DMIF, University of Udine

Time Series

Andrea Brunello
andrea.brunello@uniud.it

May 2022

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, time series involve very large datasets.

For example, 1 hour of electrocardiogram (ECG) recordings amount for around 1 Gigabyte of data.

Moreover, time series data may have to be collected at a very high pace.

Why studying time series?

There are several reasons pushing the study of time series:

- prediction of the future based on the past
- control of the process generating the series
- discovery of anomalous behaviours
- description of the salient features of the series

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

φ_i is a *stochastic sequence*, or residual term or noise, which follows a probability law

Internal structure of a 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 (e.g., sweater sales are likely to have an increase during winter)
- *Cyclical variations*: oscillatory fluctuations in the time series with an unknown and non-predictable duration, but typically of more than one year (e.g., business cycles)
- *Unexpected movements*: for instance, noise or anomalies

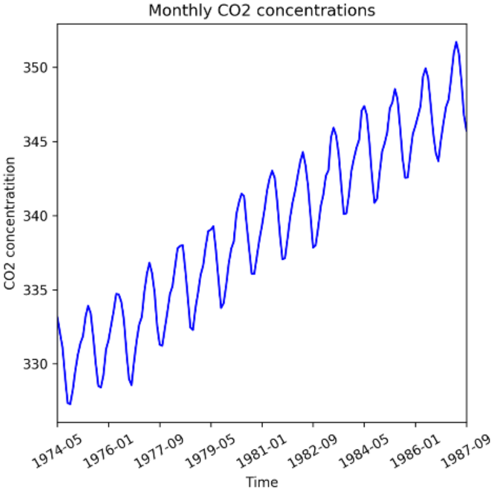


Figure 1.5: Time series of CO2 readings with an upward trend

Zooming in makes the general trend impossible to see:

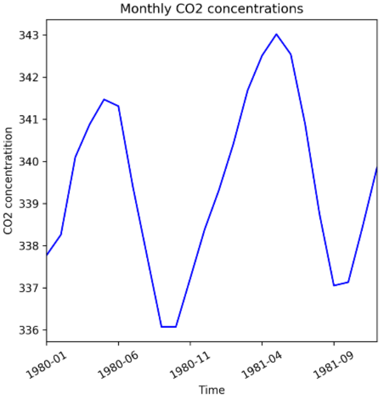


Figure 1.6: Shorter run of CO2 readings time series which is not able to reveal general trend

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

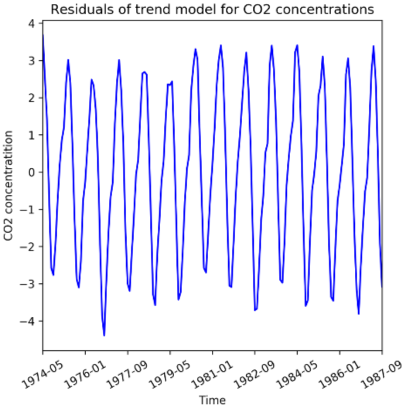


Figure 1.7: Residuals from a linear model of the general trend in CO2 readings

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:

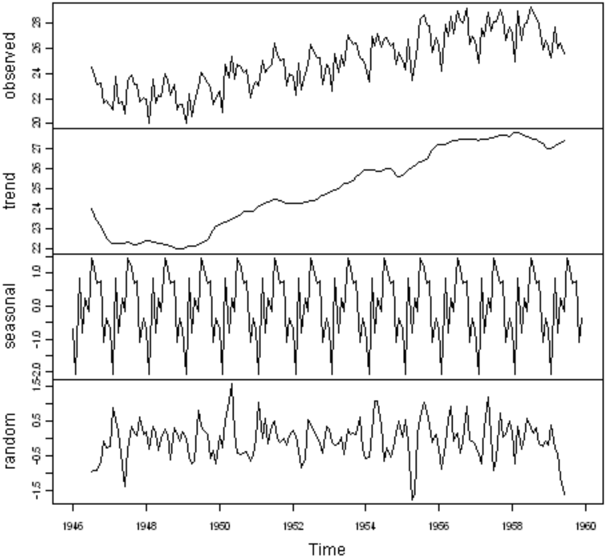
- Additive model
- Multiplicative model

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 (ϵ_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.


observed
trend
seasonal
random
1946
1948
1950
1952
1954
1956
1958
1960
Time

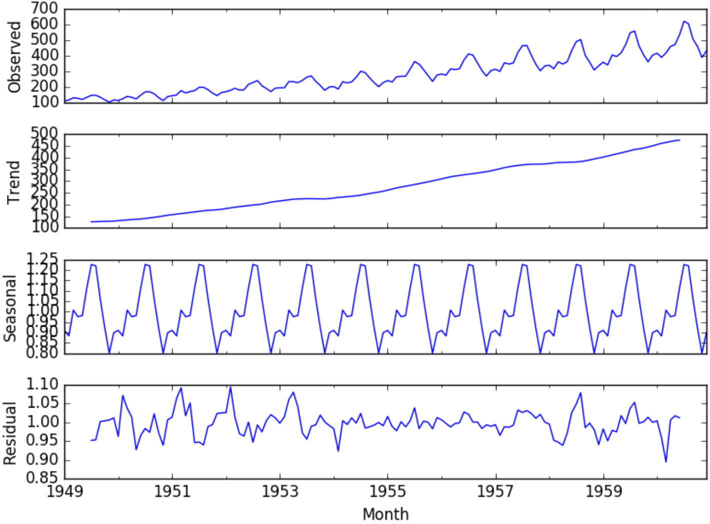
Decomposition: multiplicative model

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:

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



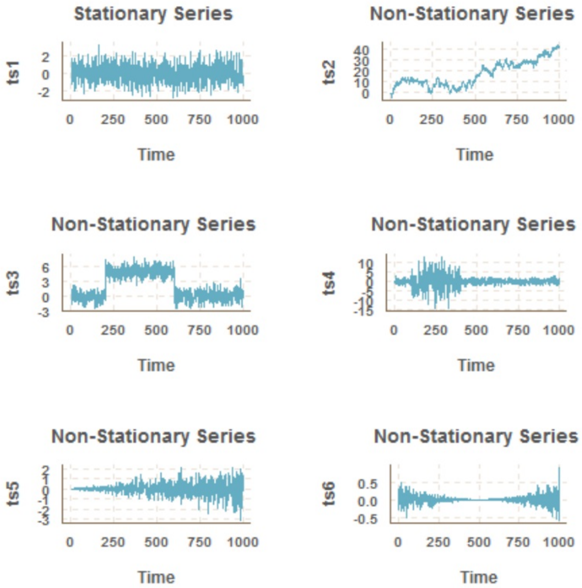
Observed
700
600
500
400
300
200
100
Trend
500
450
400
350
300
250
200
150
100
Seasonal
1.25
1.20
1.15
1.10
1.05
1.00
0.95
0.90
0.85
0.80
Residual
1.10
1.05
1.00
0.95
0.90
0.85
1949
1951
1953
1955
1957
1959
Month

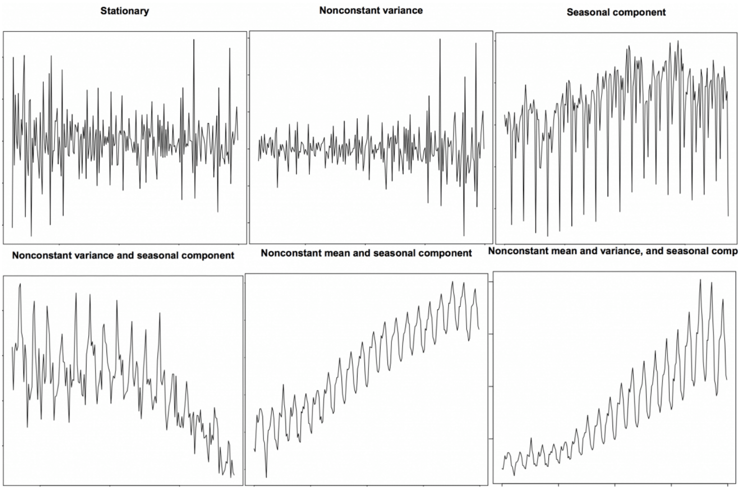
Stationary time series

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

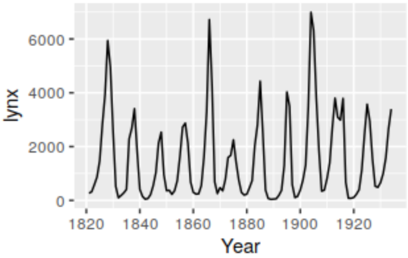
- Thus, time series with changing levels, trends, seasonality, or increasing/decreasing variance are not stationary
- On the other hand, a white noise series is stationary
- Also a time series with cyclic variations is stationary, since cycles are aperiodic and of no fixed length (unpredictable)

Stationary and non-stationary time series




Stationary
Nonconstant variance
Seasonal component
Nonconstant variance and seasonal component
Nonconstant mean and seasonal component
Nonconstant mean and variance, and seasonal comp

Actually, stationarity depends also on the underlying process that generates the time series.



At first glance, the strong cycles might appear to make the depicted time series non-stationary. However, they are actually aperiodic, since they represent the variations in an observed lynx population which, in the long-term, have no predictable timing. Hence the series can be considered to be stationary.

The importance of stationarity

The **ARMA** (AutoRegressive Moving Average) model is a tool for understanding and, perhaps, predicting future values in a time series.

It turns out that any stationary data can be approximated with the ARMA model, thanks to the Wold decomposition theorem.

That is why ARMA models are very popular and that is why we need to make sure that the series is stationary to use these models.

Some other models, e.g., **ARIMA** (AutoRegressive Integrated Moving Average) are based on the assumption that the time series can be rendered approximately stationary through the use of *differencing*.

The importance of stationarity (2)

Intuitively, non-stationary time series have undefined means and infinite variances.

Thus, they cannot be described by relying on mean, variance, and autocorrelation.

In addition, approaches like ARMA are essentially linear regression models that utilize the lag(s) of the series itself as predictors, and we know that linear regression works best if the predictors are not correlated against each other.

Stationarizing the series solves this problem since it removes any persistent autocorrelation, thereby making the predictors (lags of the series) in the forecasting models nearly independent.

Stationarity: last words

Despite its importance, stationarity is not required for *every* time series-related task. For instance, you do not need to stationarize time series to perform clustering.

Observe that the importance of stationarity holds (up to a certain extent) even if more sophisticated machine learning techniques are used for the prediction, such as recurrent neural networks (e.g., LSTM):

- The thing is that, in your training set, you are still relying, say, "on the past N values" of the series to predict the next, $N+1$ value
- Thus, when you apply the learned model on a new series of N values to perform the prediction, you are implicitly assuming the same relationship to hold between the new N predictor points and the following $N+1$ point you are trying to estimate

Keogh Eamonn's Time Series Mining Tutorial

Notebook: Time Series Modeling and Forecasting