Time Series Autoregression

Forecasting Time Series

"Prediction is very difficult, especially if it's about the future."

- Nils Bohr, Nobel laureate in Physics

Stationarity

Recall: in time series the data follow a chronological ordering. The previous models we studied did not consider an ordering over data - we could randomly "*reshuffle*" the data and everything would be ok!

Stationarity is an important property of time series data that allows us to "*reshuffle*" it

In simple terms, a stationary time series has properties (e.g., mean or variance) that do not depend on time. In particular: it doesn't have an obvious trend or seasonality.

CAMBRIDGE SPARK

Stationarity

In simple terms, a stationary time series has properties that do not depend on time. In particular: it doesn’t have an obvious trend or seasonality.

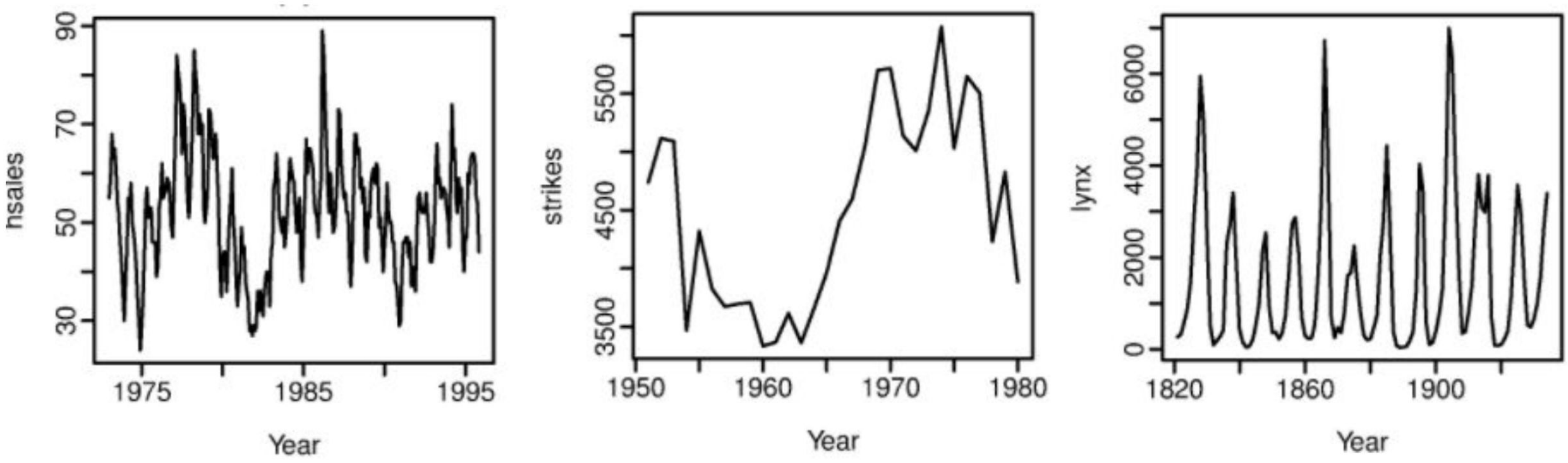
Often we want our time series model to look like the following:

Observations = seasonal comp. + trend comp. + stationary noise

General idea : remove the "non-stationary" components using various methods: rolling average, differencing, etc.

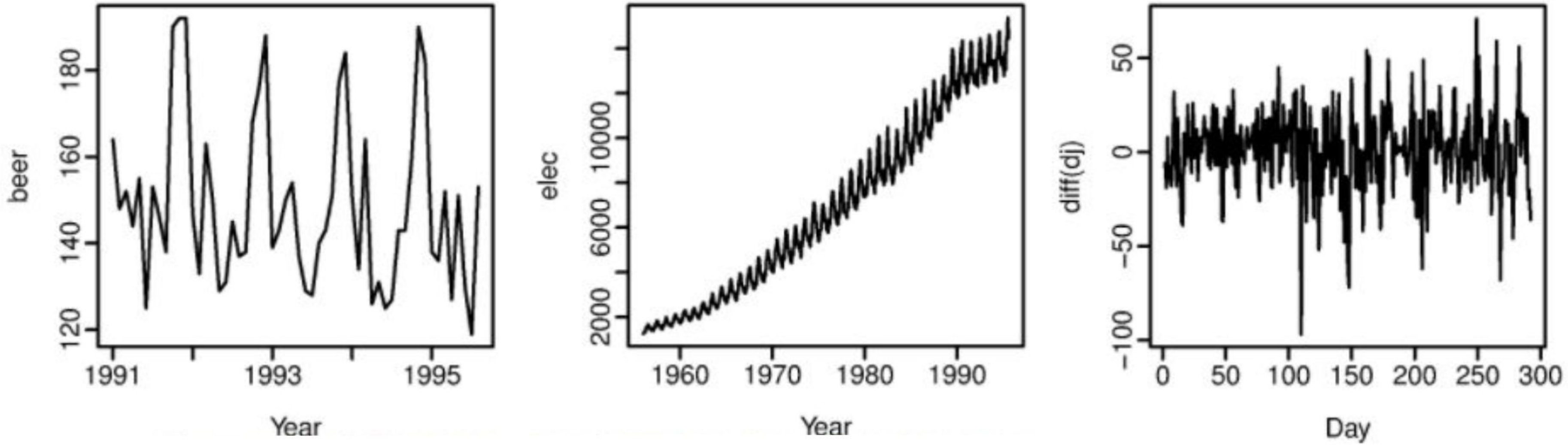
CAMBRIDGE SPARK

Stationary or not?



CAMBRIDGE SPARK

Stationary or not?



Time series modelling, the basics

Simple modelling steps:

1. Do differencing with ideal lag (based on autocorrelation),
2. If the result looks stationary (no time-dependent pattern), apply a time series model
3. If does not look stationary? try differencing again.

Autoregression model (AR)

- Closely related to differencing!
- An autoregression model makes the assumption that past observations are useful to predict the value in the future
- Previous time steps become input to a regression equation to predict the value at the next time step

Autoregression model (AR)

For a stationary time series, model y_t as a linear combination of p past values:

$$y_t = w_0 + w_1 y_{t-1} + \cdots + w_p y_{t-p} + \text{noise}_t$$

Autoregression model (AR)

For a stationary time series, model y_t as a linear combination of p past values:

$$y_t = w_0 + w_1 y_{t-1} + \cdots + w_p y_{t-p} + \text{noise}_t$$

This is just a linear regression where each column of the feature matrix is the lagged time-series with an increasing lag.

$$X = \begin{pmatrix} 1 & y_{t-1} & y_{t-2} & \cdots & y_{t-p} \\ 1 & y_{t-2} & y_{t-3} & \cdots & y_{t-p-1} \\ & & \vdots & & \end{pmatrix}$$

CAMBRIDGE SPARK

Autoregression model (AR)

For a stationary time series, model y_t as a linear combination of p past values:

$$y_t = w_0 + w_1 y_{t-1} + \cdots + w_p y_{t-p} + \text{noise}_t$$

Intuition: values at previous time points are informative when predicting the next value.

For example: sales in the past *p* days might be useful when predicting sales tomorrow.

This is exactly what an AR(*p*) model will fit!

CAMBRIDGE SPARK

Hands-on session

time_series_autoregression.ipynb