3 Box-Jenkins methodology

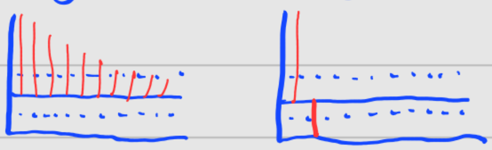
Step 1: model selection

↳ Plot data & correlograms

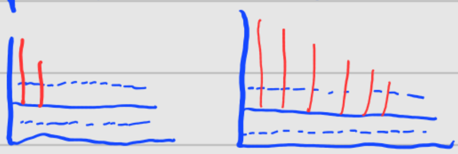
↳ If non-seasonal, use the following RoT's to determine ARMA(p,q):

i. If no sig. ACF or PACF spikes then it's white noise & not an ARMA process

ii. If ACF decreases slowly & lin. through zero, or a wave-like pattern then not stationary (try differencing)

iii. 

AR(2) process because of SPACF cut-off.

iv. 

MA(2) process b/c of SACF cut-off

v. Both ACF & PACF's converge to zero following some exp. or damped sine wave ∴ ARMA(p,q) model.

↳ Need to do more analysis (auto.arima()) for p & q values

Step 2: model estimation

↳ Estimate the specified ARIMA model using the Arima() function from forecast package

Step 3: diagnostic checks

↳ If in step 1 you selected more than one model use model selection criterion to make a final decision (the smaller the better)

↳ The residuals should behave as white noise (i.e. show ACF & PACF for residuals)

↳ Check for normality as well (espec. if t is small)

↳ If you're satisfied do one final check compare model selection criteria against an overparametised model (i.e. AR(p+1))

↳ If you've chosen the "correct" model then this term should be insignificant.

Key ARMA formulae

MA(1): $y_t = \theta_1 \varepsilon_{t-1}, \ \varepsilon_t \sim WN(0, \sigma^2)$

$E(y_t) = 0$

$Var(y_t) = \sigma^2(1+\theta_1^2)$

every MA(1) but the first is zero. memory of MA(1) is only 1 period long

$y_{T+h} = \varepsilon_{T+h} + \theta_1 \varepsilon_{T+h-1}, \ h > 1$

$\hookrightarrow E(y_{T+h}) = E_T(\varepsilon_{T+h}) + \theta E_T(\varepsilon_{T+h-1}) = 0$

$e_{T+h} = y_{T+h} - E_T(y_{T+h})$

$\hookrightarrow E(e_{T+h}) = E(\varepsilon_{T+h} + \theta \varepsilon_{T+h-1}) = 0$

$Var(e_{T+h}) = \sigma^2(1+\theta_1^2)$

AR(1): $\varphi_1 y_{t-1} + \varepsilon_t, \ |\varphi| < 1, \ \varepsilon_t \sim WN(0, \sigma^2)$

$E(y_t) = 0, \ Var(y_t) = \frac{\sigma^2}{1-\varphi_1^2}$

$y_{t+h} = \varphi_1 y_{T+h-1} + \varepsilon_{T+h}, \ h > 1$

$\hookrightarrow E(y_{T+h}) = \varphi_1 E_T(y_{T+h-1}) + \underbrace{E(\varepsilon_{T+h})}_{=0} = \varphi_1^h y_T$

$e_{T+h} = y_{T+h} - E_T(y_{T+h}) = \varphi_1^{h-1}\varepsilon_{T+1} + \varphi_1^{h-2}\varepsilon_{T+2} + \dots + \varphi_1 \varepsilon_{T+h-1} + \varepsilon_{T+h}$

$\hookrightarrow E(e_{T+h}) = 0$

$Var(e_{T+h}) = \sigma^2 \frac{1-\varphi_1^{2h}}{1-\varphi_1^2} \xrightarrow[h\to\infty]{} \frac{\sigma^2}{1-\varphi_1^2}$