Date
Page

1.2.1) The goal of maximum likelihood is to find the optimal way to fit a distribution to the data.

Let's say there are n observations
We first collect the first p observations in the sample. $(Y_1, Y_2 \ldots Y_p)$ in a $(p \times 1)$ vector y_p.

For a gaussian AR(p) process

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t$$

where $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$

$\therefore$ y_p has mean

$$\mu = \frac{c}{1-\phi_1-\phi_2-\cdots-\phi_p}$$

Let $\sigma^2 V_p$ denote the $(p \times p)$ variance - covariance matrix of $(Y_1, Y_2 \ldots Y_p)$

$$\sigma^2 V_p = \begin{bmatrix} E(Y_1-\mu)^2 & E(Y_1-\mu)E(Y_2-\mu) & \cdots & E(Y_1-\mu)E(Y_p-\mu) \\ E(Y_2-\mu)(Y_1-\mu) & E(Y_2-\mu)^2 & \cdots & E(Y_2-\mu)E(Y_p-\mu) \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ E(Y_p-\mu)(Y_1-\mu) & E(Y_p-\mu)(Y_2-\mu) & & E(Y_p-\mu)^2 \end{bmatrix}$$

$$\sigma^2 V_p = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix}$$

The density of first p observations is then:

$$f_{Y_p Y_{p-1} \cdots Y_1}(y_p, y_{p-1}, \ldots y_1; \theta) =$$

$$(2\pi)^{-p/2} |\sigma^{-2} V_p^{-1}|^{1/2} \exp\left[\frac{-1}{2\sigma^2}(y_p-\mu_p)' V_p^{-1}(y_p-\mu_p)\right]$$

$$= (2\pi)^{-p/2} (\sigma^{-2})^{p/2} |V_p^{-1}|^{1/2} \exp\left[\frac{-1}{2\sigma^2}(y_p-\mu_p)' V_p^{-1}(y-\mu_p)\right]$$

For the remaining observations in the sample $(Y_{p+1}, Y_{p+2} \ldots Y_T)$, conditional on the first $t-1$ observations, the t^{th} observation is gaussian with mean

$$c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p}.$$

& variance σ^2. Only the p most recent observations matter for this distribution. hence for $t > p$

$$f_{Y_t | Y_{t-1}, \ldots Y_1}(y_t | y_{t-1}, \ldots y_1; \theta) = f(Y_t | Y_{t-1} \ldots Y_{t-p}; \theta)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_t - c - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p})^2}{2\sigma^2}\right]$$

The likelihood function for the complete sample is then

$$f_{Y_T, Y_{T-1} \ldots Y_1}(y_T, y_{T-1}, \ldots y_1; \theta) =$$

$$f_{Y_p Y_{p-1} \ldots Y_1}(y_p, y_{p-1} \ldots, y_1; \theta)$$

$$\times \prod_{t=p+1}^{T} f_{Y_t | Y_{t-1}, \ldots, Y_{t-p}}(y_t | y_{t-1} \ldots y_{t-p}; \theta)$$

classmate
Date
Page

The log likelihood is therefore

$$L(\theta) = \log f_{Y_T, Y_{T-1} \dots Y_1}(y_T, y_{T-1} \dots, y_1 ; \theta)$$

$$= \frac{-p}{2}\log(2\pi) - \frac{p}{2}\log(\sigma^2) + \frac{1}{2}\log|V_p^{-1}| - \frac{1}{2\sigma^2}(y_p - \mu_p)'$$

$$V_p^{-1}(y - \mu_p) - \left(\frac{T-p}{2}\right)\log 2\pi - \frac{T-p}{2}\log(\sigma^2)$$

$$- \sum_{t=p+1}^{T} \frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2}{2\sigma^2}$$

The value of $\phi_1, \phi_2 \dots \phi_p$ that maximizes the above equation are the same as thoes that minimize

$$\sum_{t=p+1}^{T} (y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2$$

This is an exactly similar equation like OLS
∴ The conditional MLE of these parameters can be obtained from an OLS regression of y_t on a constant & p of it's own lagged terms.

∴ The 2 estimates are exactly the same.

classmate

Date
Page

12.2) Consider simple MA(1) model, $Y_t = e_t - \theta e_{t-1}$

The true lag-1 autocorrelation in this model is $\rho_1 = \frac{-\theta}{1+\theta^2}$

If we equate ρ_1 to r_1, we get a quadratic equation in θ.

if $|r_1| < 0.5$, then only one of the 2 real solutions satisfies the invertibility condition $|\theta| < 1$

That solution is $\hat{\theta} = \left(-1 + \sqrt{1-4r_1^2}\right) / 2r_1$

But if $|r_1| = 0.5$, no invertible solution exists & if $|r_1| > 0.5$ then no real solution at all exists & the method of moments fail to give any estimator of θ.

With higher order MA(q) models, the set of equations for estimating $\theta_1, \dots \theta_p$ is highly non-linear & could be only solved numerically

Hence model estimation procedure is a bit harder

2) Consider now the MA(1) model

$$Y_t = e_t - \theta e_{t-1}$$

This can be written as

$$Y_t = -\theta Y_{t-1} - \theta^2 Y_{t-2} - \theta^3 Y_{t-3} - \dots + e_t$$

So a least square estimator of θ can be abtained by finding the value of θ that minimizer

$$S_c(\theta) = \Sigma\left[Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} \dots \right]^2$$

But this is non-linear in θ & the infinite series causes. technical problems