Date: Thursday, November 8, 2018

Recap

- $\{Y_t\} \sim ARMA(p,q)$:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

where $\{\varepsilon_t\} \sim WN(0, \sigma^2)$

$\Leftrightarrow$

$$\phi(B) Y_t = \theta(B) \varepsilon_t$$

where

- $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$
- $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q$

- An $ARMA(p,q)$ model is stationary iff the zeros of $\phi(z)$ all lie outside the unit circle in the complex plane.

- An $ARMA(p,q)$ model is invertible iff the zeros of $\theta(z)$ all lie outside the unit circle in the complex plane.

⏱ the watch sweats

In order to estimate $(\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \sigma^2)$ with observed data $\{y_1, y_2, \dots, y_n\}$ we can use maximum likelihood estimation or least squares estimation.

Least squares estimation

Here, we want to find values of the parameters that minimizes errors. Specifically, sum of squared errors:

$$\boxed{S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2}$$

Notice that this does not depend on σ, but the LS estimates of σ, once the other parameters have been estimated, is given by:

(Gives us an idea of how well our model fits data kinda R^2)

$$\boxed{\hat{\sigma} = \sqrt{\frac{S(\hat{\phi}, \hat{\theta})}{n-p-q}}}$$

Sum of squared residuals

Once a model has been fit to the data, we should verify that the assumptions it makes are valid. Specifically, we assume that $\{\varepsilon_t\} \sim WN(0, \sigma^2)$.

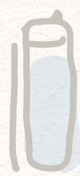
We can think of residuals as sample estimates of the error terms, and so we should expect them to behave like $WN(0, \sigma^2)$.

We can check this by determining whether the residuals $\{e_t\}$:

- have zero mean
- have constant variance
- are uncorrelated
- follow a normal distribution ← specifically when MLE is used.

Generally speaking, we refer to this step as "verification". Thus, we can think of modeling as a 4 step procedure:

1. Order selection.
2. Modeling Fitting (Estimation).
3. Verification.
4. Forecasting.



water bottle!

$\text{diff}(y_t) = y_t - y_{t-1}$

Trade off: Normal distribution vs AIC

1st. MLE if not, try LSE

[R]

[break]

log-likelihood : bigger is better

aic: function of log-likelihood but penalizes overfitting. smaller is better. (only for MLE)

σ^2 : |smaller| is better.

More on order selection:

- Use $\hat{\sigma}^2$, $\ln L(\cdot)$, and AIC to help choose optimal orders p,q.

$\hat{\sigma}^2$ ↓ smaller better; $\ln L(\cdot)$ ↓ bigger better; AIC ↓ smaller better

- Note that AIC penalizes you for over-complicating your model, and hence protects you from overfitting. The other metrics don't do this. (is the difference significant?)*

(These don't say anything about predicting the future. Maybe only AIC because it protects you from over fitting.)

"Akaike Information Criterion"

$\hat{\phi}, \hat{\theta}, \hat{\sigma}$

- $AIC = -2 \ln L(\cdot) + 2(p+q+1)$

penalty term

- $AICC = -2 \ln L(\cdot) + \dfrac{2(p+q+1)n}{n-p-q-2}$

corrected

Still penalizes for a complicated model, but adjusts to the amount of data. Not as harsh if n is big.

- These both depend on $\ln L(\cdot)$, which means we cannot calculate this in the context of Least Squares Estimation.

* To formally compare models we can use a likelihood ratio test (LRT).

ARMA(1,2) MA(1) ✓ AR(3) MA(1) ✗

It's important to note that the models being compared must be "nested" within one another.

H_0: reduced model and full model fit the data equally well

H_A: full model fits the data better than the reduced one.

Test Statistic

$$\boxed{D = -2 \log \left[\frac{L(\text{reduced model})}{L(\text{full model})} \right] \sim \chi^2_{(m_F - m_R)}}$$

where m_F = # of parameters in full model

m_R = # of parameters in reduced model

↳ Large values of D provide evidence against H_0.

[R]