Lecture 24

ML Estimation

Arnab Hazra

Problem of MoM for other models

- AR(p) models are basically linear models, and the Yule-Walker estimators are essentially least squares estimators.
- If we use MoM for MA or ARMA models, we will not get optimal estimators because such processes are nonlinear in the parameters.
- For the MA(1) model $X_t = W_t + \theta W_{t-1}$, we can write as

$$X_t = \sum_{j=1}^{\infty} (-\theta)^j X_{t-j} + W_t,$$

  which is nonlinear in θ.
- Here $\gamma(0) = \sigma_W^2(1 + \theta^2)$ and $\gamma(1) = \theta\sigma_W^2$.
- Hence, $\hat{\rho}(1) = \hat{\gamma}(1)/\hat{\gamma}(0) = \hat{\theta}/(1 + \hat{\theta}^2)$. Because $|\hat{\theta}/(1 + \hat{\theta}^2)| \leq 1/2$ but $|\hat{\rho}(1)|$ does not necessarily satisfy this condition, MoM is problematic for this model.

ML estimation

- We first focus on the causal AR(1) case. Let $X_t = \mu + \phi(X_{t-1} - \mu) + W_t$.
- Here $W_t \overset{IID}{\sim} \mathrm{N}(0, \sigma_W^2)$. The likelihood is $L(\mu, \phi, \sigma_W^2) = f(X_1, \ldots, X_T | \mu, \phi, \sigma_W^2)$.
- Due to AR(1) structure, $L(\mu, \phi, \sigma_W^2) = f(X_1) f(X_2 | X_1) \ldots f(X_T | X_{T-1})$.
- Because $X_t | X_{t-1} \sim \mathrm{N}(\mu + \phi(X_{t-1} - \mu), \sigma_W^2)$, we have

$$f(X_t | X_{t-1}) = f_W[(X_t - \mu) - \phi(X_{t-1} - \mu)].$$

- Overall

$$L(\mu, \phi, \sigma_W^2) = f(X_1) \prod_{t=2}^{T} f_W[(X_t - \mu) - \phi(X_{t-1} - \mu)].$$

- Considering $f(X_1)$, our approach is called unconditional least square, and after ignoring it, it is called unconditional least square.

Calculation of likelihood

- To find $f(X_1)$, we can use the causal representation $X_1 = \mu + \sum_{j=0}^{\infty} \phi^j W_{1-j}$ to see that X_1 is normal, with mean μ and variance $\sigma_w^2/(1-\phi^2)$.
- Finally, the likelihood is

$$L(\mu, \phi, \sigma_W^2) = (2\pi\sigma_W^2)^{-T/2}(1-\phi^2)^{1/2} \exp\left[-\frac{S(\mu,\phi)}{2\sigma_W^2}\right],$$

where $S(\mu, \phi) = (1-\phi^2)(X_1-\mu)^2 + \sum_{t=2}^{T}[(X_t-\mu) - \phi(X_{t-1}-\mu)]^2$.

- $S(\mu, \phi)$ is called the unconditional sum of squares. The MLE of σ_W^2 is $\hat{\sigma}_W^2 = T^{-1} S(\hat{\mu}, \hat{\phi})$ where $\hat{\mu}$ and $\hat{\phi}$ are the MLEs of μ and ϕ, respectively.
- We obtain this by maximizing $l(\mu, \phi) = \log[T^{-1} S(\mu, \phi)] - T^{-1}\log(1-\phi^2)$.

Conditional likelihood

- $L(\mu, \phi, \sigma_W^2)$ is a complicated function of the parameters, and hence, the minimization of $l(\mu, \phi)$ or $S(\mu, \phi)$ is accomplished numerically.
- In the case of AR models, conditional on initial values, they are linear models.
- We can drop the term in the likelihood that causes the nonlinearity.
- Conditioning on X_1, the conditional likelihood becomes

$$L(\mu, \phi, \sigma_W^2 | X_1) = \prod_{t=2}^{T} f_W[(X_t - \mu) - \phi(X_{t-1} - \mu)].$$

- Finally, the conditional likelihood is

$$L(\mu, \phi, \sigma_W^2 | X_1) = (2\pi\sigma_W^2)^{-(T-1)/2} \exp\left[-\frac{S_c(\mu, \phi)}{2\sigma_W^2}\right],$$

where $S_c(\mu, \phi) = \sum_{t=2}^{T}[(X_t - \mu) - \phi(X_{t-1} - \mu)]^2$.

Maximum conditional likelihood estimates

- $S_c(\mu, \phi)$ is called the conditional sum of squares. The MLE of σ_W^2 is $\hat{\sigma}_W^2 = (T-1)^{-1} S_c(\hat{\mu}, \hat{\phi})$ where $\hat{\mu}$ and $\hat{\phi}$ are the MLEs of μ and ϕ, respectively.

- Following the results from least squares estimation, we have

$$\hat{\mu} = \frac{\overline{X}_{(2)} - \hat{\phi}\,\overline{X}_{(1)}}{1-\hat{\phi}}, \quad \hat{\phi} = \frac{\sum_{t=2}^{T}(X_t - \overline{X}_{(2)})(X_{t-1} - \overline{X}_{(1)})}{\sum_{t=2}^{T}(X_{t-1} - \overline{X}_{(1)})^2},$$

where $\overline{X}_{(2)} = \frac{\sum_{t=2}^{T} X_t}{T-1}$ and $\overline{X}_{(1)} = \frac{\sum_{t=2}^{T} X_{t-1}}{T-1}$.

MLE for ARMA(p, q) model

- For a normal ARMA(p, q) model, let $\boldsymbol{\beta} = (\mu, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$ be the ($p + q + 1$)-dimensional vector of the model parameters.
- The likelihood can be written as

$$L(\boldsymbol{\beta}, \sigma_W^2) = f(X_1)f(X_2|X_1)f(X_3|X_1, X_2)\ldots f(X_T|X_1, \ldots, X_{T-1}).$$

- The conditional distribution of X_t given $X_{t-1}, \ldots, X_1$ is Gaussian with mean X_t^{t-1} and variance P_t^{t-1}. Recall that $P_t^{t-1} = \gamma(0)\prod_{j=1}^{t-1}(1 - \phi_{j,j}^2)$.
- For ARMA model, $\gamma(0) = \sigma_W^2 \sum_{j=0}^{\infty} \psi_j^2$, in which case we may write

$$P_t^{t-1} = \sigma_W^2 \sum_{j=0}^{\infty} \psi_j^2 \prod_{j=1}^{t-1}(1 - \phi_{j,j}^2) \stackrel{def}{=} \sigma_W^2 r_t.$$

- Here $r_1 = \sum_{j=0}^{\infty} \psi_j^2$ and $r_{t+1} = r_t(1 - \phi_{t,t}^2)$.

Calculation of likelihood

- Finally, the likelihood is

$$L(\beta, \sigma_W^2) = (2\pi\sigma_W^2)^{-T/2}[r_1(\beta)\dots r_T(\beta)]^{-1/2} \exp\left[-\frac{S(\beta)}{2\sigma_W^2}\right],$$

  where $S(\beta) = \sum_{t=1}^{T}[X_t - X_t^{t-1}(\beta)]^2/r_t(\beta)$.

- The MLE of σ_W^2 is $\hat{\sigma}_W^2 = T^{-1}S(\hat{\beta})$ where $\hat{\beta}$ is the MLE of β, respectively.

- We obtain this by maximizing $l(\beta) = \log[T^{-1}S(\beta)] + T^{-1}\sum_{t=1}^{T}\log(r_t(\beta))$.

- We can use Newton-Raphson and Fisher-scoring algorithms.

Thank you!