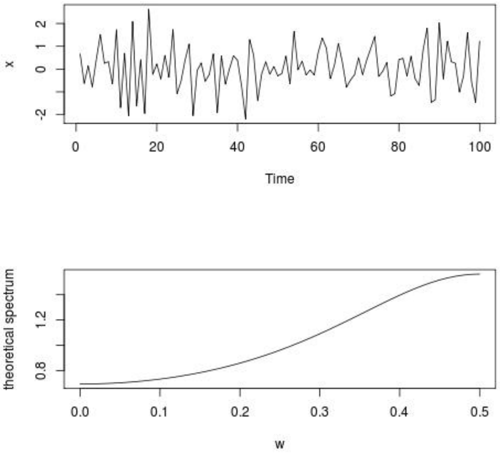
$x_t = w_t$ (all frequencies equally important)





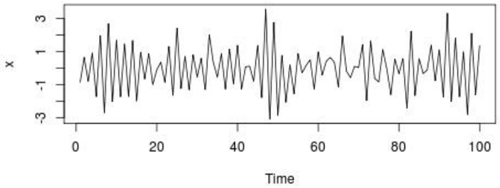
The spectral density

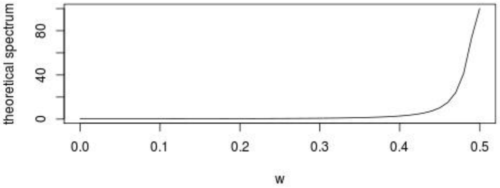
$x_t = -0.2x_{t-1} + w_t$ (more weight on high frequencies)



The spectral density

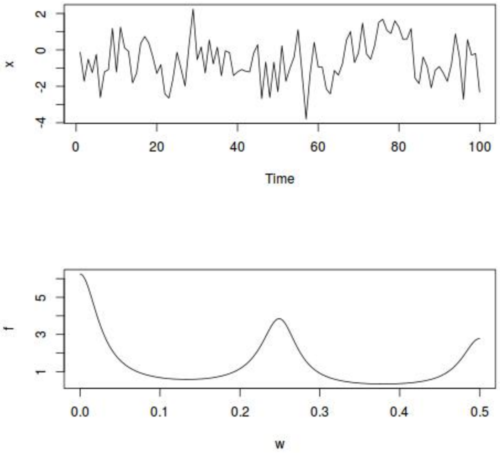
$x_t = -0.9x_{t-1} + w_t$ (wiggly, high weight on high frequencies)





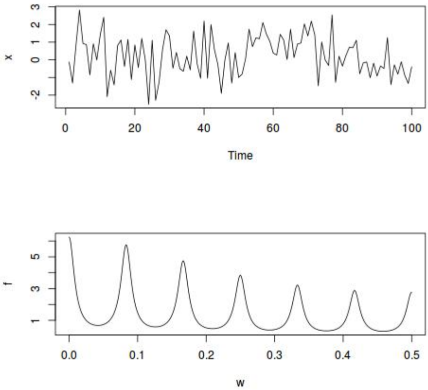
The spectral density

$(1 - 0.2B)(1 - 0.5B^4)x_t = w_t$ (high weight on frequency $1/4 = 0.25$)



The spectral density

$(1 - 0.2B)(1 - 0.5B^{12})x_t = w_t$
(high weights on multiples of frequency $1/12 \approx 0.08$)



News of today

- The spectral distribution
- The spectral density
- The spectral density for an ARMA process

Analysis of Time Series, L11

Rolf Larsson

Uppsala University

23 april 2025

Today

- 4.3: Periodogram and Discrete Fourier Transform (DFT)
- 4.4: Nonparametric Spectral Estimation
- 4.5: Parametric Spectral Estimation
- Menti

Periodogram and DFT

Observations $x_1, ..., x_n$.

Definition (4.1)

The *discrete Fourier transform* (DFT) is defined as

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t e^{-2\pi i \omega_j t}, \quad j = 0, 1, ..., n-1,$$

where the frequencies $\omega_j = j/n$ are called the *Fourier* or *fundamental frequencies*.

Inverse DFT

$$x_t = n^{-1/2} \sum_{j=0}^{n-1} d(\omega_j) e^{2\pi i \omega_j t}.$$

Periodogram and DFT

Observations $x_1, \dots, x_n$, $\omega_j = j/n$.

Definition (4.2)

The *periodogram* is defined as

$$I(\omega_j) = |d(\omega_j)|^2, \quad j = 0, 1, \dots, n-1.$$

- Assume $j \neq 0$. It follows that (why?)

$$I(\omega_j) = \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{-2\pi i \omega_j h}.$$

- Recall: for the spectral density, we have

$$f(\omega_j) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega_j h}.$$

Periodogram and DFT

The periodogram $I(\omega_j)$ is an estimate of the spectral density $f(\omega_j)$. Properties?

- Let $\omega_{j:n}$ be a sequence of fundamental frequencies such that $\omega_{j:n} \to \omega$ as $n \to \infty$.
- Asymptotic unbiasedness: as $n \to \infty$,

$$E\{I(\omega_{j:n})\} \to f(\omega).$$

- Asymptotic distribution: For a linear process, as $n \to \infty$,

$$I(\omega_{j:n}) \xrightarrow{d} \frac{\chi_2^2}{2} f(\omega).$$

- Approximate $1 - \alpha$ confidence interval

$$\frac{2I(\omega_{j:n})}{\chi_2^2\left(1 - \frac{\alpha}{2}\right)} \leq f(\omega) \leq \frac{2I(\omega_{j:n})}{\chi_2^2\left(\frac{\alpha}{2}\right)}$$

Periodogram and DFT

- Let $\omega_{j:n}$ be a sequence of fundamental frequencies such that $\omega_{j:n} \to \omega$ as $n \to \infty$.
- Asymptotic distribution: For a linear process, as $n \to \infty$,

$$I(\omega_{j:n}) \xrightarrow{d} \frac{\chi_2^2}{2} f(\omega).$$

- Hence,

$$\mathrm{var}\{I(\omega_{j:n})\} \to f(\omega)^2 \neq 0.$$

- The periodogram is *not* a consistent estimator of the spectral density!
- The solution is smoothing!

Nonparametric Spectral Estimation

- Smoothed periodogram

$$\bar{f}(\omega) = \frac{1}{L} \sum_{k=-m}^{m} I\left(\omega_j + \frac{k}{n}\right),$$

  where $L = 2m + 1$ and $\omega_j = j/n$ is close to ω.
- For large n,

$$\bar{f}(\omega) \approx \frac{\chi^2_{2L}}{2L} f(\omega).$$

- Hence,

$$\mathrm{var}\{\bar{f}(\omega)\} \approx \frac{1}{L} f(\omega)^2 \to 0 \quad \text{as } L \to \infty.$$

- Dilemma:
  - L large gives small variance, large bias.
  - L small gives small bias, large variance.

Nonparametric Spectral Estimation

- For large n,

$$\bar{f}(\omega) \approx \frac{\chi^2_{2L}}{2L} f(\omega).$$

- Approximate $1 - \alpha$ confidence interval

$$\frac{2L\bar{f}(\omega)}{\chi^2_{2L}\left(1 - \frac{\alpha}{2}\right)} \leq f(\omega) \leq \frac{2L\bar{f}(\omega)}{\chi^2_{2L}\left(\frac{\alpha}{2}\right)}.$$

- Equivalent to

$$\log\{\bar{f}(\omega)\} - \log\left\{\frac{\chi^2_{2L}\left(1 - \frac{\alpha}{2}\right)}{2L}\right\} \leq \log\{f(\omega)\}$$

$$\leq \log\{\bar{f}(\omega)\} + \log\left\{\frac{2L}{\chi^2_{2L}\left(\frac{\alpha}{2}\right)}\right\}.$$

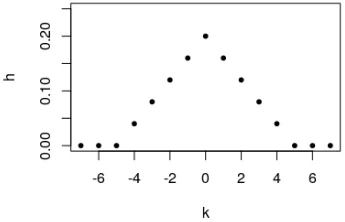
Nonparametric Spectral Estimation

More general:

- Smoothed periodogram

$$\hat{f}(\omega) = \sum_{k=-m}^{m} h_k I\left(\omega_j + \frac{k}{n}\right),$$

  where $\sum_{k=-m}^{m} h_k = 1$.
- Example of $\{h_k\}$:



Nonparametric Spectral Estimation

- Smoothed periodogram, where $\sum_{k=-m}^{m} h_k = 1$,

$$\hat{f}(\omega) = \sum_{k=-m}^{m} h_k I\left(\omega_j + \frac{k}{n}\right),$$

- Assume that if $n, m \to \infty$ such that $\frac{m}{n} \to 0$, then $\sum_{k=-m}^{m} h_k^2 \to 0$.
- If so, then as $n \to \infty$,

$$E\left\{\hat{f}(\omega)\right\} \to f(\omega)$$

and

$$\hat{f}(\omega) \approx \frac{\chi^2_{2L_h}}{2L_h} f(\omega), \quad L_h = \left(\sum_{k=-m}^{m} h_k^2\right)^{-1}.$$

- With $h_k = 1/L$, we have $\sum_{k=-m}^{m} h_k^2 = 1/L$ and $L_h = L$.

Nonparametric Spectral Estimation

- Smoothed periodogram

$$\hat{f}(\omega) = \sum_{|k| \le m} h_k I(\omega_j + k/n)$$

- Inserting $I(\omega) = \sum_{|h|<n} \hat{\gamma}(h) e^{-2\pi i \omega h}$ with $\omega = \omega_j + k/n$ yields (why?)

$$\hat{f}(\omega) = \sum_{h} g(h/n) \hat{\gamma}(h) e^{-2\pi i \omega_j h},$$

where $g(h/n) = \sum_{|k| \le m} h_k e^{-2\pi i k h/n}$.

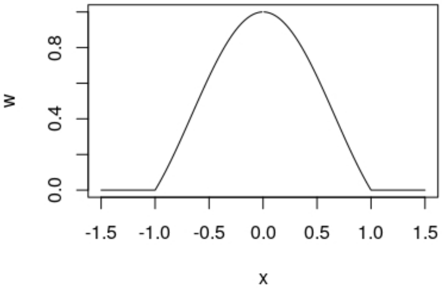
- Suggests estimators of the form

$$\tilde{f}(\omega) = \sum_{|h| \le r} w(h/r) \hat{\gamma}(h) e^{-2\pi i \omega h}.$$

- The function $w(x)$ is called the *lag window*.

Nonparametric Spectral Estimation

- The lag window $w(\cdot)$ has to satisfy
  - (*i*) $w(0) = 1$,
  - (*ii*) $|w(x)| \leq 1$ and $w(x) = 0$ for $|x| > 1$,
  - (*iii*) $w(x) = w(-x)$.
- Example of $w(x)$:



Nonparametric Spectral Estimation

- The *smoothing window*

$$W_r(\omega) = \sum_{|h| \le r} w(h/r) e^{-2\pi i \omega h}.$$

- Inversion formula (not in the book)

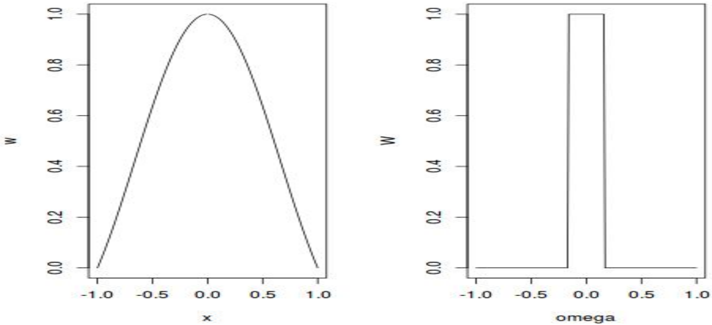
$$w(h/r) = \int_{-1/2}^{1/2} W_r(\omega) e^{2\pi i \omega h} d\omega.$$

- Corresponds to $g(h/n) = \sum_{|k| \le m} h_k e^{-2\pi i k h/n}$,
  i.e $W_r(\omega)$ is the “continuous counterpart” of h_k.
- $W_r(\omega)$ “smooths over frequencies” and $w(x)$ “smooths over lags”.

Nonparametric Spectral Estimation

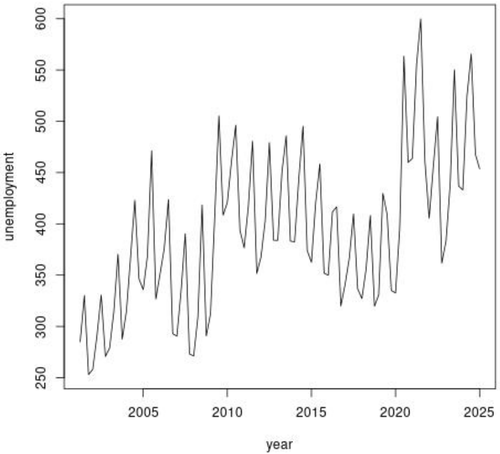
Default in R: The sinc lag window, Daniell smoothing window (here $r = 3$)

$$w(x) = \frac{\sin(\pi x)}{\pi x}, \qquad W_r(\omega) = \begin{cases} r, & |\omega| \le 1/(2r), \\ 0, & \text{otherwise.} \end{cases}$$

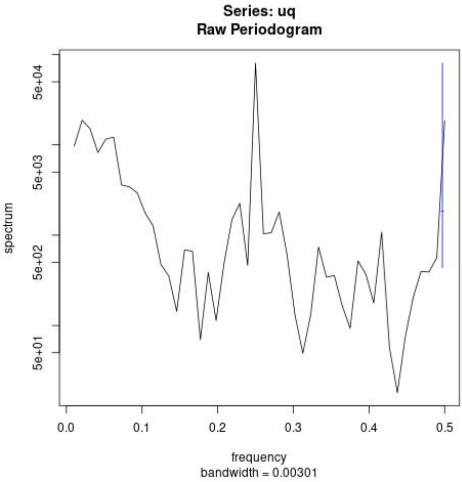


Nonparametric Spectral Estimation

Unemployment, quarterly data (period length 4, and maybe 40).



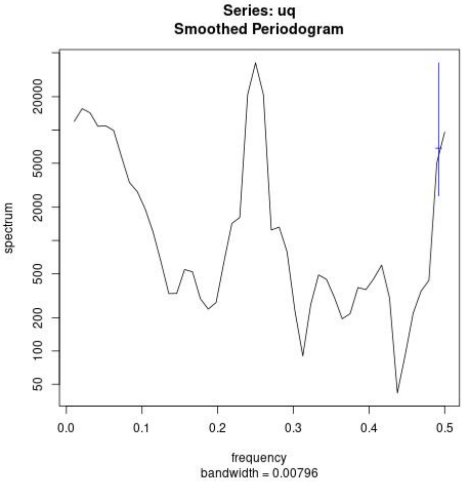
```
> spec.pgram(uq)
```



Ragged!

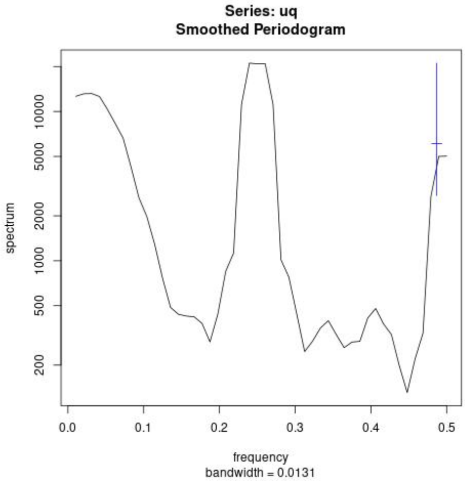
(Observe the log scale on the y axis and the 95% c.i. lenght in blue.)

```
> spec.pgram(uq,spans=2)
```



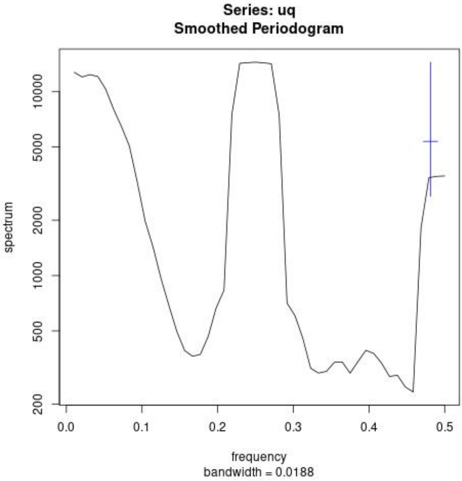
More smooth! (And shorter 95% c.i. lenght.)

```
> spec.pgram(uq,spans=4)
```



Even more smooth!

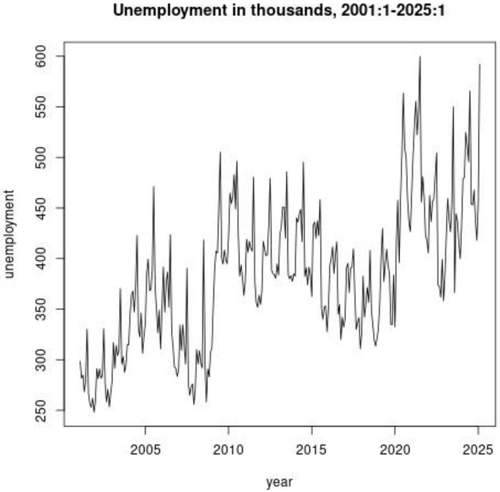
```
> spec.pgram(uq,spans=6)
```



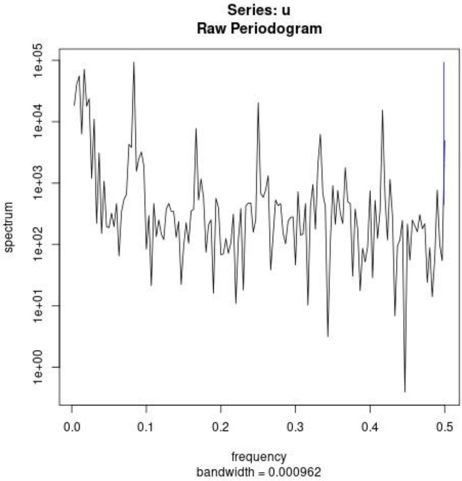
Too much smoothing?

Nonparametric Spectral Estimation

Unemployment, monthly data (period length 12, and maybe 120).

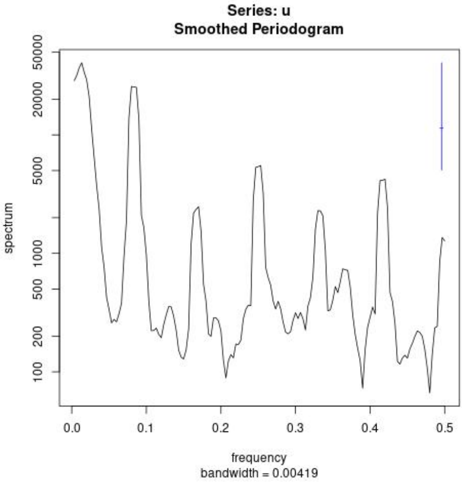


```
> spec.pgram(u)
```



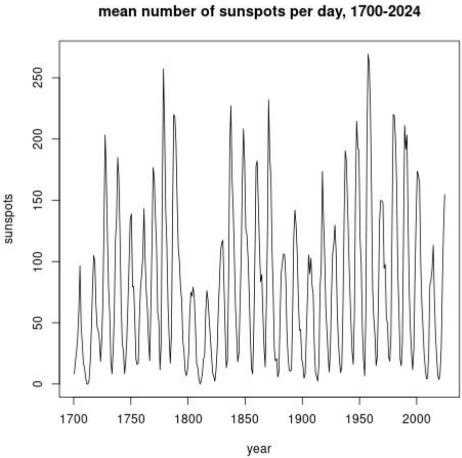
Ragged!

```
> spec.pgram(u,spans=4)
```



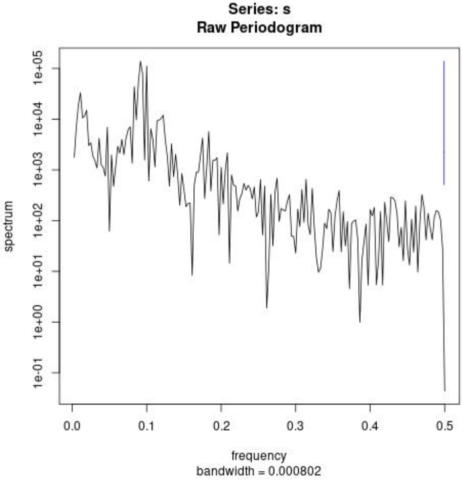
Peaks at multiples of $1/12 \approx 0.08$ and one at about $1/120 \approx 0.008$.

Nonparametric Spectral Estimation



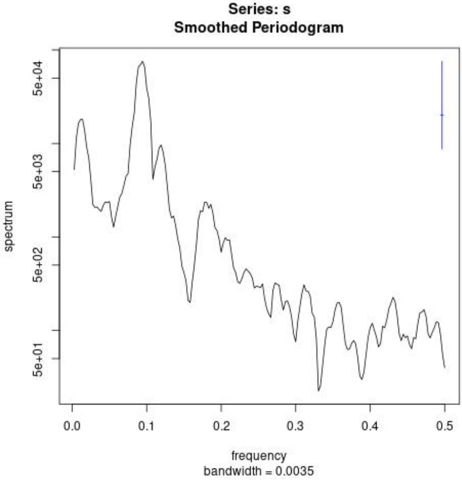
Period length 10 or 11.

```
> spec.pgram(s)
```



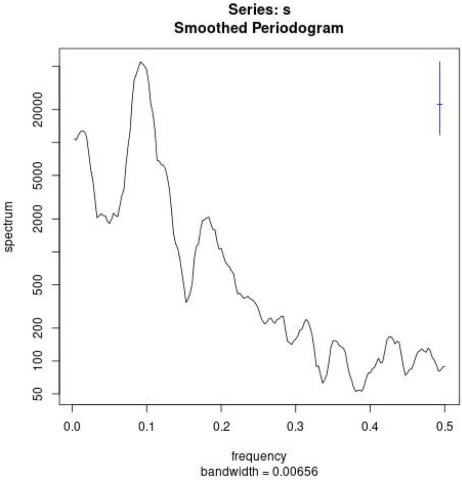
Ragged!

```
> spec.pgram(s,spans=4)
```



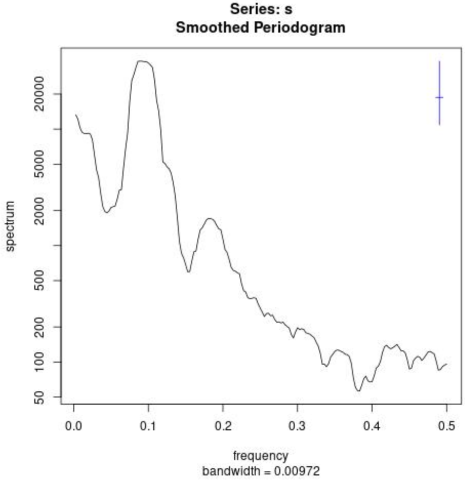
Still too ragged?

```
> spec.pgram(s,spans=8)
```



Maybe the smoothing is good enough. Main peaks at 0.1 and 0.2. Peak at around 0.01 maybe because of a 100 year period?

```
> spec.pgram(s,spans=12)
```



Too much smoothing?

Parametric Spectral Estimation

- AR(p): $\phi(B)x_t = w_t$, i.e.

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + w_t.$$

- Spectral density

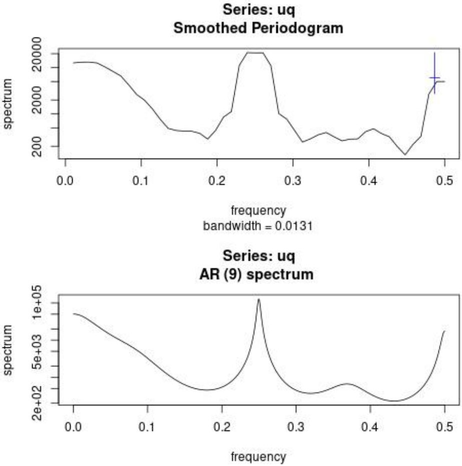
$$f(\omega) = \frac{\sigma_w^2}{|\phi(e^{-2\pi i\omega})|^2}.$$

- Idea: Fit an AR(p) model to data and estimate $f(\omega)$ by inserting the parameter estimates.
- The order p may be found by minimizing AIC.
- In R: `spec.ar(x)`
- Approximate $1-\alpha$ confidence interval (as $n, p \to \infty$, $p^3/n \to 0$)

$$\frac{\hat{f}(\omega)}{1+\sqrt{\frac{2p}{n}}z_{\alpha/2}} \le f(\omega) \le \frac{\hat{f}(\omega)}{1-\sqrt{\frac{2p}{n}}z_{\alpha/2}}.$$

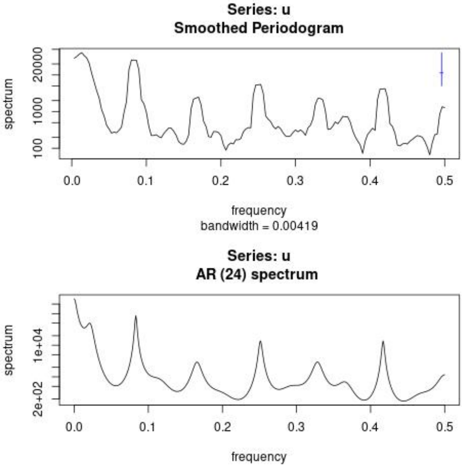
Parametric Spectral Estimation

For the quarterly unemployment series
(the upper figure is the non parametric estimate with spans=4):



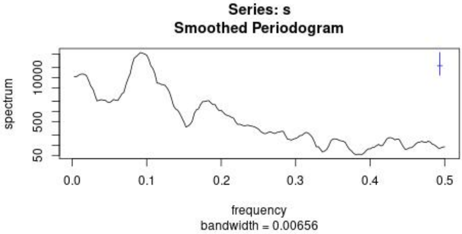
Parametric Spectral Estimation

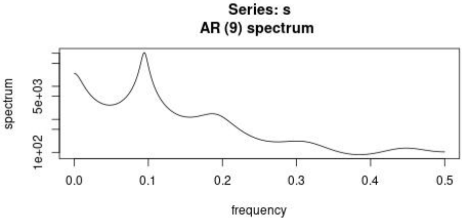
For the monthly unemployment series
(the upper figure is the non parametric estimate with spans=4):



Parametric Spectral Estimation

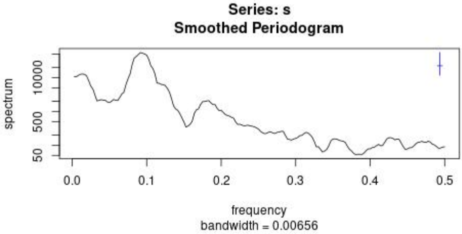
For the sunspot series
(the upper figure is the non parametric estimate with spans=8):

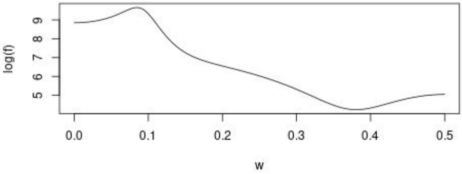




Parametric Spectral Estimation

Comparing to the spectral density for the estimated ARIMA$(3,0,0) \times (0,0,1)_{10}$ model (too smooth?):





News of today

- The periodogram
  - definition
  - asymptotic distribution
  - non consistency
- Non parametric spectral estimation
  - smoothing the periodogram
  - confidence interval
  - the lag window
  - the smoothing window
- Parametric spectral estimation
  - Fitting an AR model
  - confidence interval

Analysis of Time Series, L12

Rolf Larsson

Uppsala University

25 april 2025

Today

- 4.6: Cross Spectra
- 4.7: Linear Filters
- Wavelets (not in the book)

Cross Spectra

Let $\mathbf{x}_t = (x_{t1}, x_{t2}, ..., x_{tp})'$, $\boldsymbol{\mu} = E(\mathbf{x})$ and define the autocovariance matrix

$$\boldsymbol{\Gamma}(h) = E\{(\mathbf{x}_{t+h} - \boldsymbol{\mu})(\mathbf{x}_t - \boldsymbol{\mu})'\}$$

with elements $\gamma_{jk}(h)$, $j, k = 1, ..., p$.

Theorem (Property 4.8)

If, for all j, k,

$$\sum_{h=-\infty}^{\infty} |\gamma_{jk}(h)| < \infty,$$

then the spectral density matrix $\mathbf{f}(\omega)$ satisfies

$$\boldsymbol{\Gamma}(h) = \int_{-1/2}^{1/2} e^{2\pi i\omega h}\mathbf{f}(\omega)d\omega, \quad h = 0, \pm 1, \pm 2, ...,$$

$$\mathbf{f}(\omega) = \sum_{h=-\infty}^{\infty} \boldsymbol{\Gamma}(h)e^{-2\pi i\omega h}, \quad -1/2 \leq \omega \leq 1/2.$$

Cross Spectra

Bivariate process (x_t, y_t).

- Autocovariance matrix

$$\mathbf{\Gamma}(h) = \begin{pmatrix} \mathrm{cov}(x_{t+h}, x_t) & \mathrm{cov}(x_{t+h}, y_t) \\ \mathrm{cov}(y_{t+h}, x_t) & \mathrm{cov}(y_{t+h}, y_t) \end{pmatrix} = \begin{pmatrix} \gamma_{xx}(h) & \gamma_{xy}(h) \\ \gamma_{yx}(h) & \gamma_{yy}(h) \end{pmatrix}$$

- Spectral density matrix

$$\mathbf{f}(\omega) = \begin{pmatrix} f_{xx}(\omega) & f_{xy}(\omega) \\ f_{yx}(\omega) & f_{yy}(\omega) \end{pmatrix},$$

$$\gamma_{xy}(h) = \int_{-1/2}^{1/2} e^{2\pi i\omega h} f_{xy}(\omega)d\omega,\ f_{xy}(\omega) = \sum_{h=-\infty}^{\infty} \gamma_{xy}(h)e^{-2\pi i\omega h}.$$

- *Squared coherence*

$$\rho^2_{y\cdot x}(\omega) = \frac{|f_{yx}(\omega)|^2}{f_{xx}(\omega)f_{yy}(\omega)}.$$

Cross Spectra

- DFT $\mathbf{d}(\omega_j) = (d_1(\omega_j), d_2(\omega_j), ..., d_p(\omega_j))'$, where

$$d_k(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_{tk} e^{-2\pi i \omega_j t}, \quad j = 0, 1, ..., n-1,$$

  where $\omega_j = j/n$.
- Raw periodogram $\mathbf{I}(\omega_j) = \mathbf{d}(\omega_j)\mathbf{d}^*(\omega_j)'$.
- Smoothed periodogram

$$\begin{pmatrix} \hat{f}_{xx}(\omega_j) & \hat{f}_{xy}(\omega_j) \\ \hat{f}_{yx}(\omega_j) & \hat{f}_{yy}(\omega_j) \end{pmatrix} = \hat{\mathbf{f}}(\omega_j) = \sum_{k=-m}^{m} h_k \mathbf{I}\left(\omega_j + \frac{k}{n}\right),$$

  where $\sum_{k=-m}^{m} h_k = 1$.
- Estimated squared coherence

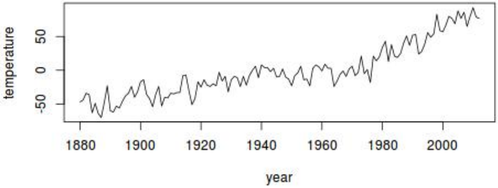
$$\hat{\rho}^2_{y \cdot x}(\omega) = \frac{|\hat{f}_{yx}(\omega)|^2}{\hat{f}_{xx}(\omega)\hat{f}_{yy}(\omega)}.$$

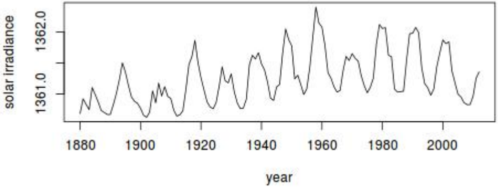
Cross Spectra

- $$\hat{\rho}^2_{y\cdot x}(\omega) = \frac{|\hat{f}_{yx}(\omega)|^2}{\hat{f}_{xx}(\omega)\hat{f}_{yy}(\omega)}.$$
- For large n,
$$|\hat{\rho}_{y\cdot x}(\omega)| \approx N\left(|\rho_{y\cdot x}(\omega)|, \frac{\{1-\rho^2_{y\cdot x}(\omega)\}^2}{2L_h}\right),$$
where
$$L_h = \left(\sum_{k=-m}^{m} h_k^2\right)^{-1}$$
- May be used to construct confidence intervals for $\rho^2_{y\cdot x}(\omega)$.

Cross Spectra

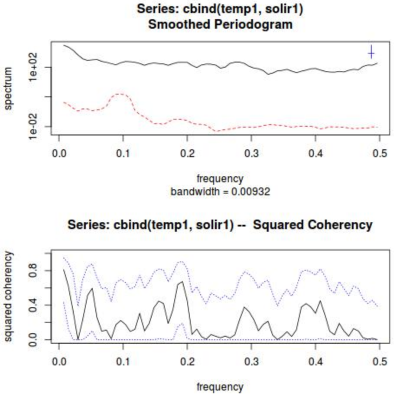
Yearly mean temperature and solar irradiance, 1880-2012.





Cross Spectra

Smoothing with spans=4



Significance (lower dotted curve above 0) basically only at 0 (common trend).

R code:

```
> par(mfrow=c(2,1))
> s=spec.pgram(cbind(temp1,solir1),spans=4)
> plot(s,plot.type="coh")
```

Linear Filters

- Input series $\{x_t\}$.
- A *linear filter* $\{a_t\}$ such that $\sum_{j=-\infty}^{\infty} |a_j| < \infty$.
- Output series $\{y_t\}$ such that we have the convolution

$$y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j}.$$

- Examples:
  - Moving average, e.g. $a_0 = a_1 = a_2 = a_3 = \frac{1}{4}$ (all the other $a_j = 0$).
  - Difference, e.g. $a_0 = 1$, $a_1 = -1$ (all the other $a_j = 0$).

Linear Filters

Frequency response function

$$A_{yx}(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{-2\pi i\omega j}.$$

Theorem (Property 4.9)

The spectral density $f_{yy}(\omega)$ of the filtered output y_t is related to the spectral density $f_{xx}(\omega)$ of the input x_t through

$$f_{yy}(\omega) = |A_{yx}(\omega)|^2 f_{xx}(\omega).$$

Prove that a causal ARMA process $\phi(B)x_t = \theta(B)w_t$ has spectral density

$$f(\omega) = \sigma_w^2 \frac{|\theta(e^{-2\pi i\omega})|^2}{|\phi(e^{-2\pi i\omega})|^2}.$$

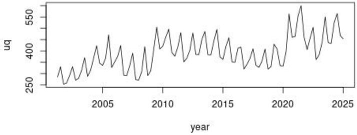
Linear Filters

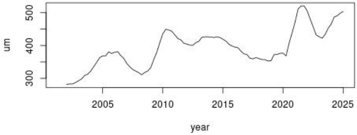
Calculate $A_{yx}(\omega)$ and $|A_{yx}(\omega)|^2$ for

1. the moving average filter $a_0 = a_1 = a_2 = a_3 = \frac{1}{4}$ (all the other $a_j = 0$) and show that $|A_{yx}(1/4)|^2 = 0$.
2. the difference filter $a_0 = 1$, $a_1 = -1$ (all the other $a_j = 0$) and show that $|A_{yx}(0)|^2 = 0$.

Linear Filters

The quarterly unemployment series and its yearly moving average:
In R: `um=filter(uq,c(rep(1/4,4)),sides=1)`

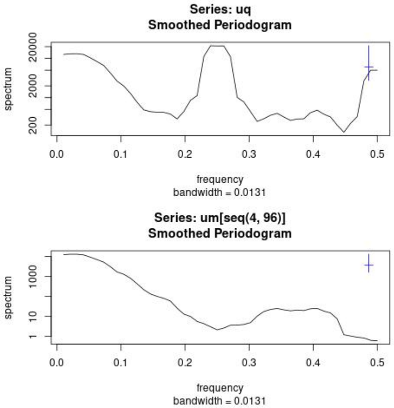




The season is smoothed out!

Linear Filters

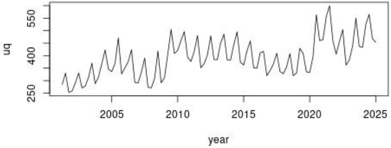
Corresponding non parametric spectral estimates, spans=4:

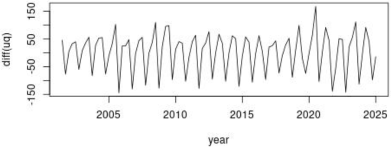


The peak at 0.25 has disappeared!

Linear Filters

The quarterly unemployment series and its difference:

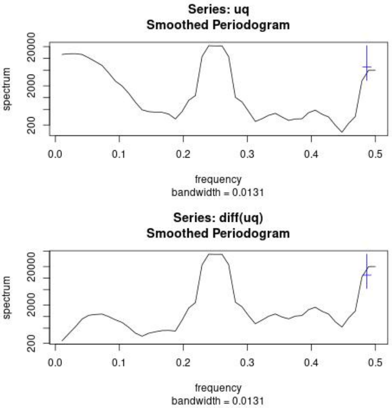




The trend is differenced out!

Linear Filters

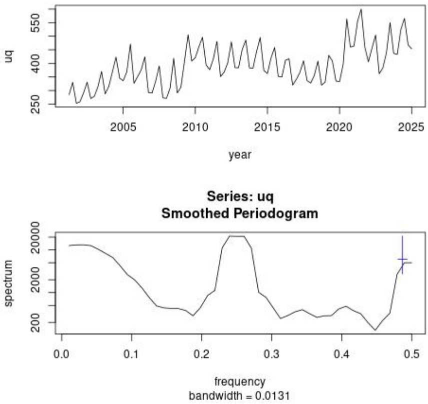
Corresponding non parametric spectral estimates, spans=4:



The peak at 0 has disappeared!

Wavelets

The unemployment series and its estimated periodogram (spans=4):



Different trend behavior for different years.

Wavelets

Chap. 2:

- "Exact" regression of x_t on periodic functions:

$$x_t = \sum_{j=1}^{n/2} \left\{ \hat{\beta}_{1j} \cos(2\pi\omega_j t) + \hat{\beta}_{2j} \sin(2\pi\omega_j t) \right\}$$

  where $\omega_j = j/n$.

- *The periodogram* consists of the estimated weights

$$P(\omega_j) = \hat{\beta}_{1j}^2 + \hat{\beta}_{2j}^2.$$

- The *basis functions* $\{cos(2\pi\omega_j t), \sin(2\pi\omega_j t)\}$ are periodic and *orthogonal*.
- $f(t)$ and $g(t)$ are orthogonal if $\int f(t)g(t)dt = 0$.

Wavelets

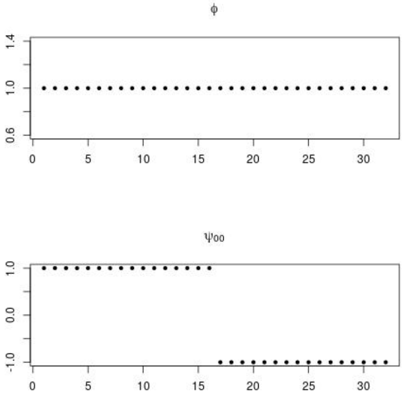
- A linear combination of orthogonal, *periodic* functions is not a good description of data if the trends or periodicities are not the same during the sampling interval.
- If so, it is better to use a linear combination of orthogonal, *non periodic functions*!
- *The wavelet transform* is

$$x_t = s\phi(t) + \sum_{j=0}^{m-1} \sum_{k=0}^{2^j-1} d_{jk}\psi_{jk}(t),$$

  where $n = 2^m$ and $\phi(t)$ (father wavelet) and the $\psi_{jk}(t)$ (mother wavelets) together form an orthonormal basis.
- Truncating the sum gives a smoothed version of the series.
- One simple set of wavelets are the *Haar wavelets*, as illustrated below.

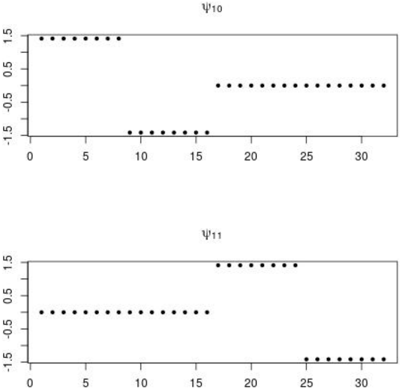
Wavelets

$n = 32$, Haar wavelets $\phi(t)$ and $\psi_{00}(t)$



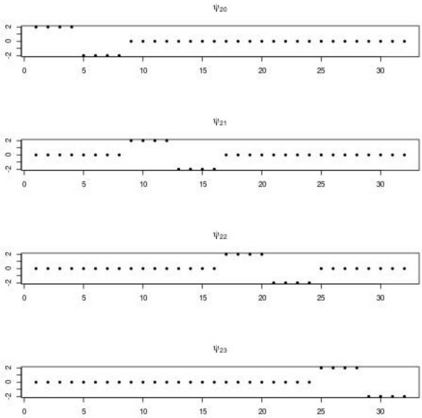
Wavelets

$n = 32$, Haar wavelets $\psi_{10}(t)$ and $\psi_{11}(t)$



Wavelets

$n = 32$, Haar wavelets $\psi_{20}(t)$, $\psi_{21}(t)$, $\psi_{22}(t)$ and $\psi_{23}(t)$



Wavelets

- $$x_t = s\phi(t) + \sum_{j=0}^{m-1} \sum_{k=0}^{2^j-1} d_{jk}\psi_{jk}(t).$$
- Example with $n = 8 = 2^3$, Haar wavelet:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \sqrt{2} & 0 & 2 & 0 & 0 & 0 \\ 1 & 1 & \sqrt{2} & 0 & -2 & 0 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & 2 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & -2 & 0 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & 2 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & -2 & 0 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & 2 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} s \\ d_{00} \\ d_{10} \\ d_{11} \\ d_{20} \\ d_{21} \\ d_{22} \\ d_{23} \end{pmatrix}$$

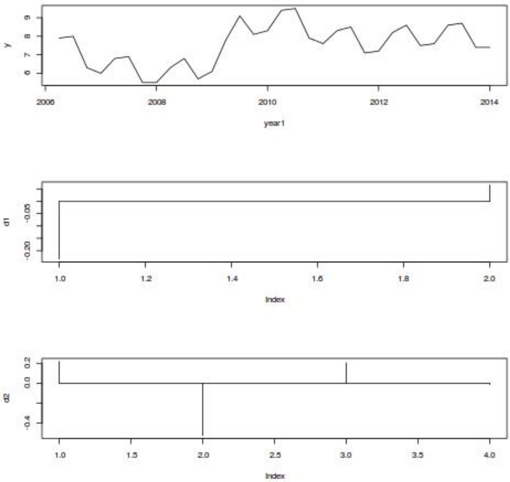
Wavelets

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \sqrt{2} & 0 & 2 & 0 & 0 & 0 \\ 1 & 1 & \sqrt{2} & 0 & -2 & 0 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & 2 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & -2 & 0 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & 2 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & -2 & 0 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & 2 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} s \\ d_{00} \\ d_{10} \\ d_{11} \\ d_{20} \\ d_{21} \\ d_{22} \\ d_{23} \end{pmatrix}$$

$$\mathbf{x}^{(8\times 1)} = \begin{pmatrix} \phi^{(8\times 1)} & \psi_0^{(8\times 1)} & \Psi_1^{(8\times 2)} & \Psi_2^{(8\times 4)} \end{pmatrix} \begin{pmatrix} s \\ d_{00} \\ \mathbf{d}_1^{(2\times 1)} \\ \mathbf{d}_2^{(4\times 1)} \end{pmatrix}$$

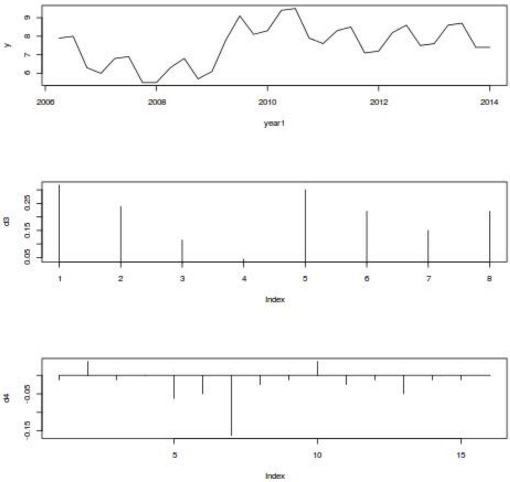
Wavelets

The unemployment series, 2006:1-2013:4, the data, $\mathbf{d_1}$ and $\mathbf{d_2}$:



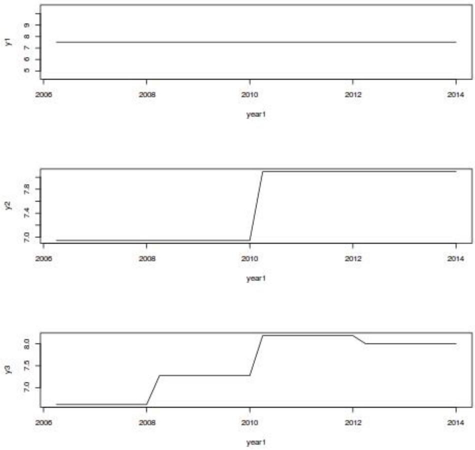
Wavelets

The unemployment series, 2006:1-2013:4, the data, $\mathbf{d_3}$ and $\mathbf{d_4}$:



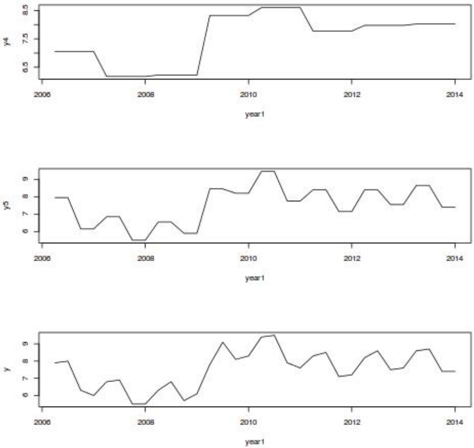
Wavelets

Unemployment, smoothed series with 1-3 components.



Wavelets

Unemployment, smoothed series with 4-6 components.



Wavelets

Some references:

- Shleicher, C. (2002) An Introduction to Wavelets for Economists, working paper.
- Crowley, P.M. (2007) A guide to Wavelets for economists, *Journal of Economic Surveys*, 21, 207-267.
- Zwanzig, S., Mahjani, B. (2020) *Computer Intensive Methods in Statistics*, CRC Press.

News of today

- Cross spectra, coherency
- Linear filters:
  - high-pass (e.g. differences)
  - low-pass (e.g. moving averages)
- Wavelets

Analysis of Time Series, L13

Rolf Larsson

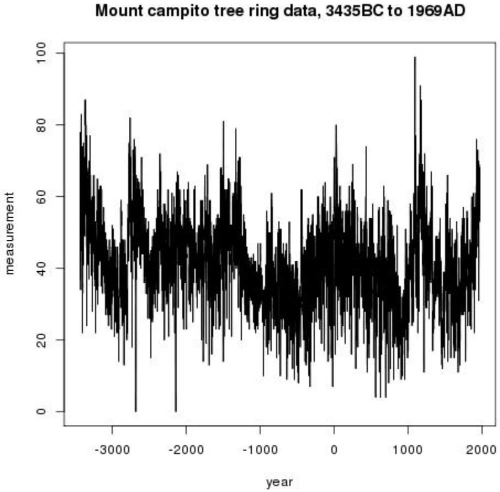
Uppsala University

29 april 2025

Today

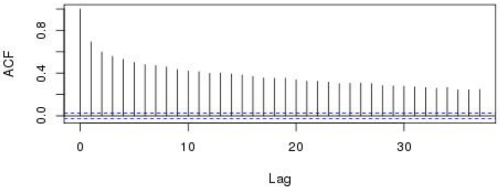
- 5.1: Long memory ARMA and Fractional Differencing
- 5.4: Threshold models
- Menti

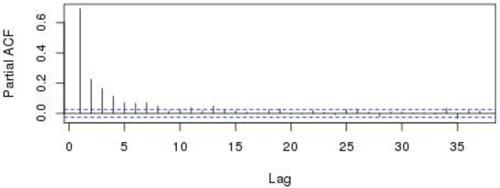
Fractional differencing



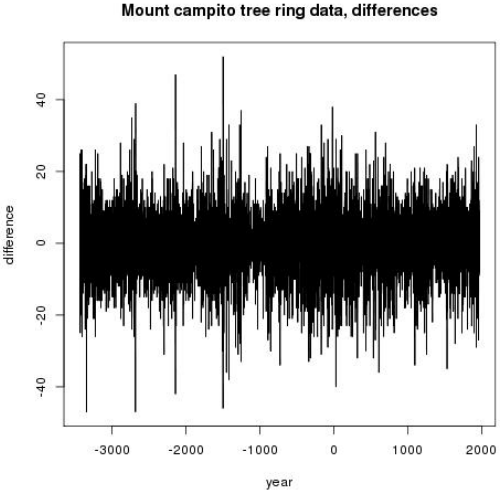
Fractional differencing

Mount Campito, ACF (slowly decreasing) and PACF (not cutting off)



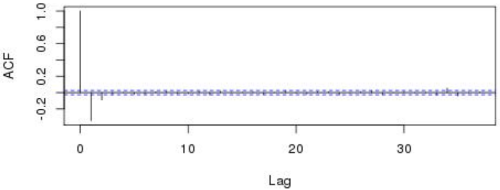


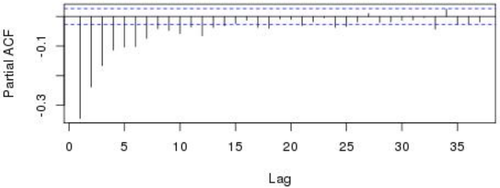
Fractional differencing



Fractional differencing

Mount Campito differences, ACF (cuts off?) and PACF





Fractional differencing

- Consider the model $(1-B)^d x_t = w_t$:
  - Random walk for $d = 1$
  - White noise for $d = 0$
  - What about $0 < d < 1$?
- WLOG: Restrict to $|d| < 1/2$.
- Binomial expansion

$$(1-B)^d = 1 - dB + \binom{d}{2} B^2 - \binom{d}{3} B^3 + \ldots = \sum_{j=0}^{\infty} \pi_j B^j,$$

where (why?)

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}, \quad \Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx.$$

Fractional differencing

- MA representation

$$x_t = (1-B)^{-d} w_t = \sum_{j=0}^{\infty} \psi_j(d) w_{t-j},$$

where

$$\psi_j(d) = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}.$$

- Inserting in $\gamma(h) = \sigma_w^2 \sum_{j=0}^{\infty} \psi_j(d)\psi_{j+h}(d)$, it may be proved that

$$\gamma(h) = \sigma_w^2 \frac{\Gamma(h+d)\Gamma(1-2d)}{\Gamma(h+1-d)\Gamma(1-d)\Gamma(d)}.$$

Fractional differencing

- Hence (why?),

$$\rho(h) = \frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(h+1-d)\Gamma(d)}.$$

- Notation: $f(z) \sim g(z)$ means that $\frac{f(z)}{g(z)} \to 1$ as $z \to \infty$.
- Stirling's formula: $\Gamma(z) \sim \sqrt{2\pi} z^{z-1/2} e^{-z}$ as $z \to \infty$ implies (why?)

$$\rho(h) \sim Ch^{2d-1}, \quad \text{as } h \to \infty.$$

- Hence, for $d \geq 0$ (why?),

$$\sum_{h=-\infty}^{\infty} |\rho(h)| = \infty.$$

Estimation

- Recall: $(1-B)^d x_t = w_t$. Assume that w_t is normal white noise.
- The MLE of d is found by maximum likelihood, approximately obtained by minimizing

$$Q(d) = \sum_t w_t^2.$$

- Use the R package fracdiff!

Prediction

- AR representation

$$w_t = (1-B)^d x_t = \sum_{j=0}^{\infty} \pi_j(d) x_{t-j},$$

where

$$\pi_j(d) = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}.$$

- Truncated forecast

$$\tilde{x}^n_{n+m} = -\sum_{j=1}^{m-1} \pi_j(\hat{d}) \tilde{x}^n_{n+m-j} - \sum_{j=m}^{\infty} \pi_j(\hat{d}) x_{n+m-j}.$$

Prediction

- MA representation

$$x_t = (1-B)^{-d} w_t = \sum_{j=0}^{\infty} \psi_j(d) w_{t-j},$$

where

$$\psi_j(d) = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}.$$

- MSPE

$$P_{n+m}^n = \hat{\sigma}_w^2 \sum_{j=0}^{m-1} \psi_j^2(\hat{d}).$$

Spectral estimation

- Spectral density based on the AR representation

$$f(\omega) = \frac{\sigma_w^2}{|\pi(e^{-2\pi i\omega})|^2} = \frac{\sigma_w^2}{|\sum_{k=0}^{\infty} \pi_k e^{-2\pi i k\omega}|^2}.$$

- It follows that (why?)

$$f(\omega) = \sigma_w^2\{4\sin^2(\pi\omega)\}^{-d},$$

i.e. for ω small, $f(\omega) \approx \sigma_w^2\{4(\pi\omega)^2\}^{-d} = C\omega^{-2d}$.

- Parametric spectral estimate

$$\hat{f}(\omega) = \hat{\sigma}_w^2\{4\sin^2(\pi\omega)\}^{-\hat{d}},$$

i.e. for ω small, $\hat{f}(\omega) \approx C\omega^{-2\hat{d}}$.

Spectral estimation

- $\hat{f}(\omega) \approx C\omega^{-2\hat{d}}$ for ω small.
- Hence,
$$\log\{\hat{f}(\omega)\} \approx \log C - 2\hat{d}\log\omega.$$
- This suggests an alternative estimation method:
Estimate d via a linear regression of $\log\{\hat{f}(\omega)\}$ on $\log\omega$ for ω "small".

Fitting ARMA(2,2) to differenced Mount Campito data:

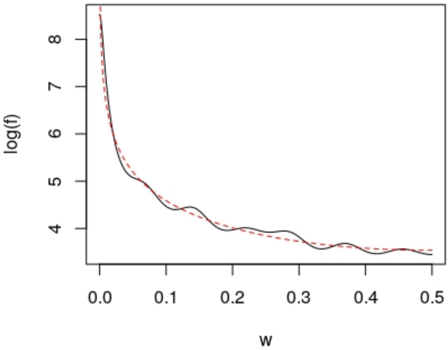
```
> a1=arima(dy,order=c(2,0,2),include.mean=FALSE);a1

Call:
arima(x = dy, order = c(2, 0, 2), include.mean = FALSE)

Coefficients:
         ar1      ar2      ma1     ma2
      1.1512  -0.2216  -1.7007  0.7059
s.e.  0.0367   0.0260   0.0322  0.0313

sigma^2 estimated as 63.93:  log likelihood = -18902.95,
aic = 37815.9
```

Fitting $(1-B)^d x_t = w_t$ to demeaned Mount Campito data (in `x`):

```
> xm=x-mean(x)
> library(fracdiff)
> m=fracdiff(xm)
> m$d
[1] 0.4472056
> m$sigma
[1] 7.991763
```

Tree rings

Estimated log spectral densities:
Parametric AR estimate (AR(14)) in black
Parametric fractional estimate dashed in red



R code for the plot:

```
> s=spec.ar(x,plot=FALSE)
> f=m$sigma^2*(4*sin(pi*s$freq)^2)^(-m$d)
> plot(s$freq,log(s$spec),type='l',xlab='w',ylab='log(f)')
> lines(s$freq,log(f),lty=2,col='red')
```

General ARFIMA

The general ARFIMA(p, d, q) model is defined as

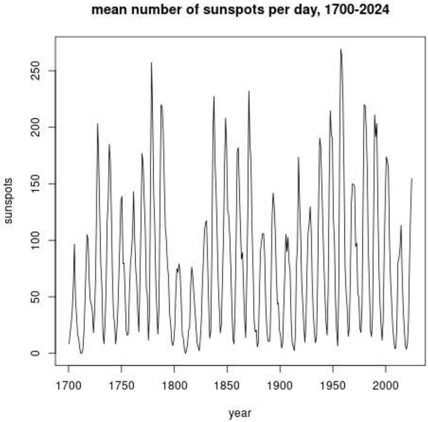
$$\phi(B)\nabla^d(x_t - \mu) = \theta(B)w_t$$

where

$$\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p, \quad \theta(B) = 1 + \theta_1 B + ... + \theta_q B^q.$$

Threshold models

Can the behavior of a series be different above and below a certain threshold?



Threshold models

Threshold AutoRegressive (TAR) model.

- General: If the regions R_1,...,R_r are mutually exclusive and exhaustive, we may define

$$x_t = \alpha^{(j)} + \phi_1^{(j)} x_{t-1} + ... + \phi_p^{(j)} x_{t-p} + w_t^{(j)},$$

  where $(x_{t-1}, ..., x_{t-p}) \in R_j$, $j = 1, 2, ..., r$.

- Special case with $r = 2$, $p = 2$:

$$x_t = \alpha^{(j)} + \phi_1^{(j)} x_{t-1} + \phi_2^{(j)} x_{t-2} + w_t^{(j)},$$

  where $R_1 = \{x_{t-1} < c\}$, $R_2 = \{x_{t-1} \geq c\}$, $j = 1, 2$.

Threshold models

- Special case with $r = 2$, $p = 2$:

$$x_t = \alpha^{(j)} + \phi_1^{(j)} x_{t-1} + \phi_2^{(j)} x_{t-2} + w_t^{(j)},$$

where $R_1 = \{x_{t-1} < c\}$, $R_2 = \{x_{t-1} \geq c\}$, $j = 1, 2$.

- Let $\delta = 1$ if $x_{t-1} \geq c$ and 0 otherwise.
- Equivalent to regression model:

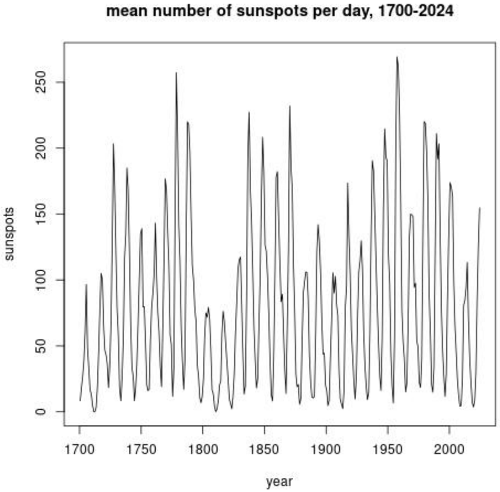
$$\begin{aligned} x_t = &\alpha^{(1)}(1-\delta) + \phi_1^{(1)}(1-\delta)x_{t-1} + \phi_2^{(1)}(1-\delta)x_{t-2} \\ &+ \alpha^{(2)}\delta + \phi_1^{(2)}\delta x_{t-1} + \phi_2^{(2)}\delta x_{t-2} + w_t, \end{aligned}$$

where

$$w_t = (1-\delta)w_t^{(1)} + \delta w_t^{(2)} \sim N(0, \sigma^2)$$

with $\sigma^2 = (1-\delta)^2\sigma_1^2 + \delta^2\sigma_2^2, \quad \sigma_j^2 = \mathrm{var}(w_t^{(j)}), \; j = 1, 2.$

Threshold models



Fit the TAR model

$$x_t = \alpha^{(j)} + \phi_1^{(j)} x_{t-1} + \phi_2^{(j)} x_{t-2} + w_t^{(j)},$$

where $R_1 = \{x_{t-1} < 75\}$, $R_2 = \{x_{t-1} \geq 75\}$, $j = 1, 2$.

In R:

```
> length(x)
[1] 325
> x0=x[seq(3,325)]
> x1=x[seq(2,324)]
> x2=x[seq(1,323)]
> d=(sign(x1-74.99)+1)/2
> x11=(1-d)*x1;x12=d*x1;x21=(1-d)*x2;x22=d*x2;d2=1-d;
> m=lm(x0~0+d2+x11+x21+d+x12+x22);summary(m)
```

```
Call:
lm(formula = x0 ~ 0 + d2 + x11 + x21 + d + x12 + x22)

Residuals:
Min       1Q  Median      3Q     Max
-69.277 -15.507  -3.303  13.873  86.909

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
d2  23.25044    3.46327   6.713 8.78e-11 ***
x11  1.79571    0.11271  15.932  < 2e-16 ***
x21 -1.02830    0.07749 -13.271  < 2e-16 ***
d   34.28875    6.42086   5.340 1.77e-07 ***
x12  1.23227    0.05477  22.499  < 2e-16 ***
x22 -0.59786    0.04488 -13.322  < 2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.73 on 317 degrees of freedom
Multiple R-squared:  0.9405,Adjusted R-squared:  0.9394
F-statistic: 835.6 on 6 and 317 DF,  p-value: < 2.2e-16
```

Threshold models

Smooth Transition AutoRegressive (STAR) model:

- $$\begin{aligned} x_t = {} & \alpha^{(1)} + \phi_1^{(1)} x_{t-1} + \ldots + \phi_p^{(1)} x_{t-p} \\ & + (\alpha^{(2)} + \phi_1^{(2)} x_{t-1} + \ldots + \phi_p^{(2)} x_{t-p}) f(x_{t-1}) + w_t, \end{aligned}$$

  where

  $$f(x) = \frac{1}{1 + e^{(c-x)/\eta}}.$$

- Approaches a TAR model as $\eta \searrow 0$.
- Explanation: for η small, $f(x) \approx 0$ if $x < c$, $f(x) \approx 1$ if $x > c$.

```
> m=nls(x0~a1+f11*x1+f21*x2+(a2+f12*x1+f22*x2)*1/(1+exp((75-x1)/eta)),
+       start=list(a1=23.37,f11=1.79,f21=-1.026,a2=11.06,f12=-0.557,f22=0.4
> summary(m)

Formula: x0 ~ a1 + f11 * x1 + f21 * x2 + (a2 + f12 * x1 + f22 * x2) *
1/(1 + exp((75 - x1)/eta))

Parameters:
    Estimate Std. Error t value Pr(>|t|)
a1  23.37309    3.46832   6.739 7.56e-11 ***
f11  1.78672    0.11257  15.872  < 2e-16 ***
f21 -1.02582    0.07761 -13.218  < 2e-16 ***
a2  11.29601    7.37206   1.532    0.126
f12 -0.55610    0.12523  -4.441 1.24e-05 ***
f22  0.42731    0.08979   4.759 2.96e-06 ***
eta  0.10543    0.90941   0.116    0.908
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 24.78 on 316 degrees of freedom
Number of iterations to convergence: 5
Achieved convergence tolerance: 3.163e-06
```

News of today

- Long memory models:
  - Fractional difference
  - Estimating the d parameter
  - Prediction via AR representation
  - MSPE via MA representation
  - Spectral estimation via AR representation
- Threshold models:
  - TAR
  - STAR

Analysis of Time Series, L14

Rolf Larsson

Uppsala University

8 maj 2025

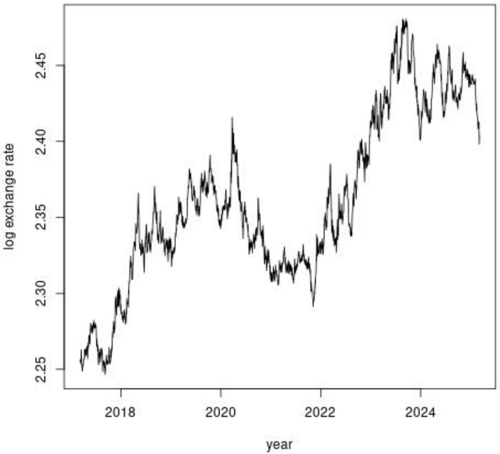
Today

5.2: Unit root testing

- DF test
- ADF test
- With deterministic terms
- KPSS test (not in book)
- Application
- Menti

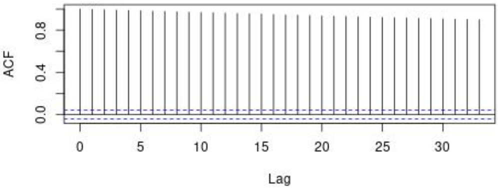
DF test

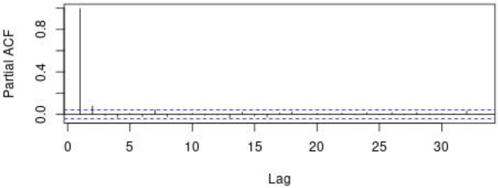
Log exchange rate, Skr/Euro, March 6 2017-March 5 2025:



DF test

Log Skr/Euro, ACF (slowly decaying) and PACF (cutting off).





```
> y1=log(y)
> arima(y1,order=c(1,0,0))

Call:
arima(x = y1, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.9979     2.3610
s.e.  0.0014     0.0347

sigma^2 estimated as 1.608e-05:  log likelihood = 8558.44,
aic = -17110.87
```

Estimated AR parameter very close to 1.

```
> arima(y1,order=c(2,0,0))

Call:
arima(x = y1, order = c(2, 0, 0))

Coefficients:
         ar1     ar2  intercept
      0.8848  0.1134     2.3610
s.e.  0.0217  0.0217     0.0389

sigma^2 estimated as 1.587e-05:  log likelihood = 8571.93,
aic = -17135.86
```

Sum of estimated AR parameters very close to 1.

DF test

- AR(1) process

$$x_t = \phi x_{t-1} + w_t, \quad t = 1, 2, ..., n.$$

- Causal if.f. $|\phi| < 1$.
- Stationary for a suitable choice of distribution for x_0 $(\mathrm{var}(x_0) = \frac{\sigma_w^2}{1-\phi^2})$ if.f. $|\phi| < 1$.
- Test H_0: $\phi = 1$ vs H_1: $|\phi| < 1$.
- A natural test statistic is $\hat{\phi} - 1$.
- The *Dickey-Fuller* (DF) test.

DF test

- Under H_0: $\phi = 1$, it follows that (why?)

$$\hat{\phi} - 1 = \frac{\sum_{t=1}^{n} x_t x_{t-1}}{\sum_{t=1}^{n} x_{t-1}^2} - 1 = \frac{\sum_{t=1}^{n} w_t x_{t-1}}{\sum_{t=1}^{n} x_{t-1}^2}.$$

- Moreover (why?), assuming $x_0 = 0$,

$$\hat{\phi} - 1 = \frac{x_n^2 - \sum_{t=1}^{n} w_t^2}{2\sum_{t=1}^{n} x_{t-1}^2}.$$

DF test

Definition (5.1)

A continuous time process $\{W(t);\ t \geq 0\}$ is called a *standard Brownian motion* if it satisfies

(*i*) $W(0) = 0$.

(*ii*) For any $0 \leq t_1 < t_2 < ... < t_n$ and integer n, $W(t_2) - W(t_1),\ W(t_3) - W(t_2),...,W(t_n) - W(t_{n-1})$ are independent.

(*iii*) $W(t + \Delta t) - W(t) \sim N(0, \Delta t)$ for $\Delta t > 0$.

One may show that for a white noise process $\{w_t\}$, as $n \to \infty$,

$$\frac{1}{\sigma_w \sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} w_j \xrightarrow{\mathcal{L}} W(t)$$

where $\lfloor a \rfloor$ is the integer part of a.

DF test

- As $n \to \infty$,

$$\frac{1}{\sigma_w \sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} w_j \xrightarrow{\mathcal{L}} W(t)$$

- It follows that (why?)

$$n(\hat{\phi} - 1) = n \frac{x_n^2 - \sum_{t=1}^n w_t^2}{2 \sum_{t=1}^n x_{t-1}^2} \xrightarrow{\mathcal{L}} \frac{W(1)^2 - 1}{2 \int_0^1 W(t)^2 dt}.$$

- No "standard" distribution!

DF test

- An alternative is the t test

$$\hat{t} = \frac{\hat{\phi} - 1}{\sqrt{s^2 / \sum_{t=2}^{n} x_{t-1}^2}},$$

where

$$s^2 = \frac{1}{n-1} \sum_{t=2}^{n} (x_t - \hat{\phi} x_{t-1})^2.$$

- One may show that, as $n \to \infty$,

$$\hat{t} \xrightarrow{\mathcal{L}} \frac{W(1)^2 - 1}{2\sqrt{\int W(t)^2 dt}}.$$

ADF test

AR(1): $x_t = \phi x_{t-1} + w_t \Leftrightarrow \nabla x_t = \gamma x_{t-1} + w_t$, $\gamma = \phi - 1$.

Extension to AR(2):

- $$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$$
- Equivalent to (why?)

  $$\nabla x_t = \gamma x_{t-1} + \psi_1 \nabla x_{t-1} + w_t,$$

  where

  $$\gamma = \phi_1 + \phi_2 - 1, \quad \psi_1 = -\phi_2.$$

ADF test

Extension to AR(p):

- $$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t$$
- Equivalent to
$$\nabla x_t = \gamma x_{t-1} + \psi_1 \nabla x_{t-1} + \ldots + \psi_{p-1} \nabla x_{t-(p-1)} + w_t$$
where
$$\gamma = \sum_{j=1}^{p} \phi_j - 1, \quad \psi_j = -\sum_{i=j+1}^{p} \phi_i.$$
- The Augmented Dickey-Fuller (ADF) test:
Test H_0: $\gamma = 0$ vs H_1: $\gamma < 0$.
- Modified test statistics, but the same limit distributions as before.

With deterministic terms

Incorporating deterministic terms:

- $x_t = \beta_0 + \phi x_{t-1} + w_t$
- Test H_0: $(\beta_0, \phi) = (0, 1)$ vs H_1: $\neg H_0$.
- $x_t = \beta_0 + \beta_1 t + \phi x_{t-1} + w_t$
- Test H_0: $(\beta_1, \phi) = (0, 1)$ vs H_1: $\neg H_0$.
- Modified test statistics, other limit distributions.

KPSS test

- Test H_0: stationarity vs H_1: non stationarity.
- Model

$$x_t = r_t + \varepsilon_t,$$
$$r_t = r_{t-1} + u_t,$$

where $\{\varepsilon_t\}$ and $\{u_t\}$ are independent white noise sequences, $t = 1, 2, ..., n$.

- Test H_0: $\mathrm{var}(u_t) = \sigma_u^2 = 0$ vs H_1: $\sigma_u^2 > 0$.
- Under H_0, $x_t = r_0 + \varepsilon_t$ is stationary.
- Let $S_t = \sum_{j=1}^{t} e_j$, $e_j = (x_j - \bar{x})$ and estimate its variance by

$$s^2(l) = \frac{1}{n}\sum_{t=1}^{n} e_t^2 + \frac{2}{n}\sum_{s=1}^{l}\left(1 - \frac{s}{l+1}\right)\sum_{t=s+1}^{n} e_t e_{t-s}.$$

KPSS test

- Let $S_t = \sum_{j=1}^{t} e_j$, $e_j = (x_j - \bar{x})$ and estimate its variance by

$$s^2(l) = \frac{1}{n}\sum_{t=1}^{n} e_t^2 + \frac{2}{n}\sum_{s=1}^{l}\left(1 - \frac{s}{l+1}\right)\sum_{t=s+1}^{n} e_t e_{t-s}.$$

- Test statistic

$$\hat{\eta} = \frac{\sum_{t=1}^{n} S_t^2}{n^2 s^2(l)}.$$

- As $n, l \to \infty$ such that $l/\sqrt{n} \to 0$,

$$\hat{\eta} \xrightarrow{d} \int_0^1 \{W(r) - rW(1)\}^2 \, dr,$$

where $W(r)$ is the standard Brownian motion.

Unit root tests for log Skr/Euro in R:

```
> y1=log(y)
> library(tseries)
> adf.test(y1)

Augmented Dickey-Fuller Test

data:  y1
Dickey-Fuller = -2.6273, Lag order = 12, p-value = 0.3128
alternative hypothesis: stationary

> kpss.test(y1)

KPSS Test for Level Stationarity

data:  y1
KPSS Level = 15.517, Truncation lag parameter = 8, p-value = 0.01

Warning message:
In kpss.test(y1) : p-value smaller than printed p-value
```

The adf test does not reject non stationarity and the KPSS test rejects stationarity.

News of today

- Testing for non stationarity (unit root)
- Test statistics
- Limit distributions (non standard)
- Extensions:
  - Allowing for autocorrelation (ADF test)
  - Allowing for deterministic terms
- Testing the null of stationarity (KPSS test)

Analysis of Time Series, L15

Rolf Larsson

Uppsala University

9 maj 2025

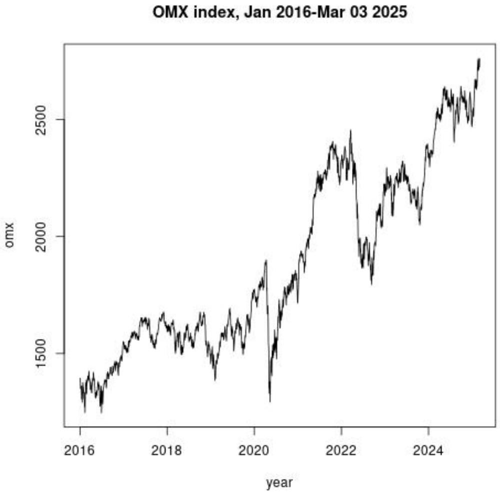
Today

5.3: GARCH models:

- ARCH
- GARCH
- other models

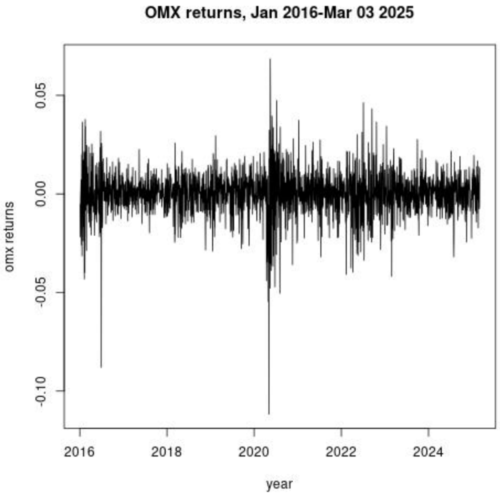
ARCH

OMX series x_t



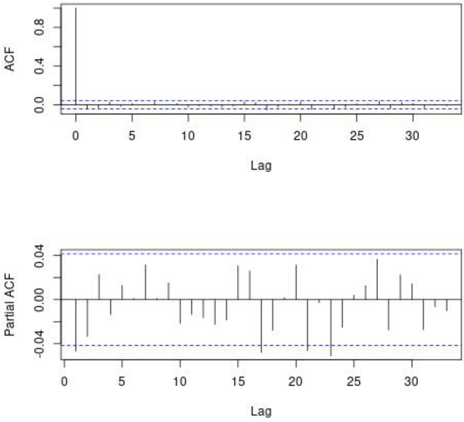
ARCH

OMX returns $y_t = \ln(x_t) - \ln(x_{t-1})$



ARCH

ACF and PACF for $y_t = \ln(x_t) - \ln(x_{t-1})$. White noise?



Box-Ljung test of H_0: The y_t are uncorrelated. Do not reject for lag length 10.

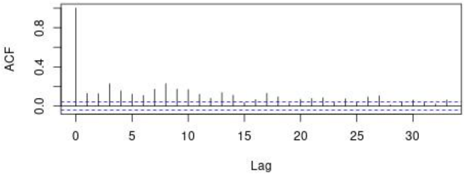
```
> Box.test(y,type="Ljung-Box")

Box-Ljung test

data:  y
X-squared = 4.8895, df = 1, p-value = 0.02702

> Box.test(y,type="Ljung-Box", lag=10)

Box-Ljung test

data:  y
X-squared = 12.755, df = 10, p-value = 0.2377
```

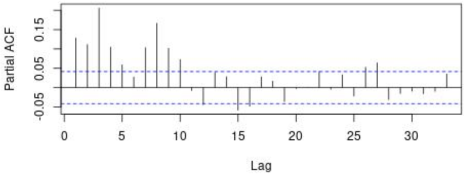
Box-Ljung test of H_0: The y_t^2 are uncorrelated. Reject for both lag lengths.

```
> Box.test(y^2,type="Ljung-Box")

Box-Ljung test

data:  y^2
X-squared = 36.764, df = 1, p-value = 1.333e-09

> Box.test(y^2,type="Ljung-Box",lag=10)

Box-Ljung test

data:  y^2
X-squared = 614.21, df = 10, p-value < 2.2e-16
```

ARCH

ACF and PACF for y_t^2. A lot of structure!





ARCH

Robert Engle, Nobel prize 2003.

https://www.nobelprize.org/nobelprizes/economic-sciences/laureates/2003/engle-photo.html

ARCH

An important formula: $E(X) = E\{E(X|Y)\}$.

In the discrete case, we have

$$E(X|Y=y) = \sum_x xP(X=x|Y=y)$$

which leads to (why?)

$$E\{E(X|Y)\} = \sum_x xP(X=x) = E(X).$$

ARCH

AutoRegressive Conditional Heteroscedasticity (ARCH) model:

- $$\begin{aligned} y_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2, \end{aligned}$$

  where the ϵ_t are i.i.d. $N(0,1)$ and $\alpha_0 > 0$, $\alpha_1 > 0$.
- It follows that
  $$y_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + v_t,$$
  where $v_t = \sigma_t^2(\epsilon_t^2 - 1)$.
- $\{y_t\}$ is an uncorrelated sequence (why?), but $\{y_t^2\}$ is not.

ARCH

Some further properties:

- $$\mathrm{Var}(y_t) = \alpha_0 + \alpha_1 E(y_{t-1}^2)$$
- If $\alpha_1 < 1$,
$$\mathrm{Var}(y_t) = \frac{\alpha_0}{1-\alpha_1}$$
- and
$$E(y_t^4) = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)},$$
implying that the kurtosis $\kappa > 3$.
- It is required that $\alpha_1^2 < 1/3$, i.e. $\alpha_1 < 0.577$.

ARCH

- Observations $y_1, y_2, ..., y_n$.
- The conditional log likelihood of $y_2, ..., y_n$ given y_1 fulfills (why?)

$$l(\alpha_0, \alpha_1) \propto -\frac{1}{2}\sum_{t=2}^{n} \log(\alpha_0 + \alpha_1 y_{t-1}^2) - \frac{1}{2}\sum_{t=2}^{n} \frac{y_t^2}{\alpha_0 + \alpha_1 y_{t-1}^2}.$$

- May be maximized with numerical methods.
- Changing the normality assumption to e.g. Student's t alters the likelihood.
- Still, it can be maximized with numerical methods.

ARCH

Extension: ARCH(m)

$$
\begin{aligned}
y_t &= \sigma_t \epsilon_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 + \ldots + \alpha_m y_{t-m}^2.
\end{aligned}
$$

GARCH

GARCH(1,1):

- $$y_t = \sigma_t \epsilon_t,$$
  $$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$
- It follows that
  $$y_t^2 = \alpha_0 + (\alpha_1 + \beta_1) y_{t-1}^2 + v_t - \beta_1 v_{t-1},$$
  where $v_t = \sigma_t^2(\epsilon_t^2 - 1)$.
- Not identified if $\alpha_1 = 0$.
- Hence, GARCH(0,1) is not a possible model.

GARCH

Some further properties:

- If $\alpha_1 + \beta_1 < 1$,
$$\mathrm{Var}(y_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$
- and
$$E(y_t^4) = \frac{3\alpha_0^2(1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)\{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2\}},$$
implying that the kurtosis $\kappa > 3$.
- It is required that $1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2 > 0$.
- The sequence $\{y_t^2\}$ has autocorrelation function
$$\rho_n = \alpha_1 \frac{1 - \beta_1^2 - \alpha_1\beta_1}{1 - \beta_1^2 - 2\alpha_1\beta_1}(\alpha_1 + \beta_1)^{n-1}.$$

GARCH

- Recall:
$$y_t^2 = \alpha_0 + (\alpha_1 + \beta_1) y_{t-1}^2 + v_t - \beta_1 v_{t-1},$$
where $v_t = \sigma_t^2(\epsilon_t^2 - 1)$.
- May show: y_t^2 is stationary if $\alpha_1 + \beta_1 < 1$.
- The *integrated* GARCH model, IGARCH, assumes that $\alpha_1 + \beta_1 = 1$:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + (1 - \beta_1) y_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

GARCH

GARCH(m, r):

$$
\begin{aligned}
y_t &= \sigma_t \epsilon_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 + \ldots + \alpha_m y_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_r \sigma_{t-r}^2,
\end{aligned}
$$

where $m \geq r$.

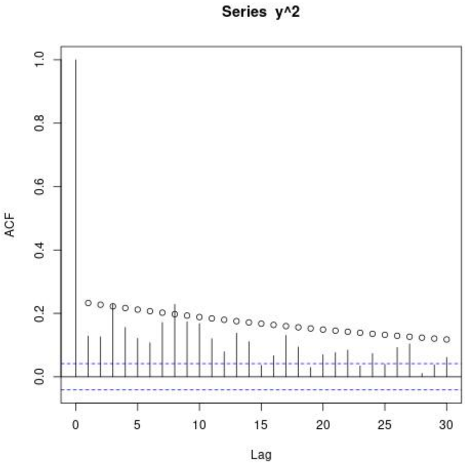
For OMX returns, R library `fGarch`:

```
> garchFit(y~garch(1,1),y,include.mean=FALSE)

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
omega  2.870e-06   8.105e-07    3.541 0.000399 ***
alpha1 9.288e-02   1.419e-02    6.546 5.91e-11 ***
beta1  8.839e-01   1.797e-02   49.185  < 2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
7123.294    normalized:  3.19717
```

`omega` is α_0.
Observe: The sum of estimated α_1 and β_1 is close to 1.

GARCH

Estimated ACF for y_t^2, compared to ACF from estimated GARCH(1,1) model with rings.



Other models

Some more extensions of GARCH$(1,1)$:

- Quadratic GARCH (QGARCH):

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \phi y_{t-1} + \beta_1 \sigma_{t-1}^2,$$

- Threshold GARCH (TGARCH):

$$\sigma_t = \alpha_0 + \alpha_1^+ y_{t-1}^+ + \alpha_1^- y_{t-1}^- + \beta_1 \sigma_{t-1},$$

  where $y_{t-1}^+ = y_{t-1} I\{y_{t-1} > 0\}$, $y_{t-1}^- = y_{t-1} I\{y_{t-1} \leq 0\}$,
- EGARCH:

$$\log(\sigma_t^2) = \alpha_0 + \alpha_1 \{|y_{t-1}| - E(|y_{t-1}|)\} + \gamma y_{t-1} + \beta \log(\sigma_{t-1}^2),$$

- GARCH-M:

$$\begin{cases} y_t = \mu + c\sigma_t^2 + \sigma_t \epsilon_t, \\ \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{cases}$$

- and many more!

Other models

Further reading:

Tsay, R.S. (2010), Analysis of Financial Time Series, 3rd ed., Wiley.

News of today

- ARCH
- GARCH
- IGARCH
- QGARCH
- TGARCH
- EGARCH
- GARCH-M
- ...

Analysis of Time Series, L16

Rolf Larsson

Uppsala University

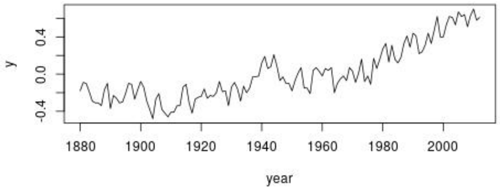
12 maj 2025

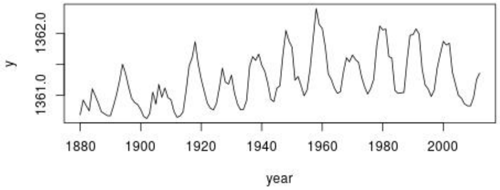
Today

- 5.5: Transfer functions
- 5.6: Multivariate ARMAX

Transfer functions

Global mean temperature (y_t) and solar irradiance (x_t), 1880-2012.





Transfer functions

- Input x_t (e.g. solar irradiance), output y_t (e.g. temperature), *both demeaned*.
- Lagged regression model

$$y_t = \sum_{j=0}^{\infty} \alpha_j x_{t-j} + \eta_t = \alpha(B) x_t + \eta_t,$$

  where $\sum_j |\alpha_j| < \infty$, x_t and η_t stationary and independent.
- How do we estimate the α_j?

Transfer functions

(*i*) Prewhitening of input:

- Find an ARMA model for the input, $\phi(B)x_t = \theta(B)w_t$, where w_t is white noise.
- Hence,

$$y_t = \alpha(B)x_t + \eta_t = \alpha(B)\frac{\theta(B)}{\phi(B)}w_t + \eta_t.$$

(*ii*) Prewhitening of output:

$$\tilde{y}_t = \frac{\phi(B)}{\theta(B)}y_t = \alpha(B)w_t + \tilde{\eta}_t,$$

where

$$\tilde{\eta}_t = \frac{\phi(B)}{\theta(B)}\eta_t, \quad w_t = \frac{\phi(B)}{\theta(B)}x_t.$$

(*iii*) The α_j are given by the CCF (why?)

$$\gamma_{\tilde{y}w}(h) = E(\tilde{y}_{t+h}w_t) = \sigma_w^2\alpha_h.$$

Transfer functions

- $y_t = \alpha(B)x_t + \eta_t$

(*i*) $y_t = \alpha(B)\frac{\theta(B)}{\phi(B)}w_t + \eta_t$

(*ii*) $\tilde{y}_t = \alpha(B)w_t + \tilde{\eta}_t$

(*iii*) CCF $\gamma_{\tilde{y}w}(h) = \sigma_w^2 \alpha_h$.
Try the representation

$$\alpha(B) = \frac{\delta(B)}{\omega(B)} B^d,$$

where

$$\delta(B) = \delta_0 + \delta_1 B + ... + \delta_s B^s,$$
$$\omega(B) = 1 - \omega_1 B - \omega_2 B^2 - ... - \omega_r B^r.$$

Identify d, s and r from the CCF.

Transfer functions

- $y_t = \alpha(B)x_t + \eta_t$

(*i*) $y_t = \alpha(B)\frac{\theta(B)}{\phi(B)}w_t + \eta_t$

(*ii*) $\tilde{y}_t = \alpha(B)w_t + \tilde{\eta}_t$

(*iii*) CCF $\gamma_{\tilde{y}w}(h) = \sigma_w^2 \alpha_h$ gives $y_t = \frac{\delta(B)}{\omega(B)}B^d x_t + \eta_t$

(*iv*) Estimate the regression $\omega(B)y_t = \delta(B)B^d x_t + \omega(B)\eta_t$ i.e.

$$y_t = \omega_1 y_{t-1} + \ldots + \omega_r y_{t-r} + \delta_0 x_{t-d} + \delta_1 x_{t-d-1} + \ldots + \delta_s x_{t-d-s} + u_t,$$

where $u_t = \omega(B)\eta_t$.

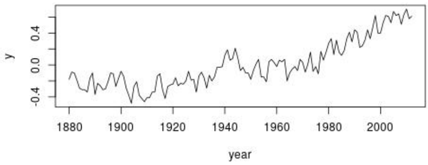
(*v*) Construct $\eta_t = \omega^{-1}(B)u_t$ and fit an ARMA model $\phi_\eta(B)\eta_t = \theta_\eta(B)z_t$ where z_t is white noise.

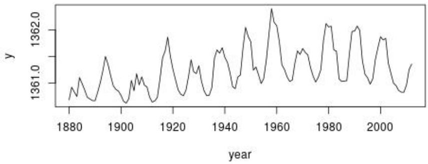
- Final model (why?):

$$\phi_\eta(B)\omega(B)y_t = \phi_\eta(B)\delta(B)B^d x_t + \omega(B)\theta_\eta(B)z_t.$$

Transfer functions

Global mean temperature (y_t) and solar irradiance (x_t), 1880-2012.
Fit a transfer function model for the years 1880-1969.

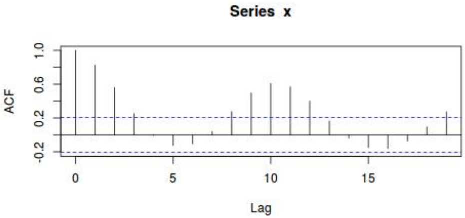


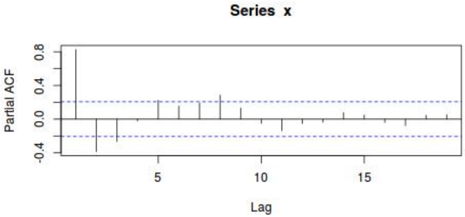


```
> x1=x[seq(1,90)]
> x=x1-mean(x1)
> y1=y[seq(1,90)]
> y=y1-mean(y1)
> par(mfrow=c(2,1))
> acf(x)
> pacf(x)
```

Transfer functions

ACF (tails off?) and PACF (cuts off after lag 3?) for demeaned solar irradiance (x_t):

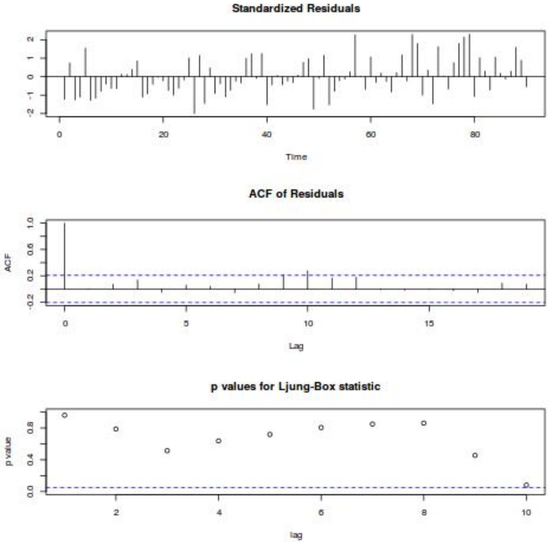




Try AR(3) without constant:

```
> m=arima(x,order=c(3,0,0),include.mean=FALSE)
> m

Call:
arima(x = x, order = c(3, 0, 0), include.mean = FALSE)

Coefficients:
         ar1      ar2      ar3
      1.0785  -0.1091  -0.2584
s.e.  0.1024   0.1553   0.1031

sigma^2 estimated as 0.03541:  log likelihood = 21.73,
aic = -35.45
```

```
> tsdiag(m)
```


Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

Transfer functions

Input: demeaned solar irradiance.

(i)
- Estimated model for input:

$$x_t = 1.0785x_{t-1} - 0.1091x_{t-2} - 0.2584x_{t-3} + w_t,$$

  i.e. $w_t = (1 - 1.0785B + 0.1091B^2 + 0.2584B^3)x_t$.
- Hence,

$$y_t = \alpha(B)\frac{1}{1 - 1.0785B + 0.1091B^2 + 0.2584B^3}w_t + \eta_t.$$

(ii) Prewhitening of output:

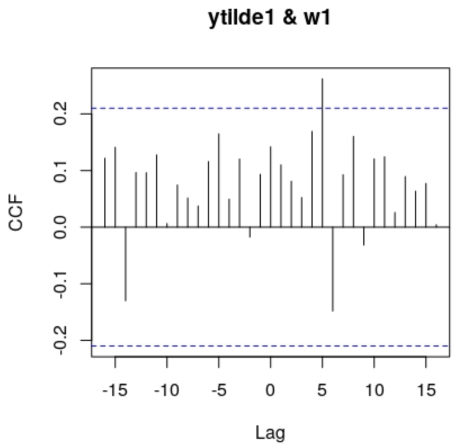
$$\tilde{y}_t = (1 - 1.0785B + 0.1091B^2 + 0.2584B^3)y_t = \alpha(B)w_t + \tilde{\eta}_t$$

(iii) The α_j are given by the CCF $\gamma_{\tilde{y}w}(h) = \sigma_w^2\alpha_h$, for $h \geq 0$.

```
> phi1=as.numeric(m$coef[1])
> phi2=as.numeric(m$coef[2])
> phi3=as.numeric(m$coef[3])
> ytilde=filter(y,filter=c(1,-phi1,-phi2,-phi3),method="convolution",sides=1)
> w=filter(x,filter=c(1,-phi1,-phi2,-phi3),method="convolution",sides=1)
> n=length(y)
> ytilde1=ytilde[seq(4,n)]
> w1=w[seq(4,n)]
> ccf(ytilde1,w1,ylab="CCF")
```

Transfer functions

CCF:



$d = 5$? Slowly decaying to the right means AR form? Trial and error!

Transfer functions

- $$\alpha(B) = \frac{\delta(B)}{\omega(B)} B^d,$$

  where

  $$\delta(B) = \delta_0 + \delta_1 B + \ldots + \delta_s B^s,$$
  $$\omega(B) = 1 - \omega_1 B - \omega_2 B^2 - \ldots - \omega_r B^r.$$

- Trial and error (regressions...):
  Try $d = 1$, $s = 0$, $r = 4$ with zero AR coefficients for lags 2,3., i.e.
  $y_t = \frac{\delta_0}{1-\omega_1 B - \omega_4 B^4} B x_t + \eta_t$.

(iv) Estimate the regression $(1 - \omega_1 B - \omega_4 B^4) y_t = \delta_0 B x_t + u_t$ i.e.

$$y_t = \delta_0 x_{t-1} + \omega_1 y_{t-1} + \omega_4 y_{t-4} + u_t,$$

where $u_t = (1 - \omega_1 B - \omega_4 B^4)\eta_t$.

```
> y0=y[seq(5,n)]
> y1=y[seq(4,n-1)]
> y4=y[seq(1,n-4)]
> x1=x[seq(4,n-1)]
> r=lm(y0~x1+y1+y4-1);summary(r)

Call:
lm(formula = y0 ~ x1 + y1 + y4 - 1)

Residuals:
      Min        1Q    Median        3Q       Max
-0.211752 -0.074580  0.000089  0.074073  0.206602

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
x1  0.05830    0.02931   1.989   0.0500 *
y1  0.60071    0.08371   7.177 2.76e-10 ***
y4  0.19932    0.08188   2.434   0.0171 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09714 on 83 degrees of freedom
Multiple R-squared:  0.6243,Adjusted R-squared:  0.6107
F-statistic: 45.98 on 3 and 83 DF,  p-value: < 2.2e-16
```

Transfer functions

(iv) Estimated regression

$$y_t = 0.058x_{t-1} + 0.60y_{t-1} + 0.20y_{t-4} + u_t,$$

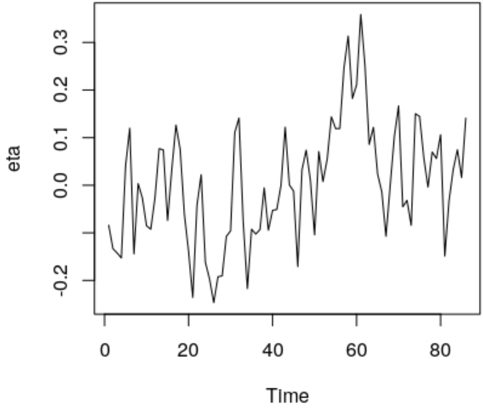
where $u_t = (1 - 0.60B - 0.20B^4)\eta_t$.

(v) Construct $\eta_t = (1 - 0.60B - 0.20B^4)^{-1}u_t$ and fit an ARMA model $\phi_\eta(B)\eta_t = \theta_\eta(B)z_t$ where z_t is white noise.

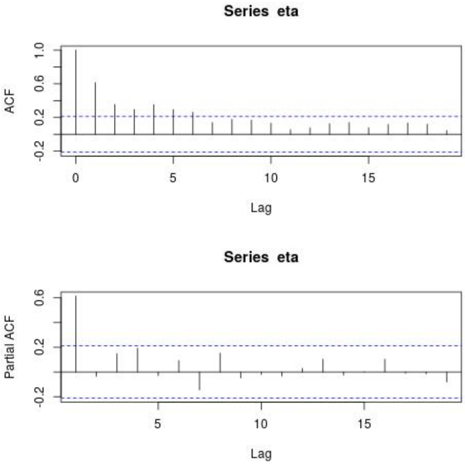
```
> u=r$res
> omega1=as.numeric(r$coef[2])
> omega2=as.numeric(r$coef[3])
> eta=filter(u,filter=c(omega1,0,0,omega2),method="recursive")
> plot(eta,type='l')
> par(mfrow=c(2,1))
> acf(eta)
> pacf(eta)
```

Transfer functions

Plot of η_t



Transfer functions



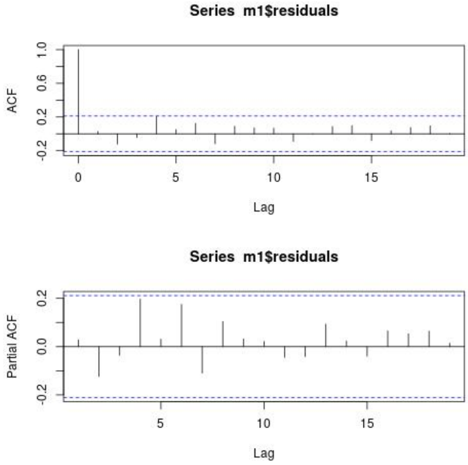
The ACF tails off and the PACF cuts off after lag 1.

Try AR(1). Model estimation for η_t:

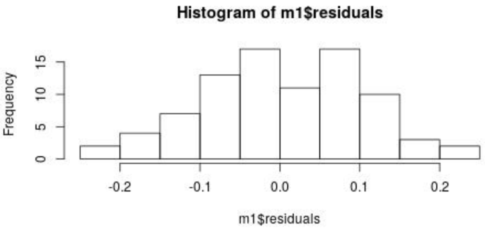
```
> m1=arima(eta,order=c(1,0,0),include.mean=FALSE);m1

Call:
arima(x = eta, order = c(1, 0, 0), include.mean = FALSE)

Coefficients:
         ar1
      0.6193
s.e.  0.0843

sigma^2 estimated as 0.00956:  log likelihood = 77.68,
aic = -151.37
```

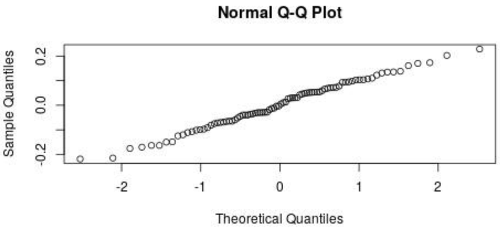
Analyse residuals of η_t model (white noise?):

```
> par(mfrow=c(2,1))
> acf(m1$res)
> pacf(m1$res)
```

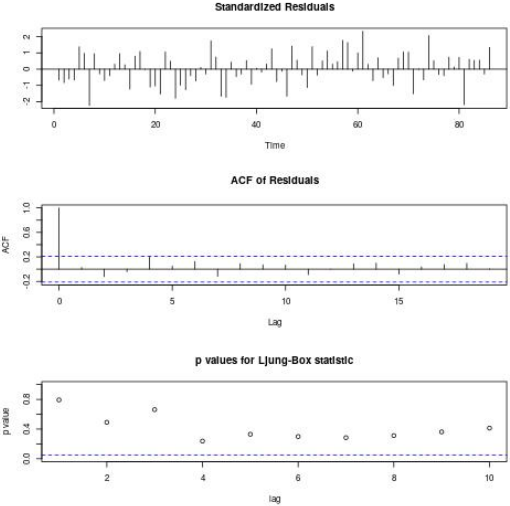


```
> hist(m1$res)
> qqnorm(m1$res)
```





```
> tsdiag(m1)
```


Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

Transfer functions

- The final model

$$\phi_\eta(B)\omega(B)y_t = \phi_\eta(B)\delta(B)B^d x_t + \omega(B)\theta_\eta(B)z_t$$

is estimated as

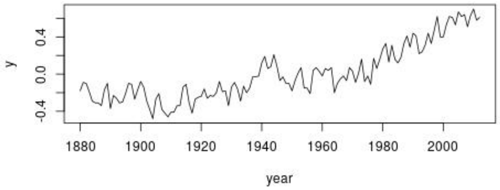
$$\begin{aligned}&(1-0.62B)(1-0.60B-0.20B^4)y_t\\ &= (1-0.62B)0.058Bx_t + (1-0.60B-0.20B^4)z_t,\end{aligned}$$

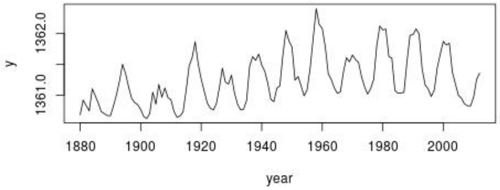
where $z_t = (1-0.62B)\eta_t$ is white noise.

- White noise properties are already checked, since the z_t are the residuals of the η_t model!

Transfer functions

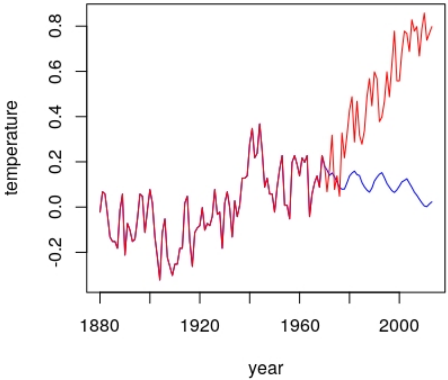
Global mean temperature (y_t) and solar irradiance (x_t), 1880-2012.





Transfer functions

Demeaned observations 1880-2013 (in red) and prediction 1970-2013 (in blue) using known x_t values and updated (predicted) y_t values.



Multivariate ARMAX

- Multivariate regression

$$\begin{cases} y_{t1} &= \beta_{11} z_{t1} + \ldots + \beta_{1r} z_{tr} + w_{t1}, \\ &\vdots \\ y_{tk} &= \beta_{k1} z_{t1} + \ldots + \beta_{kr} z_{tr} + w_{tk}. \end{cases}$$

- i.e.

$$\begin{pmatrix} y_{t1} \\ y_{t2} \\ \vdots \\ y_{tk} \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} & \ldots & \beta_{1r} \\ \beta_{21} & \beta_{22} & \ldots & \beta_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \ldots & \beta_{kr} \end{pmatrix} \begin{pmatrix} z_{t1} \\ z_{t2} \\ \vdots \\ z_{tr} \end{pmatrix} + \begin{pmatrix} w_{t1} \\ w_{t2} \\ \vdots \\ w_{tk} \end{pmatrix}$$

i.e. $\mathbf{y}_t = B\mathbf{z}_t + \mathbf{w}_t$.

- For example, with $r = 1$ and $k = 2$,

$$\begin{cases} y_{t1} &= \beta_{11} z_{t1} + w_{t1}, \\ y_{t2} &= \beta_{21} z_{t1} + w_{t2}, \end{cases} \qquad \begin{pmatrix} y_{t1} \\ y_{t2} \end{pmatrix} = \begin{pmatrix} \beta_{11} \\ \beta_{12} \end{pmatrix} z_{t1} + \begin{pmatrix} w_{t1} \\ w_{t2} \end{pmatrix}$$

Multivariate ARMAX

VAR(1):

$$\begin{pmatrix} x_{t1} \\ x_{t2} \\ \vdots \\ x_{tk} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1k} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{k1} & \phi_{k2} & \cdots & \phi_{kk} \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \\ \vdots \\ x_{t-1,k} \end{pmatrix} + \begin{pmatrix} w_{t1} \\ w_{t2} \\ \vdots \\ w_{tk} \end{pmatrix}$$

i.e. $\mathbf{x}_t = \mathbf{\Phi} \mathbf{x}_{t-1} + \mathbf{w}_t$.

Multivariate ARMAX

Extensions:

- VAR(p)

$$\mathbf{x}_t = \sum_{j=1}^{p} \mathbf{\Phi}_j \mathbf{x}_{t-j} + \mathbf{w}_t.$$

- VARX(p)

$$\mathbf{x}_t = \Gamma \mathbf{u}_t + \sum_{j=1}^{p} \mathbf{\Phi}_j \mathbf{x}_{t-j} + \mathbf{w}_t,$$

  where Γ is $k \times r$, $\mathbf{u}_t$ is $r \times 1$.

- Example with $r = 2$:

$$\Gamma \mathbf{u}_t = \begin{pmatrix} \gamma_0 & \gamma_1 \end{pmatrix} \begin{pmatrix} 1 \\ t \end{pmatrix} = \gamma_0 + \gamma_1 t.$$

Multivariate ARMAX

VARMA models:

- VARMA(1,1)

$$\mathbf{x}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{w}_t + \mathbf{\Theta}\mathbf{w}_{t-1}.$$

- Unicity problem: VARMA(1,1) with

$$\mathbf{\Phi} = \begin{pmatrix} 0 & \phi + \theta \\ 0 & 0 \end{pmatrix}, \quad \mathbf{\Theta} = \begin{pmatrix} 0 & -\theta \\ 0 & 0 \end{pmatrix}$$

is equivalent to VARMA(1,0), $\mathbf{x}_t = \mathbf{\Phi}_1\mathbf{x}_{t-1} + \mathbf{w}_t$, with

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0 & \phi \\ 0 & 0 \end{pmatrix}.$$

Why?

Multivariate ARMAX

VAR(1)

- $$\mathbf{x}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{w}_t.$$
- Equivalent: Error correction form

$$\nabla \mathbf{x}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{w}_t - \mathbf{x}_{t-1} = \mathbf{\Phi}_1\mathbf{x}_{t-1} + \mathbf{w}_t$$

  where $\mathbf{\Phi}_1 = \mathbf{\Phi} - I$, with I as the identity matrix.
- If $\mathbf{\Phi}$ has (at least) one eigenvalue that is equal to one, then $\mathbf{\Phi}_1$ has (at least) one eigenvalue that is equal to zero. Then, $\mathbf{\Phi}_1$ is singular and $\mathbf{x}_t$ is non stationary.

Multivariate ARMAX

VAR(1), dimension 2:

- Error correction form

$$\begin{pmatrix} \nabla x_{t1} \\ \nabla x_{t2} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \end{pmatrix} + \begin{pmatrix} w_{t1} \\ w_{t2} \end{pmatrix}$$

- Reduced rank (cointegration)

$$\begin{pmatrix} \nabla x_{t1} \\ \nabla x_{t2} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \end{pmatrix} + \begin{pmatrix} w_{t1} \\ w_{t2} \end{pmatrix}$$

- i.e.

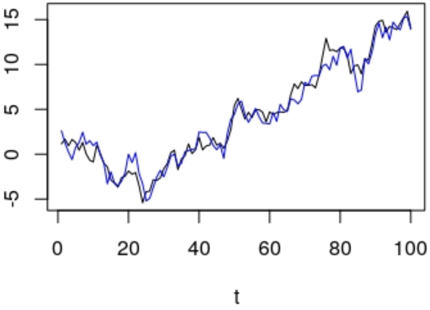
$$\begin{aligned} \nabla x_{t1} &= \alpha_1(\beta_1 x_{t-1,1} + \beta_2 x_{t-1,2}) + w_{t1}, \\ \nabla x_{t2} &= \alpha_2(\beta_1 x_{t-1,1} + \beta_2 x_{t-1,2}) + w_{t2}. \end{aligned}$$

- *Cointegrating relation* $\beta_1 x_{t-1,1} + \beta_2 x_{t-1,2}$.

Multivariate ARMAX

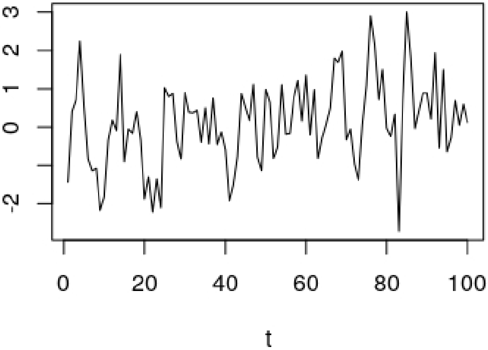
Simulation example (x_{t1} in black, x_{t2} in blue):

$$\nabla x_{t1} = w_{t1},$$
$$\nabla x_{t2} = 0.5(x_{t-1,1} - x_{t-1,2}) + w_{t2}.$$



Multivariate ARMAX

Plot of the cointegrating relation, $x_{t1} - x_{t2}$ (observe the different scale on the y axis):



Multivariate ARMAX

Clive Granger, Nobel prize 2003.

https://www.nobelprize.org/nobelprizes/economic-sciences/laureates/2003/granger-photo.html

Multivariate ARMAX

Sören Johansen

http://www.economics.ku.dk/staff/vip/?pure=en/persons/34220

News of today

- Transfer function modeling
- VAR
- VARX
- VARMA
- Cointegration

Analysis of Time Series, L17

Rolf Larsson

Uppsala University

14 maj 2025

Today

Preparation:

- Appendix B: Multivariate normal distribution
- The companion form

Chap. 6:

- State-Space Models
- The Kalman Filter
- The EM algorithm
- Maximum Likelihood

Multivariate normal distribution

p.497: Let $\mathbf{y} = (y_1, ..., y_m)'$, $\mathbf{x} = (x_1, ..., x_n)'$ and suppose that

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \right\}$$

.

Then, $\mathbf{y}|\mathbf{x} \sim N\left(\boldsymbol{\mu}_{y|x}, \Sigma_{y|x}\right)$ with

$$\boldsymbol{\mu}_{y|x} = \boldsymbol{\mu}_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x),$$
$$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}.$$

The companion form

The AR(p) model

$$x_t = \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + w_t$$

may be written in terms of one lag as

$$\begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & & \phi_p \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

or in short, $\mathbf{x}_t = \mathbf{\Phi} \mathbf{x}_{t-1} + \mathbf{w}_t$.

This is also called *the companion form*.

State-Space Models

- General model

$$\mathbf{y}_t = A_t\mathbf{x}_t + \Gamma\mathbf{u}_t + \mathbf{v}_t, \quad \text{(observation equation)}$$
$$\mathbf{x}_t = \Phi\mathbf{x}_{t-1} + \Upsilon\mathbf{u}_t + \mathbf{w}_t, \quad \text{(state equation)}$$

where $E(\mathbf{v}_t\mathbf{v}_t') = R$, $E(\mathbf{w}_t\mathbf{w}_t') = Q$.

- Example 1: MA(1) with zero mean

$$y_t = \begin{pmatrix} 1 & \theta \end{pmatrix} \begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix},$$
$$\begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} w_{t-1} \\ w_{t-2} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \end{pmatrix},$$

i.e. $A_t = \begin{pmatrix} 1 & \theta \end{pmatrix}$, $\mathbf{x}_t = \begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix}$, $\Phi = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathbf{w}_t = \begin{pmatrix} w_t \\ 0 \end{pmatrix}$,
$\Gamma = 0$, $\mathbf{v}_t = 0$, $R = 0$, $\Upsilon = 0$, $Q = \sigma_w^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

State-Space Models

General:

- $$\begin{aligned}\mathbf{y}_t &= A_t\mathbf{x}_t + \Gamma\mathbf{u}_t + \mathbf{v}_t, \quad \text{(observation equation)}\\ \mathbf{x}_t &= \Phi\mathbf{x}_{t-1} + \Upsilon\mathbf{u}_t + \mathbf{w}_t. \quad \text{(state equation)}\end{aligned}$$
- The data up to time s is $Y_s = \{\mathbf{y_1}, ..., \mathbf{y_s}\}$.
- Conditional expectation
$$\mathbf{x}_t^s = E(\mathbf{x}_t | Y_s)$$
- Mean square error matrix
$$P_t^s = E\{(\mathbf{x}_t - \mathbf{x}_t^s)(\mathbf{x}_t - \mathbf{x}_t^s)'\}$$
- *Forecasting* when $s < t$, *filtering* when $s = t$, *smoothing* when $s > t$.

The Kalman Filter

Theorem (Property 6.1 ($\Upsilon = \Gamma = 0$))

$$\begin{aligned} \mathbf{x}_t^{t-1} &= \Phi \mathbf{x}_{t-1}^{t-1}, \\ P_t^{t-1} &= \Phi P_{t-1}^{t-1} \Phi' + Q, \end{aligned}$$

where

$$\begin{aligned} \mathbf{y}_t^{t-1} &= A_t \mathbf{x}_t^{t-1}, \\ \Sigma_t &= E\{(\mathbf{y}_t - \mathbf{y}_t^{t-1})(\mathbf{y}_t - \mathbf{y}_t^{t-1})'\} = A_t P_t^{t-1} A_t' + R, \\ K_t &= P_t^{t-1} A_t' (A_t P_t^{t-1} A_t' + R)^{-1}, \\ \mathbf{x}_t^t &= \mathbf{x}_t^{t-1} + K_t(\mathbf{y}_t - A_t \mathbf{x}_t^{t-1}), \\ P_t^t &= (I - K_t A_t) P_t^{t-1}. \end{aligned}$$

The Kalman Filter

Example 1: MA(1) with zero mean

- $$y_t = \begin{pmatrix} 1 & \theta \end{pmatrix} \begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix},$$
  $$\begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} w_{t-1} \\ w_{t-2} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \end{pmatrix},$$
  i.e. $A_t = \begin{pmatrix} 1 & \theta \end{pmatrix}$, $\mathbf{x}_t = \begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix}$, $\Phi = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathbf{w}_t = \begin{pmatrix} w_t \\ 0 \end{pmatrix}$,
  $\mathbf{v}_t = 0$, $R = 0$, $Q = \sigma_w^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.
- Calculate forecasts $y_1^0, y_2^1, \ldots$ and variances of forecast errors, $\Sigma_1, \Sigma_2, \ldots$.

The EM algorithm

General:

- We observe $\mathbf{y} = (y_1, y_2, ..., y_n)$.
- These are "proxies" for the (partially) unobserved $\mathbf{x} = (x_1, x_2, ..., x_n)$, which is a sample from a distribution with parameter θ.
- Estimate θ.
- Initial estimate $\hat{\theta}_0$.
- The EM algorithm:

  E1: $\hat{\mathbf{y}}_1 = E(\mathbf{x}; \theta = \hat{\theta}_0)$
  M1: $\hat{\theta}_1$ is the MLE based on $\hat{\mathbf{y}}_1$.
  E2: $\hat{\mathbf{y}}_2 = E(\mathbf{x}; \theta = \hat{\theta}_1)$
  M2: $\hat{\theta}_2$ is the MLE based on $\hat{\mathbf{y}}_2$.
  ... Repeat until convergence.
- May be shown to give the MLE of θ under mild regularity conditions.

The EM algorithm

Example 2:

- We observe $y_1 = 1$, $y_2 = 2$, $y_3 = 6$ from a truncated Exponential distribution with parameter (intensity) λ.
  (Corresponds to x_1, x_2, x_3 from non truncated dist.)
- The distribution is truncated at 6.
- $X \sim \mathrm{Exp}(\lambda) \Rightarrow E(X|X > a) = a + \frac{1}{\lambda}$ (why?)
- Estimate λ.
- MLE for non truncated data: $\hat{\lambda} = 1/\bar{y} = 3/\sum_j y_j$.
- Initial estimate $\hat{\lambda}_0 = 3/(1+2+6) = 1/3 \approx 0.3333$.
- *The EM algorithm*:

  E1: $\hat{y}_3 = E(x_3; \lambda = 1/3) = 6 + 1/(1/3) = 9$
  M1: $\hat{\lambda}_1 = 3/(1+2+9) = 1/4 = 0.2500$
  E2: $\hat{y}_3 = E(x_3; \lambda = 1/4) = 6 + 1/(1/4) = 10$
  M2: $\hat{\lambda}_2 = 3/(1+2+10) = 3/13 = 0.2308$
  ... Going on similarly yields $\hat{\lambda}_3 = 9/40 = 0.2250$, $\hat{\lambda}_4 = 27/121 = 0.2231$, $\hat{\lambda}_5 = 81/364 = 0.2225$, $\hat{\lambda}_6 = 243/1093 = 0.2223$, ...

MLE via the EM algorithm

- Suppose that $t = 1, 2, ...n$,
  $\mathbf{y}_t = A_t \mathbf{x}_t + \mathbf{v}_t$, where $\mathbf{v}_t$ is normal with mean 0 and cov. matrix R,
  $\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t$, where $\mathbf{w}_t$ is normal with mean 0 and cov. matrix Q.
- Parameter vector Θ.
- Likelihood if the $\mathbf{x}_t$ are observed

$$L(\Theta) = f_{\mu_0, \Sigma_0}(\mathbf{x}_0) \prod_{t=1}^{n} f_{\Phi, Q}(\mathbf{x}_t | \mathbf{x}_{t-1}) \prod_{t=1}^{n} f_R(\mathbf{y}_t | \mathbf{x}_t).$$

- From the normal density function,

$$\begin{aligned} -2 \log L(\Theta) \propto \log |\Sigma_0| &+ (\mathbf{x}_0 - \boldsymbol{\mu}_0)' \Sigma_0^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}_0) \\ &+ n \log |Q| + \sum_{t=1}^{n} (\mathbf{x}_t - \Phi \mathbf{x}_{t-1})' Q^{-1} (\mathbf{x}_t - \Phi \mathbf{x}_{t-1}) \\ &+ n \log |R| + \sum_{t=1}^{n} (\mathbf{y}_t - A_t \mathbf{x}_t)' R^{-1} (\mathbf{y}_t - A_t \mathbf{x}_t). \end{aligned}$$

MLE via the EM algorithm

E step, under assumed parameter values (p.314-316):

$$
\begin{aligned}
E\left\{-2\log L(\Theta)\right\} \propto\ & \log|\Sigma_0| + \operatorname{tr}\left[\Sigma_0^{-1}\left\{P_0^n + (\mathbf{x}_0^n - \boldsymbol{\mu}_0)(\mathbf{x}_0^n - \boldsymbol{\mu}_0)'\right\}\right] \\
& + n\log|Q| + \operatorname{tr}\left[Q^{-1}(S_{11} - S_{10}\Phi' - \Phi S_{10}' + \Phi S_{00}\Phi')\right] \\
& + n\log|R| \\
& + \operatorname{tr}\left[R^{-1}\sum_{t=1}^{n}\left\{(\mathbf{y}_t - A_t\mathbf{x}_t^n)(\mathbf{y}_t - A_t\mathbf{x}_t^n)' + A_t P_t^n A_t'\right\}\right],
\end{aligned}
$$

$$
S_{11} = \sum_{t=1}^{n}(\mathbf{x}_t^n\mathbf{x}_t^{n\prime} + P_t^n),
$$

$$
S_{10} = \sum_{t=1}^{n}(\mathbf{x}_t^n\mathbf{x}_{t-1}^{n}{}' + P_{t,t-1}^n),
$$

$$
S_{00} = \sum_{t=1}^{n}(\mathbf{x}_{t-1}^n\mathbf{x}_{t-1}^{n}{}' + P_{t-1}^n).
$$

MLE via the EM algorithm

M step (minimize the expression on the previous slide):

$$\hat{\Phi} = S_{10} S_{00}^{-1},$$
$$\hat{Q} = n^{-1}(S_{11} - S_{10} S_{00}^{-1}),$$
$$\hat{R} = n^{-1} \sum_{t=1}^{n} \left\{ (\mathbf{y}_t - A_t \mathbf{x}_t^n)(\mathbf{y}_t - A_t \mathbf{x}_t^n)' + A_t P_t^n A_t' \right\}$$
$$\hat{\boldsymbol{\mu}}_0 = \mathbf{x}_0^n,$$
$$\hat{\Sigma}_0 = P_0^n.$$

This gives new assumed parameter values in the next E step.
Repeat until convergence.

News of today

- State-Space Models
- The Kalman Filter
  - Forecasting
  - Smoothing
  - MLE via the EM algorithm