ARIMA Method

Time Series Forecasting (Part 3)

Lecture Video Slides

Autocorrelation

- Can lagged data points be used to predict the next data points?
- Lag k: k period behind
- Example for monthly data:
  - Latest data is Jan 2022, Lag 1: Dec 2021, Lag 2: Nov 2021, etc.
- Autocorrelation by lag k
  - Correlation between the vector X_t and vector X_{t-k}

Example of Autocorrelation of Sales by Lags 1, 2 & 3.

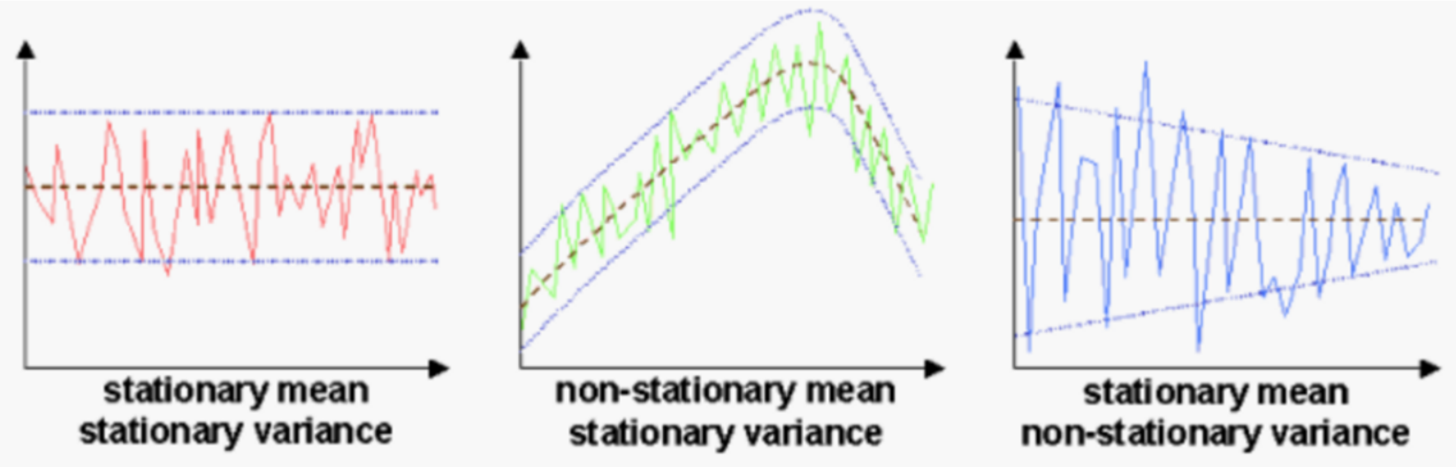
	A	B	C	D	E
1	Month	Sales	Lag 1	Lag 2	Lag 3
2	Jan-2013	226			
3	Feb-2013	254	226		
4	Mar-2013	204	254	226	
5	Apr-2013	193	204	254	226
6	May-2013	191	193	204	254
7	Jun-2013	166	191	193	204
8	Jul-2013	175	166	191	193
9	Aug-2013	217	175	166	191
10	Sep-2013	167	217	175	166
11	Oct-2013	192	167	217	175
12	Nov-2013	127	192	167	217
13	Dec-2013	148	127	192	167
14	Jan-2014	184	148	127	192
47	Oct-2016	185	175	181	179
48	Nov-2016	245	185	175	181
49	Dec-2016	177	245	185	175

Note that each successive lag just "pushes the variable down" by a row.

G	H	J
	Autocorrelation	
Lag1	0.357	=CORREL(B3:B49,C3:C49)
Lag2	0.084	=CORREL(B4:B49,D4:D49)
Lag3	0.089	=CORREL(B5:B49,E5:E49)

ARIMA(p, d, q)
Autoregressive Integrated Moving Average

- Arima requires the time series to be stationary.
- (Weak) Stationarity:
  - Mean is constant across time.
  - Every lag has a constant covariance i.e. $\forall u,v,a$, $\mathrm{cov}(x_u, x_v) = \mathrm{cov}(x_{u+a}, x_{v+a})$
    - Implies Variance is constant across time.
- d: num of differencing required for time series to become stationary.

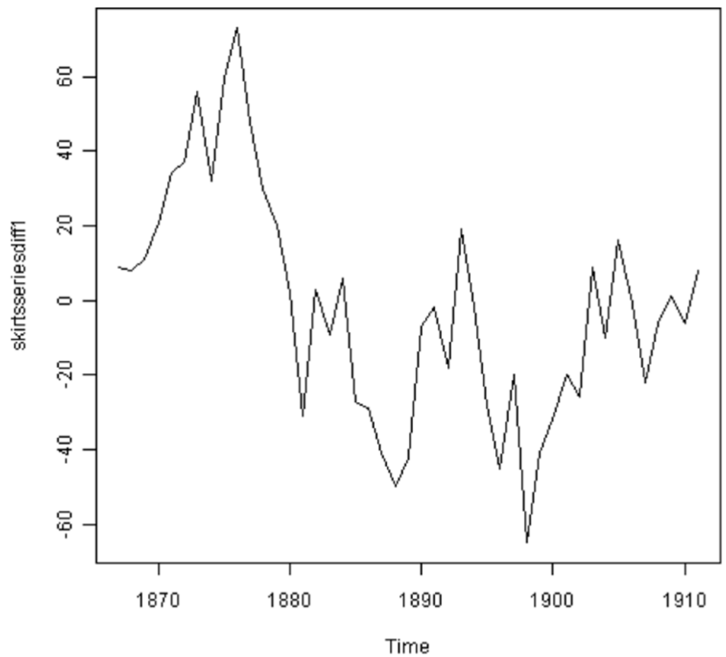


Source: https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322

ARIMA(p, d = 1, q)

Autoregressive Integrated Moving Average

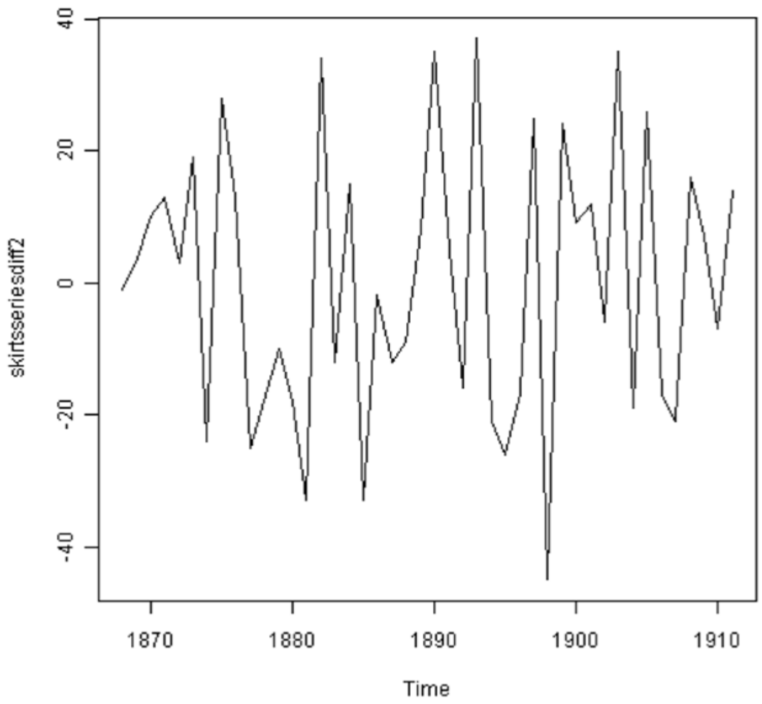
```
> skirtsseriesdiff1 <- diff(skirtsseries, differences=1)
> plot.ts(skirtsseriesdiff1)
```



Avril Coghlan (2018) pp. 49 – 50.

ARIMA(p, d = 2, q)
Autoregressive Integrated Moving Average

```
> skirtsseriesdiff2 <- diff(skirtsseries, differences=2)
> plot.ts(skirtsseriesdiff2)
```



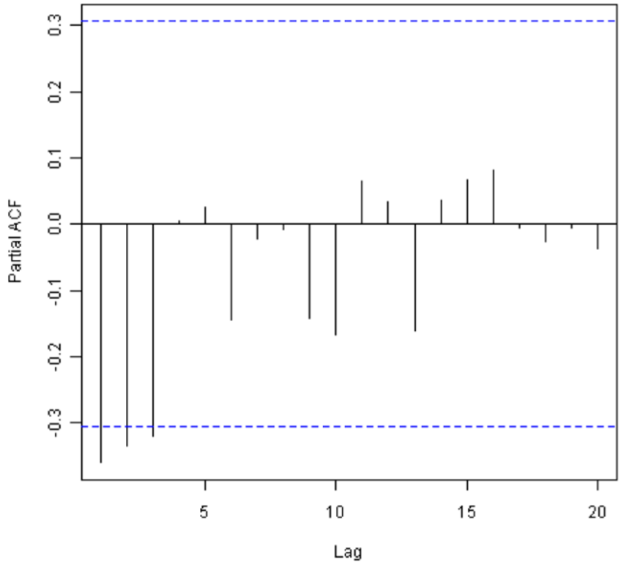
Avril Coghlan (2018) pp. 49 – 50.

ARIMA(p, d, q)
Autoregressive Integrated Moving Average

<u>**The autoregressive (AR) model**</u>: A time series modeled using an AR model is assumed to be generated as a linear function of its past values, plus a random noise/error:

$$x_t = c + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \varepsilon_t$$

Value of p estimated by viewing partial autocorrelation plot via pacf().
- See Avril Coghlan (2018) p 55.



Source: https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322

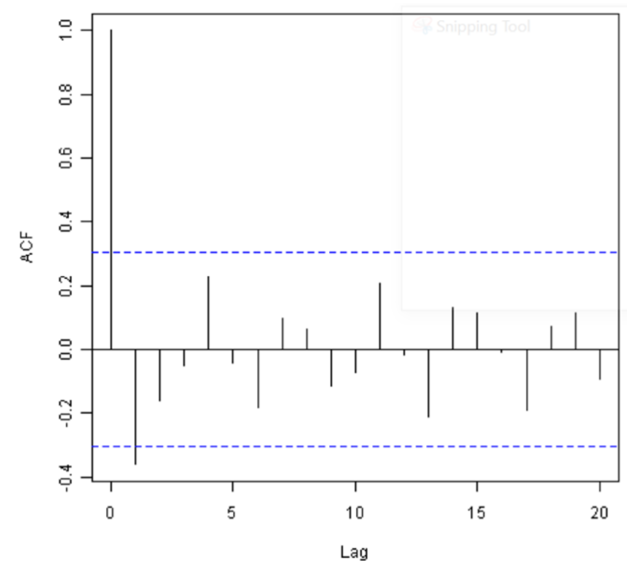
ARIMA(p, d, q)

Autoregressive Integrated Moving Average

<u>The moving average (MA) model</u>: A time series modeled using a moving average model, denoted with *MA(q)*, is assumed to be generated as a linear function of the last $q+1$ random shocks generated by ε_i, a univariate *white noise process*:

$$x_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

Value of q estimated by viewing autocorrelation plot via acf().
- See Avril Coghlan (2018) p 54.



ARIMA(p, d, q)

- First take the d differences to get stationary time series.
- Then, model the stationary time series using p previous data values and q previous random noise.

ARIMA(p, d, q)

Autoregressive Integrated Moving Average

- Instead of using eye power to view charts and then decide on p, d, q, shortcut in R:

Shortcut: the auto.arima() function

The auto.arima() function can be used to find the appropriate ARIMA model, eg., type "library(forecast)", then "auto.arima(kings)". The output says an appropriate model is ARIMA(0,1,1).

auto.arima() search for Seasonal terms too, if significant. i.e. SARIMA.

A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models we have seen so far. It is written as follows:

$$\text{ARIMA} \quad \underbrace{(p,d,q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P,D,Q)_m}_{\substack{\uparrow \\ \text{Seasonal part of} \\ \text{of the model}}}$$

where $m =$ number of observations per year. We use uppercase notation for the seasonal parts of the model, and lowercase notation for the non-seasonal parts of the model.

https://otexts.com/fpp2/seasonal-arima.html

ARIMA(p, d, q)

Autoregressive Integrated Moving Average

- Estimate the parameters of an ARIMA(p,d,q) model using the "arima()" function if you want to specify the order directly.

```
> kingstimeseriesarima <- arima(kingstimeseries, order=c(0,1,1)) # fit an ARIMA(0,1,
↪1) model
> kingstimeseriesarima
  ARIMA(0,1,1)
  Coefficients:
          ma1
        -0.7218
  s.e.   0.1208
  sigma^2 estimated as 230.4:  log likelihood = -170.06
  AIC = 344.13   AICc = 344.44   BIC = 347.56
```

Summary

- Forecasting is inherently extrapolation, unlike other models e.g. Linear Reg, CART, MARS, etc.
- MA
  - To estimate trend component
  - To decompose into trend, seasonality and random error components.
- Exponential Smoothing models
  - SES
  - Holt
  - Winters
- ARIMA
  - auto.arima() to get non-seasonal and seasonal p, d, q.
- Train-test split to follow time sequence
- Industry Practice
  - Model is just a baseline forecast
  - Adjust based on new data and new events.