Estimation of Model Parameters

(rough outline → not pursued in detail since we do not want to get into optimization)

$$X_t\text{-s given} \xrightarrow{?} \phi\text{-s}, \theta\text{-s} \ \& \ \sigma_a^2 = ?$$

If we have AR(n) model, things are easy

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_n X_{t-n} + a_t \qquad a_t \sim NIID(0, \sigma_a^2)$$

$$\underbrace{\begin{bmatrix} X_{n+1} \\ X_{n+2} \\ \vdots \\ X_N \end{bmatrix}}_{\underline{Y}} = \underbrace{\begin{bmatrix} X_n & X_{n-1} & \dots & X_1 \\ X_{n+1} & X_n & \dots & X_2 \\ - & - & - & - \\ X_{N-1} & \dots & \dots & X_{N-n} \end{bmatrix}}_{\underline{X}} \underbrace{\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix}}_{\vec{\phi}} + \underbrace{\begin{bmatrix} a_{n+1} \\ a_{n+2} \\ \vdots \\ a_N \end{bmatrix}}_{\vec{\varepsilon}}$$

$$\hat{\phi} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y}$$

$$\hat{\sigma}_a^2 = \underset{\substack{\uparrow \\ \text{unbiased}}}{\frac{1}{N-2n}} \sum_{t=n+1}^{N} (X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_n X_{t-n})$$

$\frac{1}{N-n}$ ← biased, but min variance!

– ARMA (n, n-1) model

$$X_t - \phi_1 X_{t-1} - \dots - \phi_n X_{t-n} = a_t - \theta_1 a_{t-1} - \dots - \theta_{n-1} a_{t-n+1}$$

We need to find ϕ_s^* and θ-s that will minimise residual sum of squares

$$a_t = X_t - \phi_1 X_{t-1} - \dots - \phi_n X_{t-n} + \theta_1 a_{t-1} + \dots + \theta_{n-1} a_{t-n+1}$$

$$a_{t-1} = X_{t-1} - \phi_1 X_{t-2} - \dots - \phi_n X_{t-n-1} + \theta_1 a_{t-2} + \dots + \theta_{n-1} a_{t-n}$$

$$a_{t-2} = \dots$$

Substitution of $a_{t-1}, a_{t-2}, \dots, a_0$ into (*) gives us a non-linear optimiz. problem

Basically each combination of $\vec{\phi}$-s and $\vec{\theta}$-s will give us some $RSS(\vec{\phi}, \vec{\theta})$. We need to find $\hat{\vec{\phi}}, \hat{\vec{\theta}}$ that will minimize the RSS

$$RSS(\vec{\phi}, \vec{\theta}) = \sum_{t=1}^{N} a_t^2$$

(usually you assign $a_0 = a_1 = \dots = a_{n-1} = 0$ or you can include them into optimization)

Marquard's method, Steepest descent, etc → all can be used to descend down the RSS curve!

This is a multi-modal problem & getting close to the solution is very important!

"armax" command from Matlab executes optimization used in my code "PostulateARMA"

$\hat{\vec{\phi}} \pm \Delta\phi \quad \hat{\vec{\theta}} \pm \Delta\vec{\theta}$ is given, where $\pm\Delta$-s come from local linear approximations of the problem (Jacobians take over the role of $\underline{X}$).

It is possible to define a problem like

$$(\overset{\circ}{X}_t - \mu) - \phi_1 (\overset{\circ}{X}_{t-1} - \mu) - \dots - \phi_n (X_{t-n} - \mu) =$$

$$= a_t - \theta_1 a_{t-1} - \dots - \theta_{n-1} a_{t-n+1}$$

and estimate μ-s, θ-s and ϕ-s, but ARMAX doesn't do it → you need to estimate $\mu = \bar{X}_t$ and subtract it.

How to get "close" to a solution?

Initial Guess

One way (analytically tractable & easy to understand) can be to use I.F. approximation.

$$X_t = a_t + \sum_{j=0}^{\infty} I_j X_{t-j} \qquad (*)$$

If I know I_j-s, then I also know all ϕ-s & θ-s!
If I approximate I_j-s well, then I approximate ϕ-s & θ-s well! Why I-s?

Well, $(*)$ is $AR(\infty)$ model & if I can truncate it somehow, I can get a nice AR model whose params I can get easily! Let's say I do $AR(p)$

$$X_t \approx a_t - I_1 X_{t-1} - \dots - I_p X_{t-p}$$

and using simple LS, I can find $I_1, I_2, \dots, I_p$, out of which I can look for ~~ϕ-s~~ ϕ-s and θ-s!

I will now do analysis for $ARMA(n,m)$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_n B^n) X_t = (1 - \theta_1 B - \dots - \theta_m B^m) a_t$$

$$a_t = -(I_0 + I_1 B + I_2 B^2 + \dots) X_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_n B^n) X_t = -(1 - \theta_1 B - \dots - \theta_m B^m)(I_0 + I_1 B + \dots) a_t$$

$B_0: \; I_0 = -1$

$B_1: \; \phi_1 = I_1 - \theta_1 I_0$

$B_2: \; \phi_2 = I_2 - \theta_1 I_0 - \theta_0 I_1$

B^j: $\phi_j = I_j - \theta_1 I_{j-1} - \dots - \theta_j I_0$ (a)

$\vdots$

For $j \geq \max(m+1, n) + 1$

B^j: $0 = I_j - \theta_1 I_{j-1} - \dots - \theta_m I_{j-m}$ (b)

If I have m equations (b), I can get θ-s!

Then, I can use n equations (a) to get ϕ-s!

Hence, I need at least $p = \max(m+1, n) + m$ Inverse Fcn coefficients to get these guesses (usually, this is actually $m+n$).

<u>Example</u>: Getting initial guess for Sunspot Data in Table A2 (pp. 487). I wish to fit ARMA(2,1) model

$\Rightarrow$ I need $p = \max(2,2) + 1 = 3$ I.F. params

$$X_t = 1.27 X_{t-1} - 0.5 X_{t-2} - 0.11 X_{t-3} + a_t$$

$\Rightarrow$ $I_1 \approx 1.27$ $\quad I_2 \approx -0.5$ $\quad I_3 \approx -0.11$

Since it's ARMA(2,1), for $j \geq 3$ I have

$$(1 - \theta_1 B) I_j = 0$$

<u>Indeed</u>

$$(1-\phi_1 B-\phi_2 B^2) X_t = -(1-\theta_1 B)(I_0 + I_1 B + I_2 B^2 + \dots) a_t$$

$B^0: -1 = I_0$

$B^1: \phi_1 = I_1 - \theta_1 I_0$

$B^2: \phi_2 = I_2 - \theta_1 I_1$

$B^3: 0 = I_3 - \theta_1 I_2 \quad \Rightarrow \quad \theta_1^{in} \approx \dfrac{I_3}{I_2} = 0.22$ now, plug this into top 2 eqns

$\vdots$

$\phi_1 = I_1 - \theta_1 I_0 \approx 1.49$ and $\phi_2 = I_2 - \theta_1 I_1 \approx -0.78$

$\Rightarrow$ my initial guesses are $\boxed{\phi_1 \approx 1.49 \quad \phi_2 \approx -0.78 \quad \theta_1 \approx 0.22}$

Prediction (Chap. 5)

Given $X_0, X_1, X_2, \ldots, X_t$, what is X_{t+1}? X_{t+2}? ...

Best prediction of X_{t+l} given $X_0, X_1, X_2, \ldots, X_t$ is

$$\hat{X}_t(l) = E[X_{t+l} \mid X_0, X_1, \ldots, X_t]$$

(best in the least squares sense)

Since, given $X_0, X_1, \ldots, X_t$, I also know $a_0, a_1, \ldots, a_t$

$$X_t = \sum_{i=0}^{\infty} G_i a_{t-i} \quad \text{and} \quad a_t = -\sum_{j=0}^{\infty} I_j X_{t-j}$$

$$\hat{X}_t(l) = E[X_{t+l} \mid X_0, X_1, \ldots, X_t] = E[X_{t+l} \mid a_0, a_1, \ldots, a_t]$$

(I will label this as $E[X_{t+l} \mid t]$)

Prediction through conditional expectation

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_n X_{t-n} + a_t - \theta_1 a_{t-1} - \ldots - \theta_{n-1} a_{t-n+1}$$

$$\underbrace{X_{t+1} = \phi_1 X_t + \dots + \phi_n X_{t-n+1} + a_{t+1} - \theta_1 a_t - \dots - \theta_{n-1} a_{t-n+2}}_{E[\,\cdot\,|\,t]}$$

$$\hat{X}_t(1) = E[X_{t+1} \,|\, t] = \phi_1 X_t + \dots + \phi_n X_{t-n+1} - \theta_1 a_t - \dots - \theta_{n-1} a_{t-n+2}$$

$$\hat{X}_t(2) = \phi_1 \hat{X}_t(1) + \phi_2 X_t + \dots + \phi_n X_{t-n+2} - \theta_2 a_t - \dots - \theta_{n-1} a_{t-n+3}$$

$$\vdots$$

$$\hat{X}_t(l) = \phi_1 \hat{X}_t(l-1) + \phi_2 \hat{X}_t(l-2) + \dots + \phi_n \hat{X}_t(l-n)$$

for $l > n$

Hence, we can use ARMA model to recursively evaluate predictions. Please note, ultimately, it is the AR part that drives prediction!!! Stability must play a pretty important role here. Indeed, we'll see it when we start looking at prediction using orthogonal decomposition of X_t (G.F.)

$$X_t = G_0 a_t + G_1 a_{t-1} + \dots + G_\ell a_{t-\ell} + \dots$$

$$X_{t+\ell} = G_0 a_{t+\ell} + G_1 a_{t+\ell-1} + \dots + G_\ell a_t + G_{\ell+1} a_{t-1} + \dots$$

$$E[X_{t+\ell} \mid t] = G_\ell a_t + G_{\ell+1} a_{t-1} + \dots = \hat{X}_t(\ell)$$

a_t-s happened but I cannot know them exactly (can only estimate them, since I need a few a_t-s to start the process of recursively calculating a_t-s)

$$\hat{e}_t(\ell) = -\hat{X}_t(\ell) + X_{t+\ell} = G_0 a_{t+\ell} + G_1 a_{t+\ell-1} + \dots + G_{\ell-1} a_{t+1}$$

$$Var[\hat{e}_t(\ell)] = \sigma_a^2 [G_0^2 + G_1^2 + \dots + G_{\ell-1}^2]$$

Very important because it gives us info on how accurate prediction is.

Note 1: $Var[\hat{e}_t(l)] \leq Var[\hat{e}_t(l+1)]$

The further I try to predict, the worse my prediction becomes!

Note 2: What happens ~~??~~ as $l \to \infty$ ('eventual forecast')

$$Var[\hat{e}_t(\infty)] = \sigma_a^2 \left(\sum_{l=0}^{\infty} G_l^2\right) \begin{cases} \to \text{ if stable } < \infty \\ \to \text{ if unstable } > \infty \end{cases}$$

=> Stable t-s predictions have limited variance of prediction errors

Unstable t-s predictions have variance of pred. errors growing beyond limits.

What happens with the ACTUAL prediction? (not the variance of prediction errors?)

Remember $\hat{X}_t(l) - \phi_1 \hat{X}_t(l-1) - \dots - \phi_n \hat{X}_t(l-n) = 0$

$>> \hat{X}_t(l) = C_1 \lambda_1^l + C_2 \lambda_2^l + \dots + C_n \lambda_n^l$ for some constants C

- If the model is stable $\Rightarrow \hat{X}_t(l) \xrightarrow[l\to\infty]{} 0$
- If the model is marginally stable $\Rightarrow \hat{X}_t(l) \to$ constant or sinusoid (but var $\to \infty$)
- If the model is unstable $\Rightarrow \hat{X}_t(l) \to \infty$

Updating forecasts

t: $X_0 \quad X_1 \quad \ldots \quad X_t \;|\; \hat{X}_t(1) \quad \hat{X}_t(2) \quad \ldots \quad \hat{X}_t(l) \quad \ldots$

$t+1$: $X_0 \quad X_1 \quad \ldots \quad X_t \quad X_{t+1} \;|\; \underbrace{\hat{X}_{t+1}(1) \quad \ldots \quad \hat{X}_{t+1}(l-1)}_{} \quad \ldots$

can I update these using info that just freshly arrived!?

Remember,

$$X_{t+1} = G_0 a_{t+1} + G_1 a_t + \ldots$$

$$\hat{X}_t(1) = G_1 a_t + \ldots$$

$$\Rightarrow a_{t+1} = X_{t+1} - \hat{X}_t(1)$$

Remember

$$\hat{X}_t(\ell) = G_\ell a_t + G_{\ell+1} a_{t-1} + \dots$$

$$\hat{X}_{t+1}(\ell-1) = G_{\ell-1} a_{t+1} + \underbrace{G_\ell a_t + \dots}_{\hat{X}_t(\ell)}$$

$$\Rightarrow \hat{X}_{t+1}(\ell-1) = \hat{X}_t(\ell) + G_{\ell-1} a_{t+1}$$

$$= \hat{X}_t(\ell) + G_{\ell-1}\left(X_{t+1} - \hat{X}_t(1)\right)$$