Time Series

Lluís Talavera

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
ETSEIB

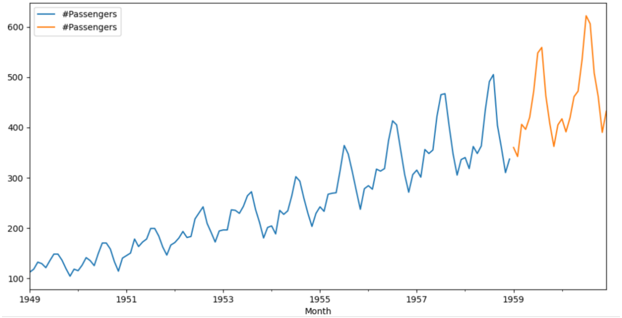
What is a time series?

It is a chronological sequence of observations of a particular variable collected at regular intervals.

The time column does not represent a variable per se: it defines a structure to order the dataset.

The goal is to build a model that forecasts future values:

- What are the expected sales volumes in a grocery store next quarter?
- What are the resale values of vehicles after leasing them out for three years?
- What are the passenger numbers for an international airline?
- What is the future electricity load in an energy supply chain?



Time Series decomposition

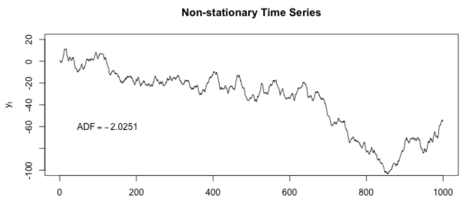
Time series data can be decomposed into three components:

- Seasonality: a periodic recurring movement. For example, temperatures in summer or winter.
- Trend: a long-term upward or downward movement.
- Noise: variability that can be explained neither by seasonality or by a trend.

Stationarity

A time series is stationary if it has no trend, i.e., basic properties of the data distribution (mean, variance) remain constant over time.

Stationary time series are easier to analyze and a requirement of some models.



We can test if a time series is stationary using the Augmented Dickey-Fuller (ADF) test.

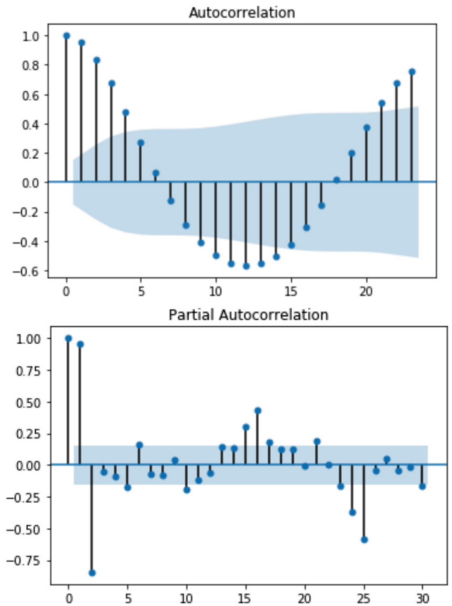
```
ADF test statistic: -2.1641431278047816
ADF p-values: 0.2195157763715047
ADF number of lags used: 12
ADF number of observations: 106
ADF critical values: {'1%': -3.4936021509366793, '5%': -2.8892174239808703, '10%': -2.58153320754717}
```

Autocorrelation

Is the correlation between a current value and past values. Can be positive or negative.

Can be detected with two graphs:

- ACF (AutoCorrelation Function)
- PACF (Partial AutoCorrelation Function)



Statistical modeling: ARIMA

An extension of the ARMA model that merges two components:

- **AR** (Auto Regressive): attempts to predict future values based on past values. AR models require the time series to be stationary.
- **MA** (Moving Average): attempts to predict future values based on past forecasting errors.

ARIMA adds an integration component:

- **I** (Integrated): if not stationary, the time series can be differenced to become stationary, i.e., compute the differences between consecutive observations.

We need to find three parameters: p is the order of the AR component, d is the number of differences and q is the order of the MA component

The most common method for manually identifying the proper orders is using ACF and PACF.

We can also use the `auto_arima` function from `pmdarima` library.

ML modeling: Lag features

We convert the time series data into a tabular classical machine learning problem by shifting, so that the model tries to predict the value at the next time (t+1) given the value at the previous time (t-1). Then, a regular regression model can be applied.

Further feature engineering can be performed, such as summary statistics or nonlinear windows (e.g. a value from last month in a daily time series).

A problem is that we need to choose the width of the sliding window.

An advantage of this method is that it can incorporate external features to the model.