Time Series Analysis
DSC534

Dr Yiolanda Englezou

Fall semester 2025-2026

Notes based on notes of Prof. Konstantinos Fokianos

Lecture 1

Time Series

Time series data are simply measurements observed sequentially in time. Some standard examples include

<u>Business</u>

Sales figures, production number, number of customers

<u>Economics</u>

Stock prices, exchange rates, interest rates

<u>Natural Sciences</u>

population size, precipitation, temperature ...

Time series data are serially correlated, that is there exist relations among observations. This is in contrast with i.i.d data.

Example: $X_1, \ldots, X_n$ random variables such that

$E(X_i) = \mu$, $\quad Var(X_i) = \sigma^2$, $\quad Cov(X_i, X_j) = \rho \neq 0$

We consider $\bar{X}$ as an estimator of μ.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Then $\quad E(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} \underbrace{E(X_i)}_{\mu} = \frac{1}{n} \cdot n\mu = \mu \qquad$ (unbiased)

$$Var(\bar{X}) = Var\left\{ \frac{1}{n} \sum_{i=1}^{n} X_i \right\} = \frac{1}{n^2} \sum_{i=1}^{n} \overbrace{Var(X_i)}^{\sigma^2} + \frac{1}{n^2} \underbrace{\sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} \overbrace{Cov(X_i, X_j)}^{\rho}}_{\text{sum has } n^2 - n = n(n-1) \text{ terms}}$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 + \frac{1}{n^2} n(n-1) \rho$$

$$= \frac{\sigma^2}{n} + \left(1 - \frac{1}{n}\right) \rho$$

From the above relation, if $\rho \neq 0$ then as $n \to \infty$ $\bar{X}$ is not consistent for μ (i.e. $\bar{X}$ is not close to μ with high probability as $n \to \infty$).

If we want to use the standard C.I. for μ

$$\bar{X} \pm t_{n-1;\, \alpha/2} \sqrt{\widehat{Var}(\bar{X})}$$

then we need to take into account estimation of ρ.

Goals of Time Series analysis

- Understand dependencies of data
- Visualisation/summary statistics
- Build suitable models to predict future data

Data Example: Quarterly earnings of J&J

Quarterly Earnings of J & J

Quarterly earnings per share for the U.S. company Johnson & Johnson. There are 84 quarters (21 years) measured from the first quarter of 1960 to the last quarter of 1980. One of the most important steps in time series analysis is to visualize the data, i.e. create a time series plot, where the quarterly earnings are ploted against time.

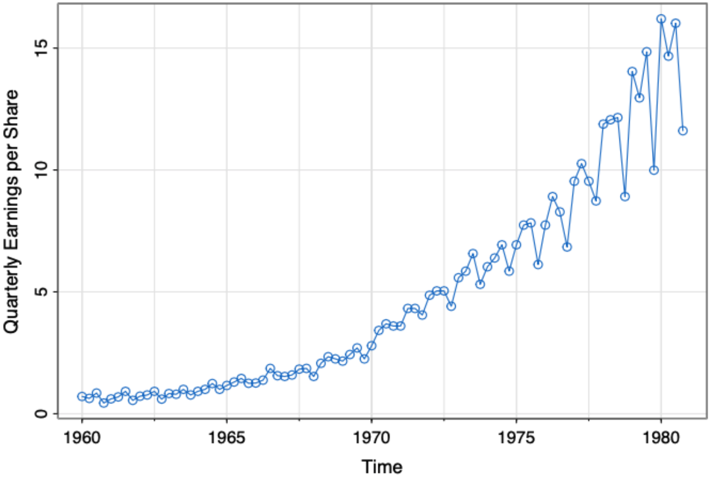
Main observations from the plot:

1. Gradually increasing trend. In general, a systematic change in the mean level of a time series that does not appear to be periodic is known as a trend. The simplest model for a trend is a linear increase or decrease.
2. Regular variation superimposed on the trend that appears to be repeated over quarters. This is known as a seasonal effect, or seasonality.

```r
library(astsa)
```

```
## Warning: package 'astsa' was built under R version 4.3.3
```

```r
tsplot(jj, col=4, type="o", ylab="Quarterly Earnings per Share")
```



Quarterly Earnings per Share
0
5
10
15
1960
1965
1970
1975
1980
Time

Dominant features of time series: trend + seasonality (periodicity)

They can modelled by a deterministic model or estimated by non-parametric methodology. In addition, we need to take into account the correlation structure of the observations.

This class
1. Focus how to describe and visualize time series data
2. Fit models
3. Generate forecasts
4. Draw conclusions based on the output.

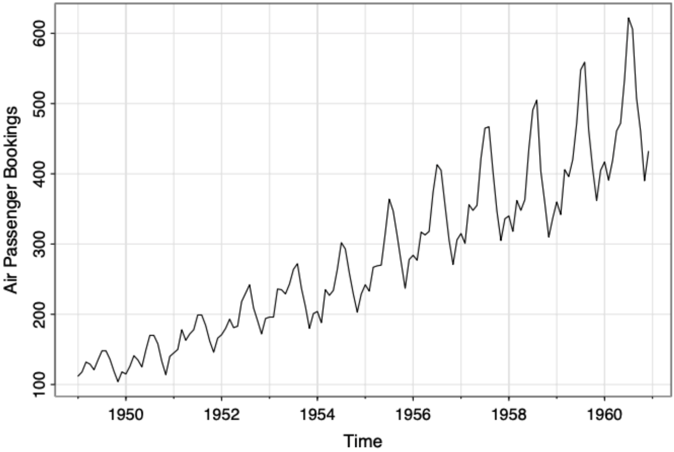
Data Example: Air Passenger bookings

Air Passenger Bookings

The numbers of international passenger bookings (in thousands) per month on an airline in the United States were obtained from the Federal Aviation Administration for the period 1949-1960. The company used the data to predict future demand before ordering new aircraft and training aircrew. The data are available as a time series in R. These data share similar characteristics with the previous example.

```
data(AirPassengers)
AirPassengers
```

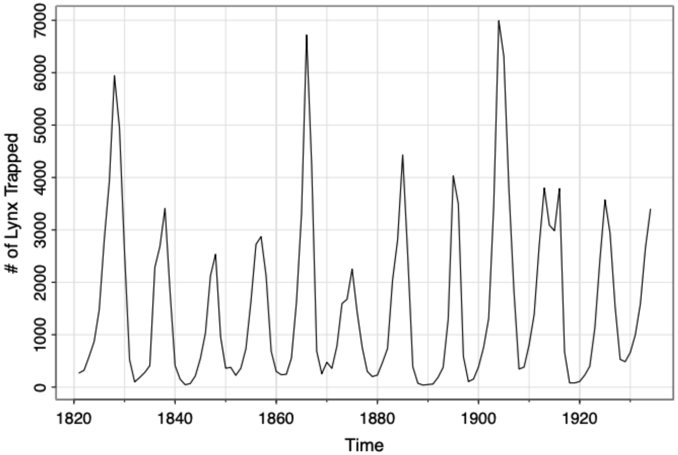
```
tsplot(AirPassengers, ylab="Air Passenger Bookings")
```



Data Example: Lynx Trappings

Annual number of lynx trappings for the years 1821-1934 in the Mackenzie River District in Canada.

```
data(lynx)
tsplot(lynx, ylab="# of Lynx Trapped")
```

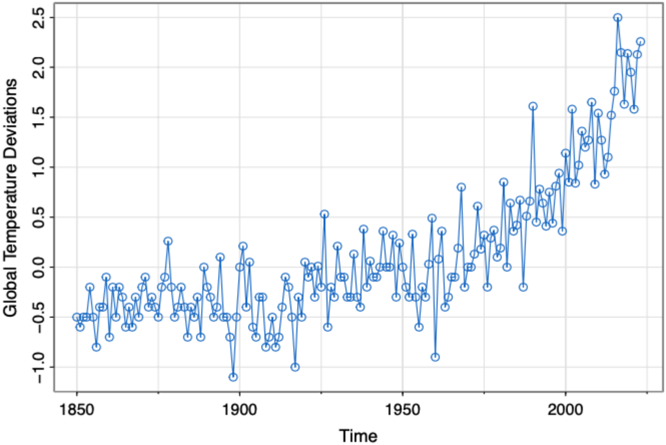


Observe that number of trapped lynx reaches high and low values every about 10 years, and some even larger figure every about 40 years. This suggests that the prominent periodicity is to be interpreted as random, but not deterministic. Understanding and modeling trend and seasonal variation is a very important aspect, much of the time series methodology is aimed at stationary series, i.e. data which do not show deterministic, but only random (cyclic) variation.

Data Example: Global Temperature

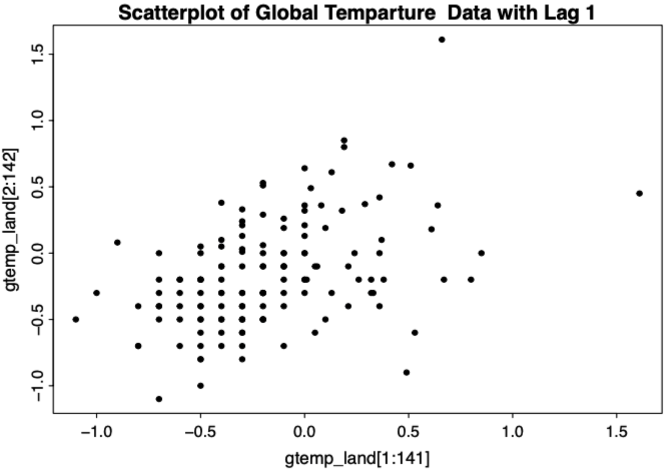
These data correspond to global mean land–ocean temperature index from 1880 to 2015, with the base period 1951–1980. In particular, the data are deviations, measured in degrees centigrade, from the 1951–1980 average. Observe an apparent upward trend in the series during the latter part of the twentieth century that has been used as an argument for the global warming hypothesis. Note also the leveling off at about 1935 and then another rather sharp upward trend at about 1970. The question of interest for global warming proponents and opponents is whether the overall trend is natural or whether it is caused by some human-induced interface. The question of trend is of more interest than particular periodicities. Instead, we address the the simpler question of analyzing the correlation of subsequent records, called auto correlations. The autocorrelation for lag 1 can be easily examined by producing a scatter plot of adjacent observations:

```
tsplot(gtemp_land, col=4, type="o", ylab="Global Temperature Deviations")
```



Data Example: Global Temperature

```r
plot(gtemp_land[1:141], gtemp_land[2:142], pch=20)
title("Scatterplot of Global Temparture  Data with Lag 1")
```



Data Example: Global Temperature

```r
cor(gtemp_land[1:141], gtemp_land[2:142])
```

```
## [1] 0.441967
```

Hence we observe positive correlation between consecutive meausrements. So if the previous observation is below/above the mean, then the next is more likely to have the same direction.

Log-returns

Suppose that X_t is the actual value of the index

$$r_t = \frac{X_t - X_{t-1}}{X_{t-1}} = \text{return}$$

Hence $1 + r_t = \frac{X_t}{X_{t-1}} \Rightarrow \log(1 + r_t) = \log\left(X_t / X_{t-1}\right) \rightarrow \text{diff}(\log)$

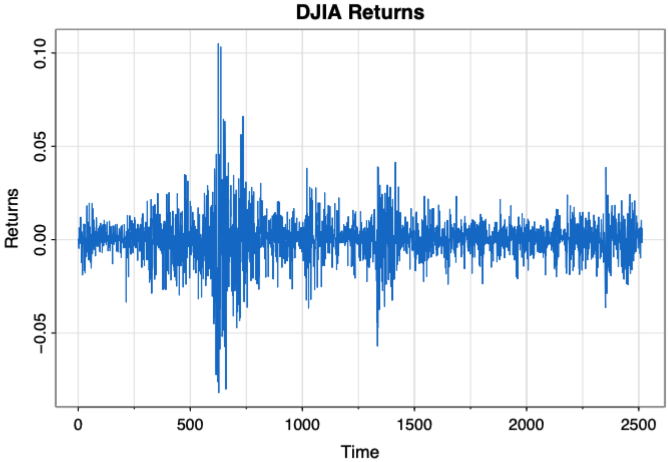
$$= \log X_t - \log X_{t-1} \simeq r_t$$

Data Example: Financial Returns

Financial returns

The figure below shows daily returns (or percent change) of the Dow Jones Industrial Average (DJIA) from April 20, 2006 to April 20, 2016. Observe the financial crisis of 2008 in the figure. In addition, observe that the mean of the series appears to be stable with an average return of approximately zero, however, highly volatile (variable) periods tend to be clustered together.

```
library(xts)          # install it if you don't have it
djiar = diff(log(djia$Close))[-1]
tsplot(as.vector(djiar$Close), xlab="Time", ylab="Returns",col=4, main="DJIA Returns")
```

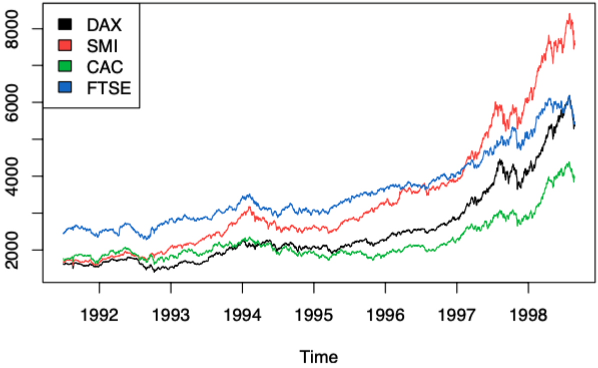


Multivariate Time Series

More than one time series (multivariate)

The data EuStockMarkets contains the daily closing prices of major European stock indices: Germany DAX (Ibis), Switzerland SMI, France CAC, and UK FTSE. The data are sampled in business time, i.e., weekends and holidays are omitted.

```
data(EuStockMarkets)
ts.plot(EuStockMarkets, col=1:4)
leg.txt <- c("DAX", "SMI", "CAC", "FTSE")
legend("topleft", leg.txt, fill=1:4)
```



Summary of the class

1.) Exploratory Data Analysis

2.) Modelling

3.) Forecasting (extrapolation)

4) Time Series Regression

5) Process Control

6) Clustering /Classification.