# Dynamic Models for Prediction

## *Tutorial 3*

# Autoregressive models

- Recall the AR(p) model

white noise

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t$$

constant

The order of the model

- Different parameters will give different time series patterns
- The variance of the error term will only change the scale not the pattern

# Autoregressive models

- Normally autoregressive models are restricted to stationary data and some constraints on the values of the parameters are needed:
  - For AR(1) model $-1<\varphi_1<1$
  - For AR(2) model $-1<\varphi_2< 1$, $\varphi_1+\varphi_2 <1$ , $\varphi_1-\varphi_2 <1$
  - For models of order greater than 2 the restrictions are more complicated

# Moving average models

- Recall the MA(q) model

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

white noise

constant

- It uses past forecast errors in a regression-like model
- Changing the parameters will result in different time series patterns

# Moving average models

- Recall the invertibility constraints
  - For an MA(1) model: $-1 < \theta_1 < 1$
  - For an MA(2) model: $-1 < \theta_2 < 1$, $\theta_1 + \theta_2 > -1$ ,

    $\theta_1 - \theta_2 < 1$
  - For more complicated conditions q>=3, R takes care of it when estimating the models.

# Non-seasonal ARIMA models

- Recall the ARIMA(p,d,q)

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t$$

The differenced series,
the d parameter
in the ARIMA

Order of the
autoregressive part

Order of the moving
average part

- auto.arima() will select the p,d,q automatically.

# auto.arima()

- Plot the usconsumption[,1]
- Use the auto.arima() to fit the data
- Which model is being suggested?
- Plot the forecast for h=5

# The effect of the ARIMA model parameters

- If $c=0$ and $d=0$, the long-term forecast will go to zero
- If $c=0$ and $d=1$, the long-term forecast will go to a non-zero constant
- If $c=0$ and $d=2$, the long-term forecasts will follow a straight line
- If $c \neq 0$ and $d=0$, the long-term forecasts will go to the mean of the data
- If $c \neq 0$ and $d=1$, the long-term forecasts will follow a straight line
- If $c \neq 0$ and $d=2$, the long-term forecasts will follow a quadratic trend

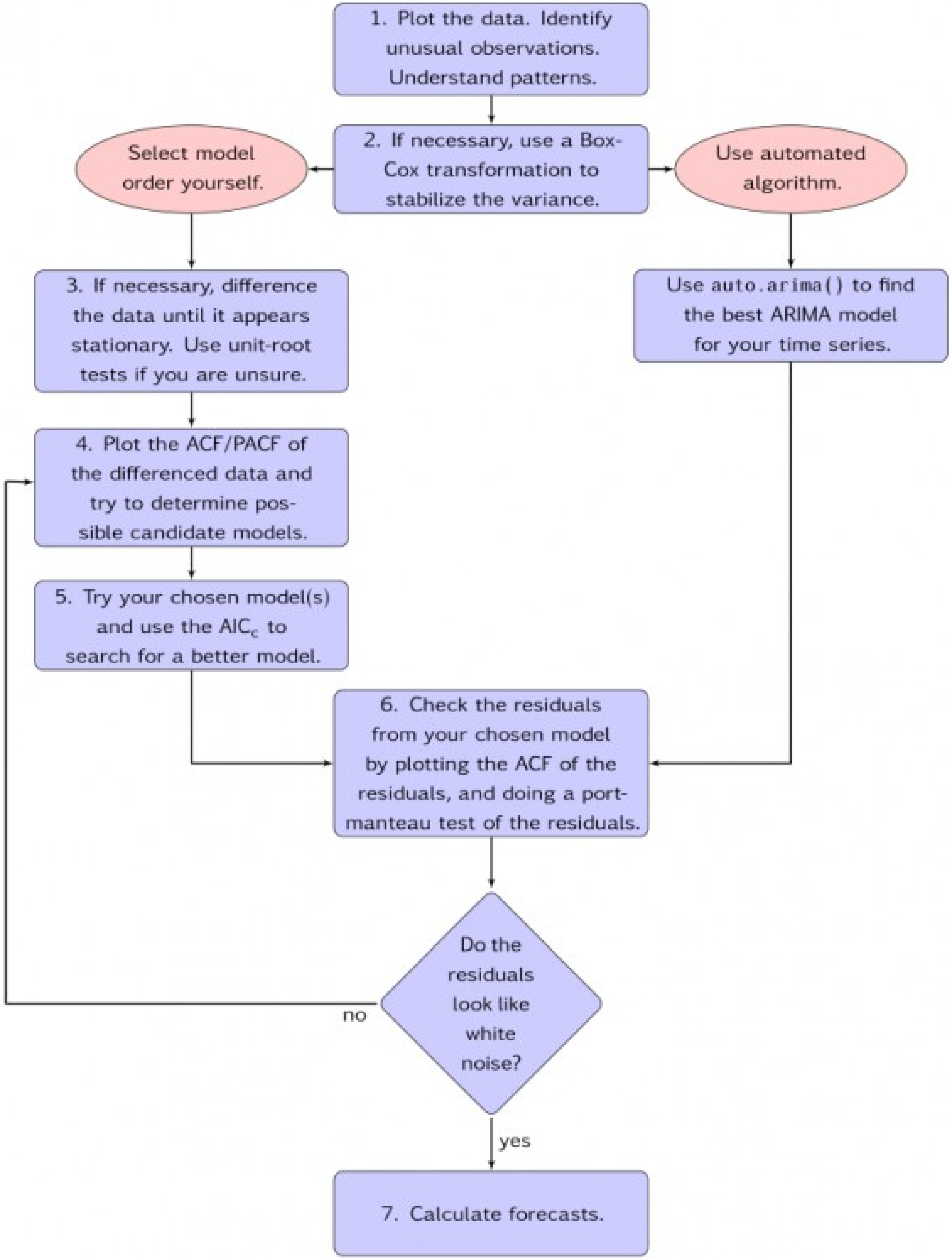
- The higher the value of d the more rapidly the prediction intervals increase in size.
- To obtain cyclic forecasts it is necessary to have $p>=2$, together with some additional conditions.

# Using ACF and PACF to find p and q

- Recall from the theory:
  - Data may follow an ARIMA(p,d,0) if:
    - The ACF is exponentially decaying or sinusoidal
    - There is a significant spike at lag p in PACF, but none beyond lag p
  - Data may follow an ARIMA(0,d,q) if:
    - The PACF is exponentially decaying or sinusoidal
    - There is a significant spike at lag q in ACF, but none beyond lag q

# auto.arima()

- Plot the Acf() and Pacf() for the usconsumption[,1]
- Does your observations agree with the model selected by auto.arima()?
- Try to plot the forecast again using the Arima() function and the model you have chosen
- What happens to the forecast if you use another model?


1. Plot the data. Identify unusual observations. Understand patterns.
2. If necessary, use a Box-Cox transformation to stabilize the variance.
Select model order yourself.
Use automated algorithm.
3. If necessary, difference the data until it appears stationary. Use unit-root tests if you are unsure.
Use auto.arima() to find the best ARIMA model for your time series.
4. Plot the ACF/PACF of the differenced data and try to determine possible candidate models.
5. Try your chosen model(s) and use the $AIC_c$ to search for a better model.
6. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a portmanteau test of the residuals.
Do the residuals look like white noise?
no
yes
7. Calculate forecasts.

# Applying the procedure

- Take the seasonally adjusted electrical equipment orders data

  ```
  eeadj <-
  seasadj(stl(elecequip,s.window="periodic"))
  ```

- Apply the procedure
- Try using the auto.arima()
- What can you conclude?

# Take a closer look...

- p-value
  - When it is small (p-value<=0.05) strong evidence against the null hypothesis
  - When it is large (p-value>0.05) weak evidence against the null hypothesis

- Always check what is the null hypothesis of your function in order to take a conclusion

# Seasonal ARIMA models

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

# Using ACF and PACF to find P and Q

- Recall that the seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF.
  - For example, an ARIMA(0,0,0)(0,0,1)_12 model will show:
    - A spike at lag 12 in the ACF but no other significant spikes.
    - The PACF will show exponential decay in the seasonal lags; that is at lags 12, 24, 36, …
  - An ARIMA(0,0,0)(1,0,0)_12 model will show:
    - Exponential decay in the seasonal lags of the ACF
    - A single significant spike at lag 12 in the PACF

# Applying the procedure to a seasonal time series

- Select the quarterly European retail trade data from 1996 to 2011 as your data set.
  - Copy paste the code of the part “Applying the procedure to a seasonal time series”
  - Can you identify the different steps and the rationale that lead to them?

- Attention: when comparing the AIC_c values the models must have the same orders of differencing! (This does not apply when you are comparing the models using a test set).

# Example 1

- For the mcopper data:

  (a) if necessary, find a suitable Box-Cox transformation for the data

  (b) fit a suitable ARIMA model to the transformed data using auto.arima()

  (c) try some other plausible models by experimenting with the orders chosen

  (d) chose what you think is the best model and check the residual diagnostics

  (e) produce forecasts of your fitted model. Do the forecasts look reasonable?

  (f) compare the results with what you would obtain using ets() (with no transformation)

# Example 2

- For the condmilk data:

  (a) Do the data need transforming? If so, find a suitable transformation

  (b) Are the data stationary? If not, find an appropriate differencing which yields stationary data

  (c) Identify a couple of ARIMA models that might be useful in describing the time series. Which of your models is the best according to their AIC_c values?

  (d) Estimate the parameters of your best model and do diagnostic testing on the residuals. Do the residuals resemble white noise? If not, try to find another ARIMA model which fits better.

  (e) Forecast the next 24 months of data using your preferred model.

  (f) Compare the forecasts obtained using ets().