DMIF, University of Udine

# Time Series

Andrea Brunello
andrea.brunello@uniud.it

May 2020

# Outline

# Introduction

# What is a time series?

A time series is a set of observations $x_t$, each one collected at a specific, usually ordered, time $t$.

A *discrete time series* is one in which the set $T$ of times at which observations are collected is discrete (e.g., sampling rate).

A *continuous time series* is obtained when observations are recorded continuously over some time interval.

# Time series everywhere

Time series are the result of repeatedly measuring observable quantities like:

- heart rate
- stock value
- engine throttle per minute
- monthly rainfall in Rome
- daily vehicle transit on a highway

Time series are involved in scientific, business and medical domains.

# Time series means big data

Usually they are very large datasets:

- 1 Hour of Electrocardiogram (ECG) data is 1 Gigabyte
- New York City Taxi dataset is 260 GB

Moreover, they may have to be collected at a very high peace.

# Why studying time series?

There are several reason pushing the study of time series:

- description of the salient features of the series (decriptive analysis)
- prediction of the future based on the past (predictive analysis)
- control of the process producing the series (e.g., finding out anomalous behaviour: maintenance/predictive maintenance)

# Time series general model

Let $s_1, \ldots, s_t$ be a discrete univariate time series describing observations on some variable made at $T$ equally spaced time points labelled with $1, \ldots, t$

A general model for the above time series can be:

$$s_i = g(i) + \varphi_i \qquad i = 1, \ldots, t$$

$g(i)$ is the *systematic part*, also called signal, which is a deterministic function of time

$\varphi_i$ is a *stochastic sequence*, or residual term or noise, which follows a probability law (often Gaussian)

# Internal structures of time series

Time series can exhibit some special characteristics:

- *Trend*: it shows the general tendency of the data to increase or decrease during a long period of time
- *Seasonality*: periodic, repetitive, and generally regular and predictable variations that occur in a time series at specific regular intervals less than a year, such as weekly, monthly, or quarterly
- *Cyclical variations*: oscillatory fluctuations in the time series with unknown duration, but typically of more than one year (e.g., stock prices)
- *Unexpected movements*: for instance, noise or anomalies

# Trend – 1

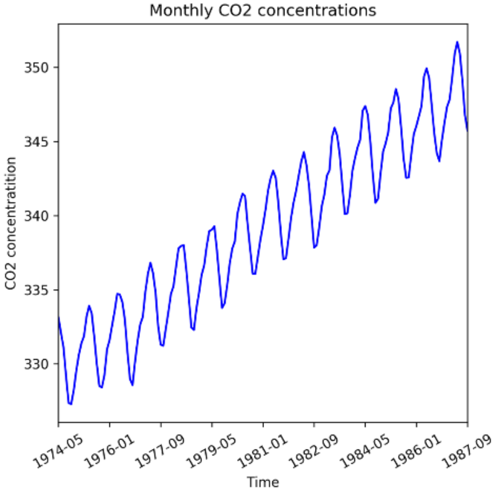



Figure 1.5: Time series of CO2 readings with an upward trend

# Trend – 2

Zooming in makes the general trend impossible to see:

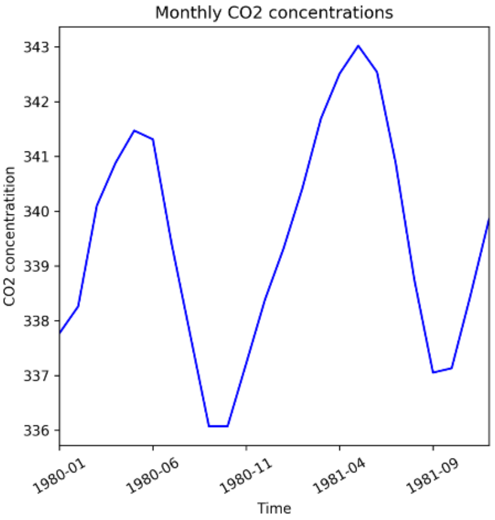
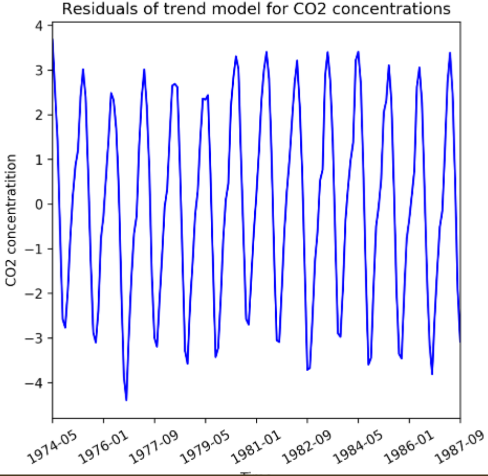



Figure 1.6: Shorter run of CO2 readings time series which is not able to reveal general trend

It is possible to compute trend line as a prediction (regression). Residuals are then useful for determining the presence of noise, seasonality effects or other movements:



# Time series decomposition

The objective of time series decomposition is to model the long-term trend and seasonality and estimate the overall time series as a combination of them. Two popular models for time series decomposition are:

- Additive model
- Multiplicative model

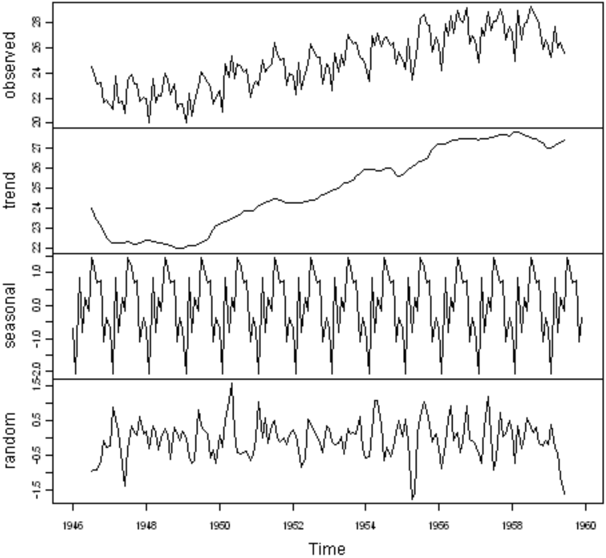
# Decomposition: additive model

The additive model formulates the original time series $(x_t)$ as the sum of the trend $(F_t)$, seasonal $(S_t)$ and random $(\epsilon_t)$ components as follows:

$$x_t = F_t + S_t + \epsilon_t$$

The additive model is applied when there is a time-dependent trend component, but independent seasonality that does not change over time.

# Additive model – Example



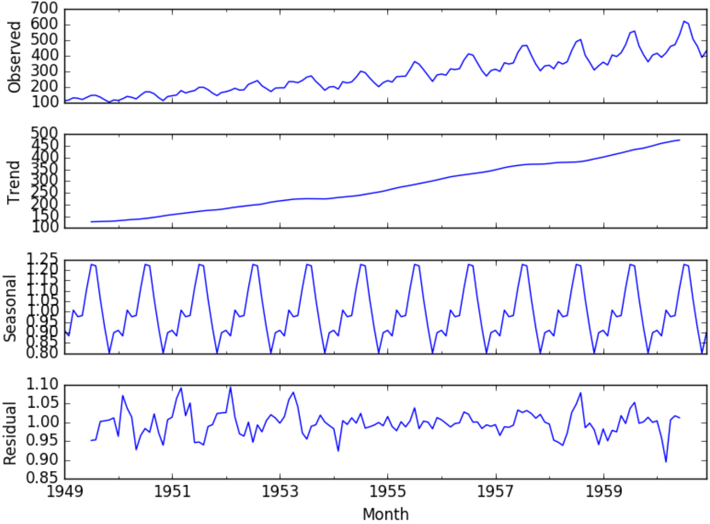
# Decomposition: multiplicative model

The multiplicative decomposition model, which gives the time series as product of the trend, seasonal, and random components is useful when there is time-varying seasonality, that depends on the trend/level of the time series:

$$x_t = F_t * S_t * \epsilon_t$$

It can be converted to an additive model of the logarithms of the individual components:

$$log(x_t) = log(F_t) + log(S_t) + log(\epsilon_t)$$



Observed
700
600
500
400
300
200
100
Trend
500
450
400
350
300
250
200
150
100
Seasonal
1.25
1.20
1.15
1.10
1.05
1.00
0.95
0.90
0.85
0.80
Residual
1.10
1.05
1.00
0.95
0.90
0.85
1949
1951
1953
1955
1957
1959
Month

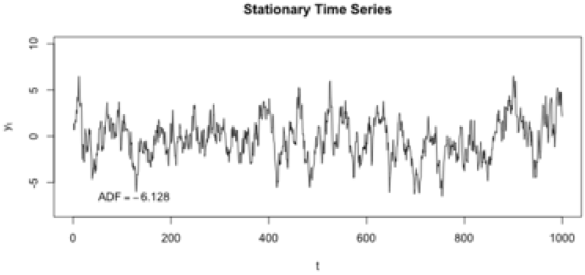
# Stationary time series

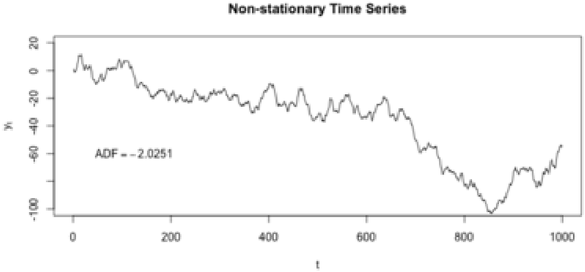
A stationary time series is one whose properties do not depend on the time at which the series is observed.

- Thus, time series with changing levels, trends, seasonality, or increasing/decreasing variance are not stationary
- On the other hand, a white noise series is stationary
- Also a time series with cyclic variations is stationary, since cycles are aperiodic and of no fixed length (unpredictable)

Stationarity is important, since most statistical forecasting methods (e.g., ARIMA) are based on the assumption that the time series can be rendered approximately stationary through the use of mathematical transformations such as *detrending* and *differencing*.

# Stationary and non-stationary time series



# Keogh Eamonn's Time Series Mining Tutorial

# Time Series Modeling and Forecasting Exercises