Time Series Analysis

Regina Kaiser
Statistic Department, Carlos III University.
e-mail: regina.kaiser@uc3m.es

2020

Syllabus.

- Chapter 1. Concepts in time series.
- Chapter 2. Univariate ARIMA models.
- Chapter 3. Model fitting and checking.
- Chapter 4. Prediction and model selection.
- Chapter 5. Outliers and influential observations.
- Chapter 6. Heterocedastic models.
- Chapter 7. Multivariate time series.

Grading

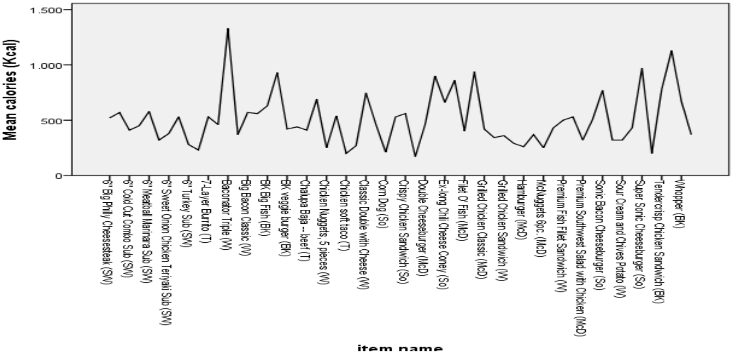
- Final exam 40 %. (optionally: two midterms with grade ≥ 7)
- Weekly laboratory 20 %
- Group empirical project 40 %.

Motivation

Time series data: is a sequence of observations taken at regular intervals of time. We can find data in:

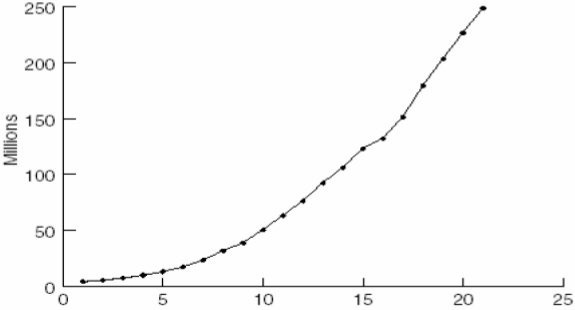
- Business (sales figures, production numbers,customer frequencies...),
- Economics (stock prices, exchanges rates, interest rates...),
- Environmetrics(precipitation,temperature or pollution recordings...),
- Natural sciencies (population sizes, sunspot activity, chemical process data...).

- Basic data analysis is based on the assumption of identically and independently distributed data (current observation has no information about future observations)



Motivation

- Time series are correlated data (current observation has information about future observations)



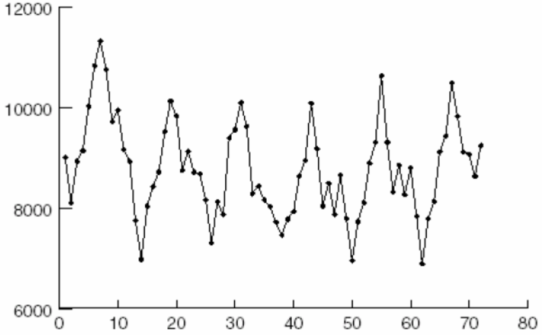
Population of the a U.S.A. at ten-year intervals, 1790-1990

Motivation

- The purpose of time series analysis is to visualize and understand these dependences in past data, and to exploit them for forecasting future values.
- While some simple descriptive techniques do often considerably enhance the understanding of the data, a full analysis usually involves modeling the stochastic mechanism that is assumed to be the generator of the observed time series.
- Once a good model is found and fitted to data, the analyst can use the model to forecast future values and produce prediction intervals, or he can generate simulations, for example to guide planning decisions.

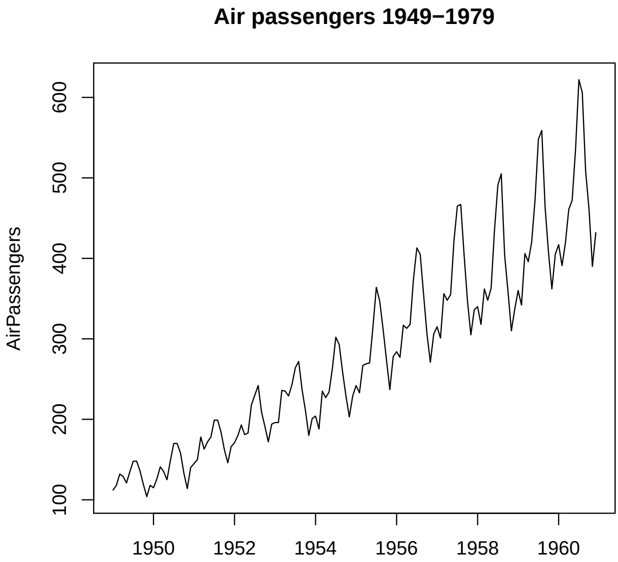
Motivation

- The dominant main features of many time series are trend and seasonal variation. These can be either be modeled deterministically by mathematical functions of time, or are estimated using non-parametric smoothing approaches.



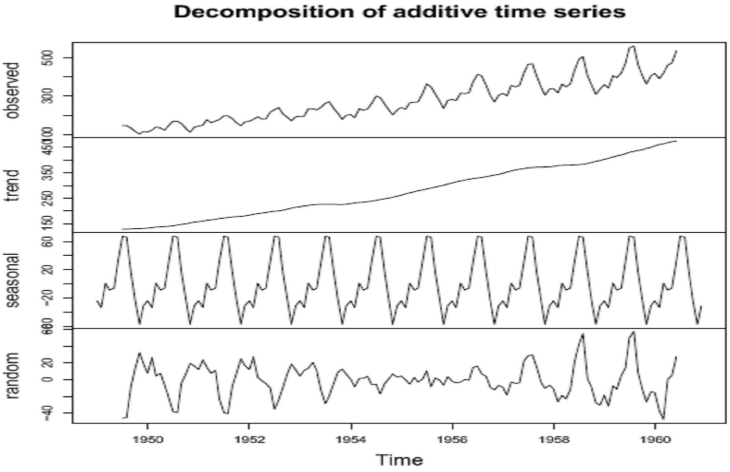
The monthly accidental deaths data, 1973-1978

Examples



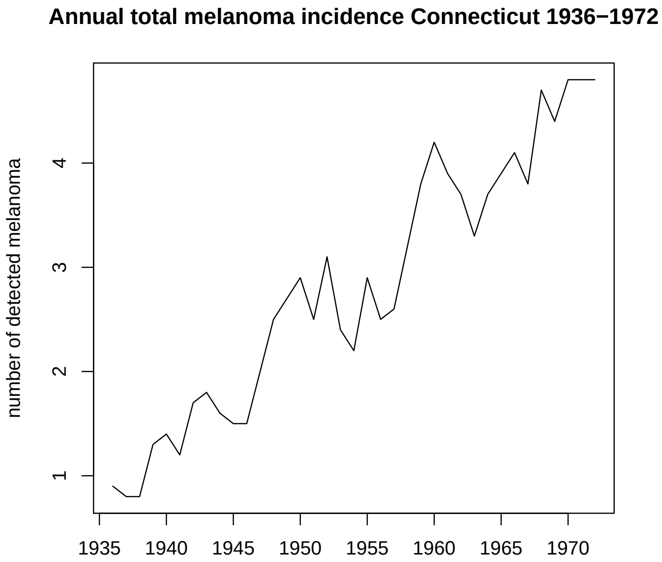
Air passengers 1949−1979
AirPassengers
100
200
300
400
500
600
1950
1952
1954
1956
1958
1960

Examples



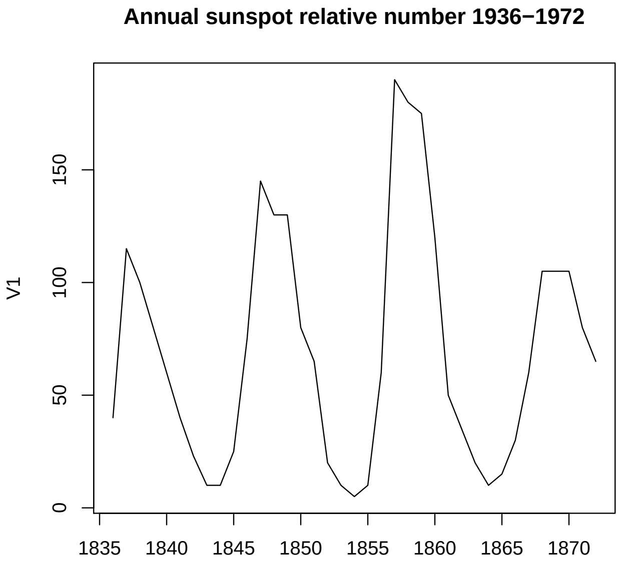
Decomposition of additive time series
observed
trend
seasonal
random
Time
1950
1952
1954
1956
1958
1960

Examples

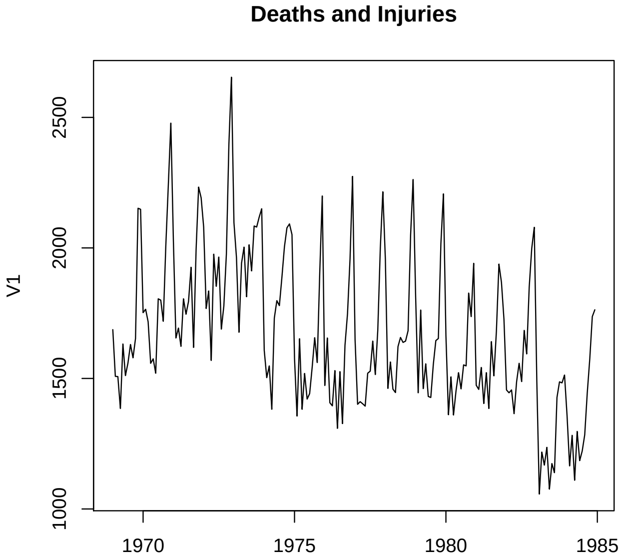


Annual total melanoma incidence Connecticut 1936−1972
number of detected melanoma
1
2
3
4
1935
1940
1945
1950
1955
1960
1965
1970

Examples

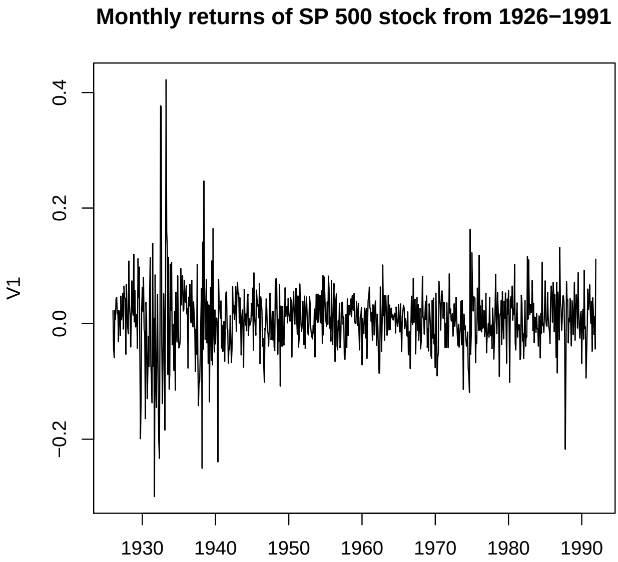

Annual sunspot relative number 1936−1972
V1
0
50
100
150
1835
1840
1845
1850
1855
1860
1865
1870

Examples



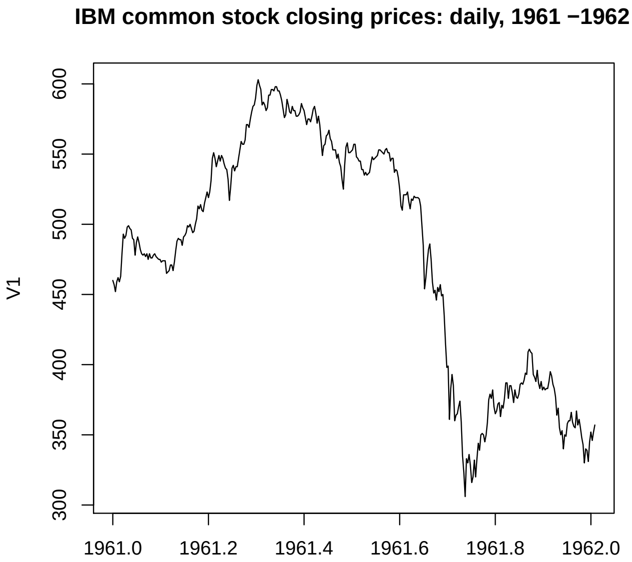
Deaths and Injuries
V1
1000
1500
2000
2500
1970
1975
1980
1985

Examples

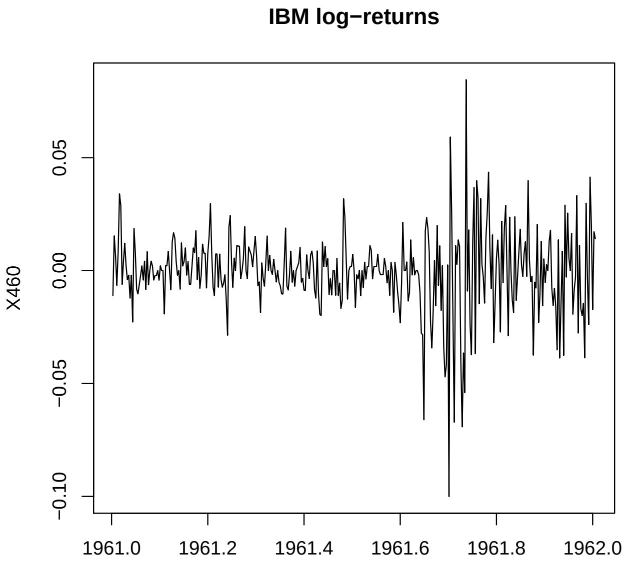


Monthly returns of SP 500 stock from 1926−1991
V1
0.4
0.2
0.0
−0.2
1930
1940
1950
1960
1970
1980
1990

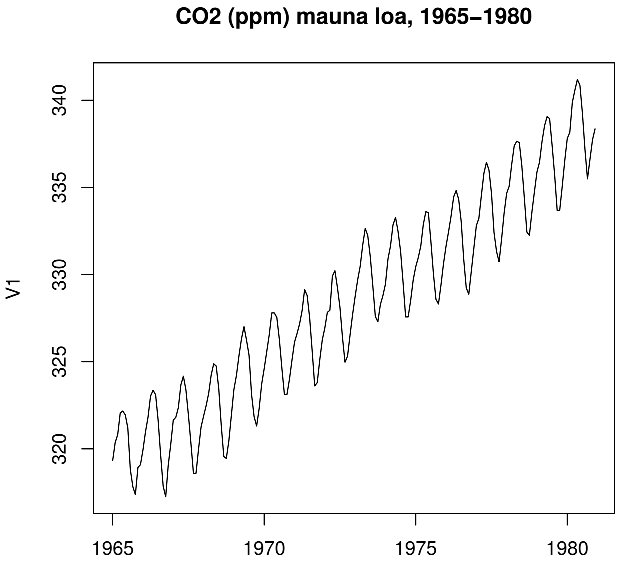
Examples


IBM common stock closing prices: daily, 1961 −1962
V1
300
350
400
450
500
550
600
1961.0
1961.2
1961.4
1961.6
1961.8
1962.0

Examples


IBM log−returns
X460
0.05
0.00
−0.05
−0.10
1961.0
1961.2
1961.4
1961.6
1961.8
1962.0

Examples



CO2 (ppm) mauna loa, 1965−1980
V1
340
335
330
325
320
1965
1970
1975
1980

Goals in time series

- Exploratory analysis: visualization, decomposition into deterministic and stochastic parts and study of the dependence structure of the data.
- Modeling: formulation of the stochastic model, estimation of the paramaters, model diagnosis and evaluation.
- Forecasting: prediction of future observations in the series.
- Time series regression: relations with explanatory series. Outliers
- Process control: optimal management and quality control. With time series becomes feasible to monitor which fluctuations are normal, and which require an intervention.

Chapter 1. Concepts in time series.

Definition of a time series.

- **Definition** A time series process is a set of random variables $\{X_t, t \in T\}$, where T is the set of times at which the process can be observed. We assume $X_t \sim F_t$.
- We exclusively consider time series processes with equidistant time intervals, as well as real-valued random variables.
- An observed time series is a realization of the n- dimensional random vector

$$X = (X_1, X_2, ..., X_n),$$

and is denoted with small letters

$$x = (x_1, x_2, ...x_n)$$

.

Definition of a time series.

- In time series we have only (a part of) one multivariate observation. Thus we have to base our inference on a sample of dimension one.Therefore, we will need to impose some conditions on the joint distribution F.
- Since most social and natural phenomena seem to evolve smoothly rather than by abrupt changes, modeling them by time homogeneous processes is a reasonable approximation.
- The most important form of time-homogeneity in time series analysis is stationarity.

Stationarity.

- In colloquial language, stationarity means that the probabilistic character of the series must not change over time, i.e. any section of the time series is "typical" for every other section with the same length.
- Mathematically, we require that for any indices s, t,and k, the observations $x_t, \ldots, x_{t+k}$ could have just as easily ocurred at times $s, \ldots, s+k$.

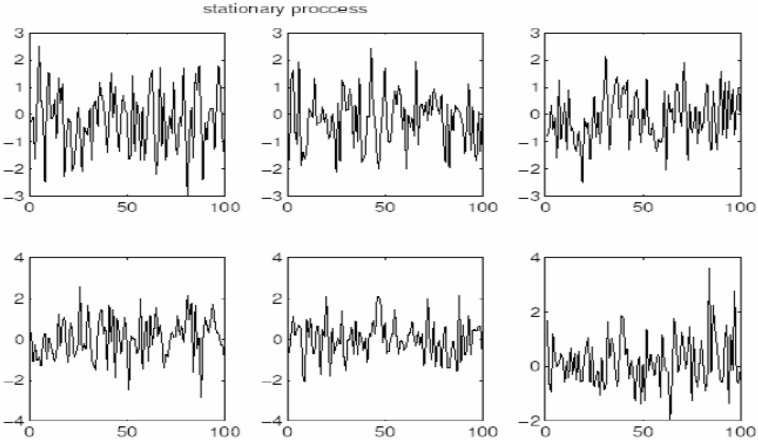
Stationarity.

- A time series is said to be Strictly stationarity if and only if the $(k+1)$-dimensional joint distribution of $X_t, \ldots, X_{t+k}$ coincides with the joint distribution of $X_s, \ldots, X_{s+k}$.
- With real data, since we only have one single observation is impossible to verify strict stationarity.
- We should, then verify the condition of Weak stationarity: constant unconditional mean and variance and dependency that only depends on the lag.

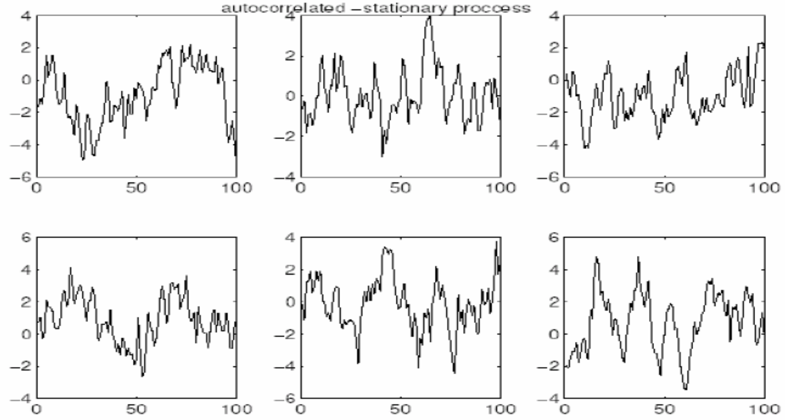
Weak Stationarity

- Some obvious violations of weak stationarity are trends, non-constant variance, deterministics seasonal variation, breaks in the data...
- Stationarity is a hypothesis which is tested on data.
- There are formal tests that focus on some very specificic non-stationarity aspect (for example DF test for the presence of a trend), but they do not test stationarity in a broad sense.
- It is, therefore, recommended assessing stationarity by visual inspection.

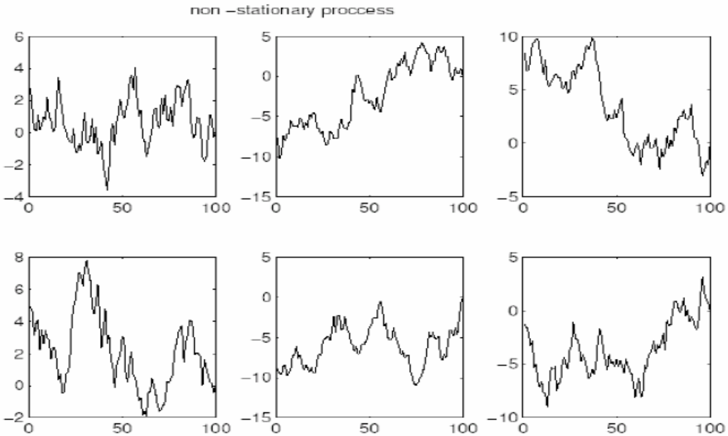
Weak stationarity



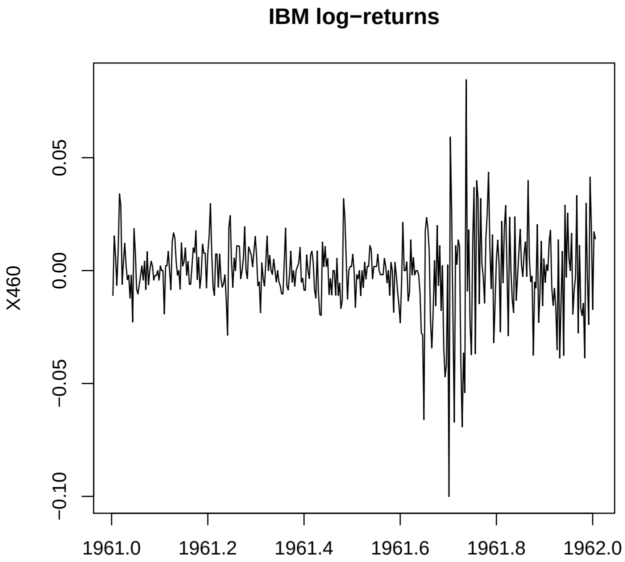
Weak stationarity



Weak stationarity

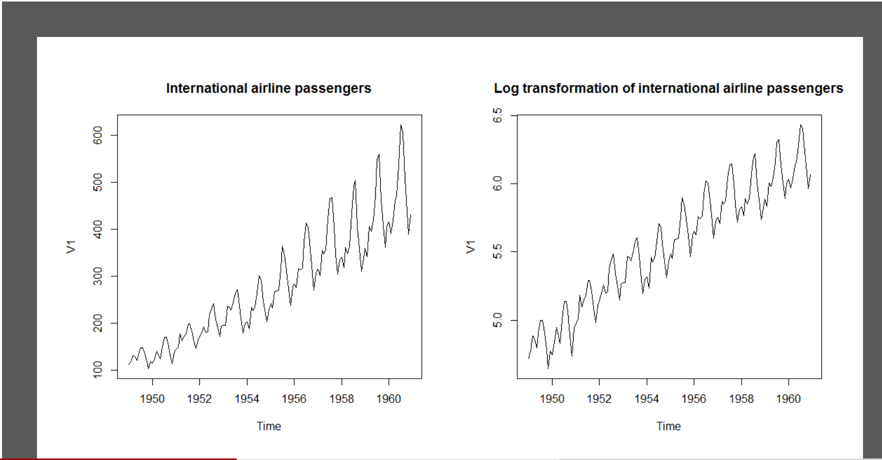


Stationarity in mean but not in variance



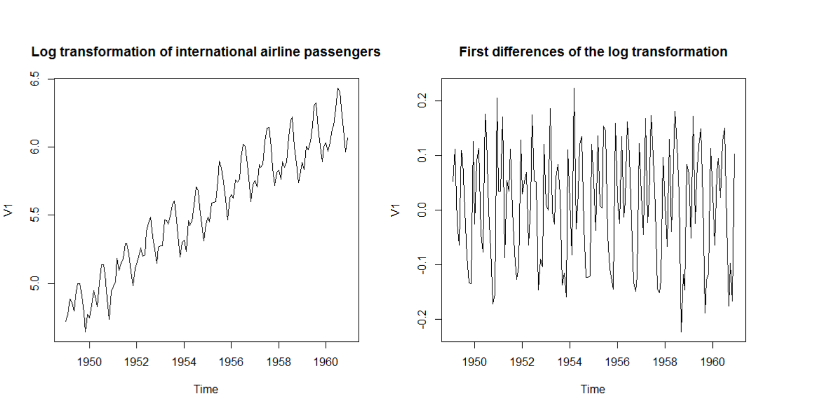
Transformations to achieve stationarity. Logarithms.

- The most popular and practically relevant transformation is $g(.) = log(.)$. It is indicated if either the variation in the series grows with increasing mean, or if the marginal distribution appears to be right-skewed.

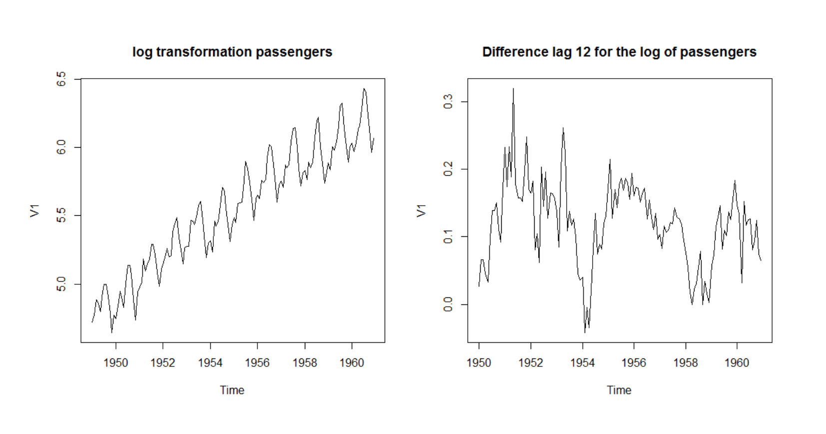


Stationarity by Regular Differences.

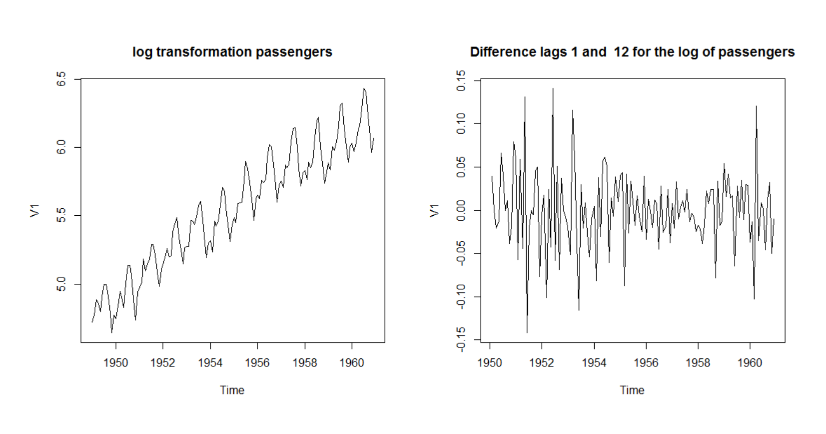
- A simple approach to achieve stationarity in mean is by taking differences. A practical interpretation of taking differences is that by doing so, the changes in the data will be monitored, but not longer the series itself.



Stationarity by Seasonal Differences.



Stationarity by Seasonal and Regular Differences.



Transformations to achieve stationarity. Differencing.

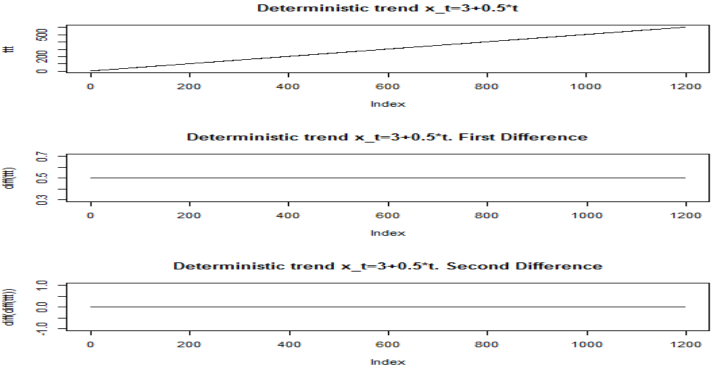
- Let B the Backward/backshift operator, such that:

$$B^j x_t = x_{t-j}$$

- We shall use the operators:
  - Regular difference: $\nabla = 1 - B$
  - Seasonal difference: $\nabla_S = 1 - B^S$, where S is the frequency.

Transformations to achieve stationarity. Differencing.

- If x_t is a deterministic trend $x_t = a + b * t$, $\nabla x_t = b$ and $\nabla^2 x_t = 0$. In general, ∇^d will reduce a polynomial of degree d to a constant.



Transformations to achieve stationarity. Log-transformation and differencing.

- If a time series that was log-transformed is differenced with lag 1, we obtain the so-called log-return, which is an approximation to the relative change,i.e. the percent increase or decrease with respect to the previous instance. In particular:

$$Y_t = log(X_t) - log(X_{t-1}) = log\left(\frac{X_t}{X_{t-1}}\right) = log\left(\frac{X_t - X_{t-1}}{X_{t-1}} + 1\right)$$

$$\approx \frac{X_t - X_{t-1}}{X_{t-1}}$$

Autocorrelation.Visualization of the dependence.

- The autocorrelation between two random variables X_{t+k} and X_t is defined as:

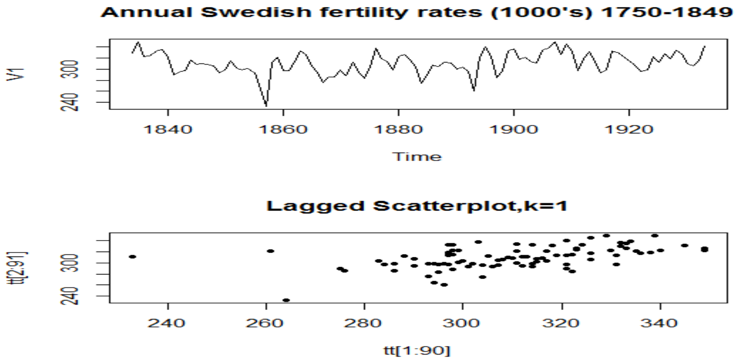
$$Cor(X_{t+k}, X_t) = \frac{Cov(X_{t+k}, X_t)}{\sqrt{Var(X_{t+k})Var(X_t)}}$$

- This is a dimensionless measure for the linear association between the two random variables. Since for stationary series, we require the moments to be non-changing over time, we can drop the index t and write the autocorrelation as a function of the lag k:

$$\rho(k) = Cor(X_{t+k}, X_t)$$

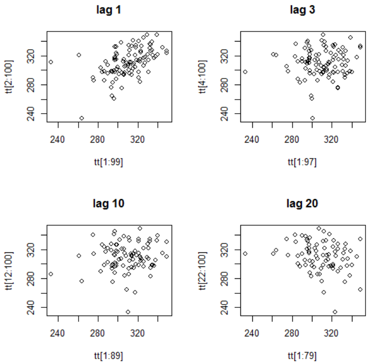
Autocorrelation. Lagged scatter plot.

In the example below the association seems linear and positive. The correlation coefficient is $0,51$, thus the relation is moderately strong.



Autocorrelation. Lagged scatter plot.

Because there are less and less data pairs at higher lags, the respective estimated correlations are less precise (one can prove that the variance of ρ_k increases with k).



Plug-in estimation of the autocorrelation function.

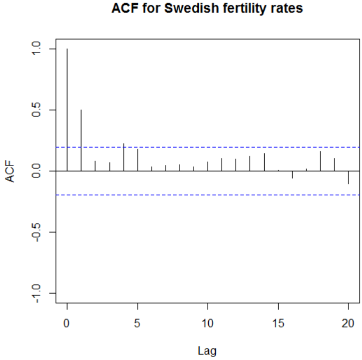
For mitigating the above problem with the lagged scatter plot, autocorrelation estimation is commonly done using the so-called plug-in approach (sample moments instead of population moments). That is, using estimated autocovariances:

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}, \quad \textit{for} \quad k = 1, \ldots, n-1,$$

where $\hat{\gamma}(k) = \frac{1}{n}\sum_{s=1}^{n-k}(x_{s+k} - \bar{x})(x_s - \bar{x})$ with $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$.

Correlogram.

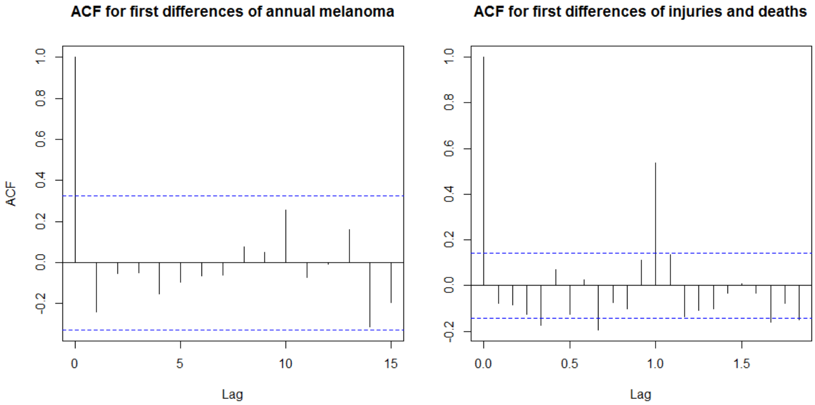
Now, we know how to estimate the autocorrelation function (ACF) for any lag k and we can introduce the correlogram, the standard means of visualization for the ACF.



Confidence bands.

- It is obvious that even for an iid series without any serial correlation, and thus $\rho(k) = 0$ for all k, the estimated autocorrelations $\hat{\rho}(k)$ will generally not be zero.
- The confidence bands are obtained from an asymptotic result: for long iid time series it can be shown that the $\hat{\rho}(k)$ approximately follow a $N(0, 1/n)$ distribution. Thus, each $\rho(k)$ lies within the interval of $\pm 1{,}96/\sqrt{n}$ with a probability of approximately 95 %.
- Statement: for any stationary series, sample autocorrelation coefficients $\hat{\rho}(k)$ that fall within the confidence band $\pm 1{,}96/\sqrt{n}$ are considered to be different from 0 only by chance, while those outside the confidence band are considered to be truly different from 0.

Correlogram and confidence bands.



Correlogram and Ljung-Box test.

- The Ljung-Box approach test the null hypothesis that a number of autocorrelation coefficients are simultaneously equal to zero. The test statistic is :

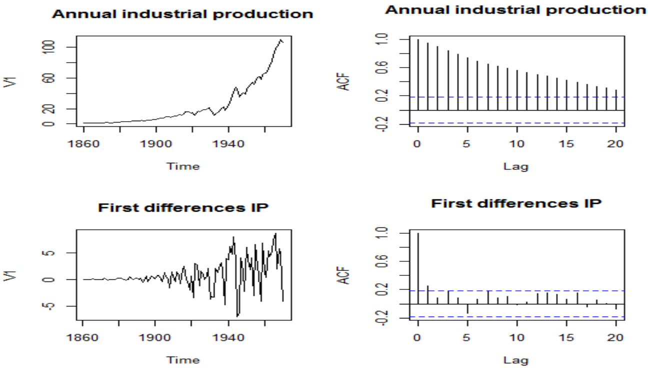
$$Q(h) = n \cdot (n+2) \cdot \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$$

- For first differences of melanoma series: $Q(24) = 16{,}687$ with critical value $\chi^2_{24}(0{,}05) = 36{,}41$, and therefore the p-value$= 0{,}8617$. No significant correlation for the first 24 lags.
- For first differences of injuries series: $Q(24) = 152{,}91$ with critical value $\chi^2_{24}(0{,}05) = 36{,}41$, and therefore the p-value$= 2{,}2e^{-16}$. We reject the null hypothesis and there is significant correlation for the first 24 lags.

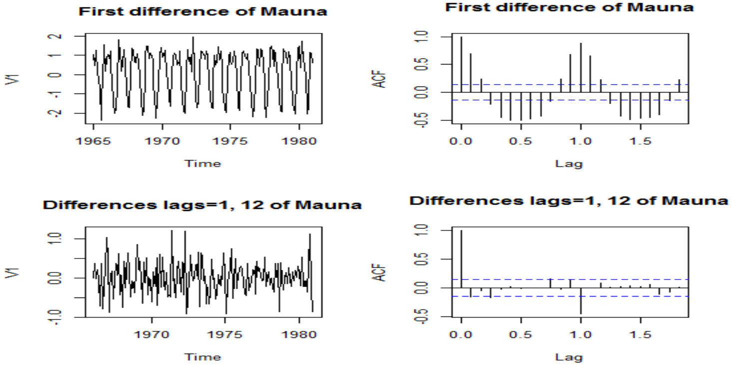
Correlogram of non-stationary series.

- Estimation of the ACF from an observed time series assumes that the underlying process is stationary. Only then we can treat pairs of observations at lag k as being probabilistically equal and compute sample covariance coefficients.
- If we apply the same formulae to non-stationary series, the ACF usually exhibits slow decay. This can serve as a second check for non-stationarity.

Correlogram of non-stationary series.

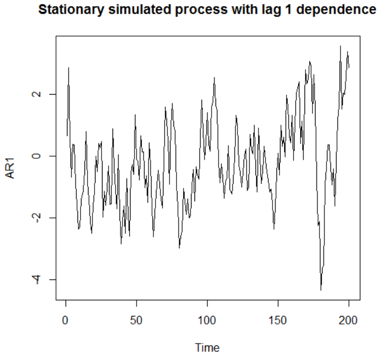


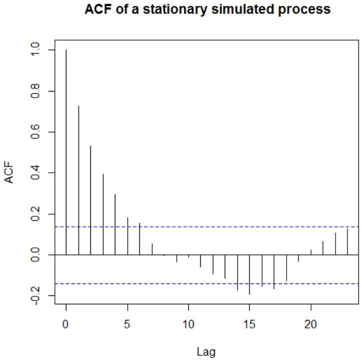
Correlogram of non-stationary series.



Partial autocorrelation and partial correlogram

Let us consider the simulated process$x_t = 0{,}75 * x_{t-1} + a_t$ where $a_t \sim N(0, 1)$. Below the plot of the simulated series and its correlogram.





Partial autocorrelation and partial correlogram

- For a process with a strong positive correlation at lag 1, the autocorrelations for lags 2 and higher will be positive as well just by propagation.
- Partial autocorrelation $\pi(k)$ measures the association between X_t and X_{t+k} with the linear dependence of X_{t+1} through X_{t+k-1} removed.

$$\pi(k) = Cor(X_{t+k}, X_t \mid X_{t+1} = x_{t+1}, \ldots, X_{t+k-1} = x_{t+k-1})$$

- $\pi(k)$ is defined for $k =\geq 1$. Remember that $\rho(k)$ is defined for $k \geq 0$ and we had that $\rho(0) = 1$).
- There is a (theoretical) relation between the partial autocorrelations $\pi(k)$ and the plain autocorrelations $\rho(1), \ldots, \rho(k)$:

$$\pi(1) = \rho(1); \quad and \quad \pi(2) = (\rho(2) - \rho(1))^2/(1 - \rho(1)^2)$$

The formula for higher lags exist, but get complicated rather quickly.

Partial autocorrelation and partial correlogram

Let us consider the simulated process$x_t = 0{,}75 * x_{t-1} + a_t$ where $a_t \sim N(0,1)$. Below the correlogram and partial correlogram.