Lecture 22

Estimation

Arnab Hazra

Introduction

- We assume we have T observations, $X_1, \ldots, X_T$, from a causal and invertible Gaussian ARMA(p, q) process.
- For the time being, the order parameters, p and q, are known.
- Our goal is to estimate the parameters, $\phi_1, \ldots, \theta_p, \theta_1, \ldots, \theta_q$, and σ_W^2.
- If $E(X_t) = \mu$, then the method of moments estimator of μ is the sample average, $\overline{X}$.
- Thus, while discussing method of moments, we will assume $\mu = 0$.

ACF of an ARMA(p, q) process (Recap)

- A causal ARMA(p, q) model $\{X_t; t = 0, \pm 1, \pm 2, \ldots\}$ can be written as a one-sided linear process $X_t = \sum_{j=0}^{\infty} \psi_j W_{t-j} = \psi(B) W_t$.

- We have

$$\gamma(h) = \mathrm{Cov}(X_{t+h}, X_t) = \sum_{j=1}^{p} \phi_j \gamma(h-j) + \sigma_w^2 \sum_{j=h}^{q} \theta_j \psi_{j-h}, \quad h \geq 0.$$

- From there, we can write $\gamma(h) - \sum_{j=1}^{p} \phi_j \gamma(h-j) = 0, \quad h \geq \max\{p, q+1\}$, with initial conditions

$$\gamma(h) - \sum_{j=1}^{p} \phi_j \gamma(h-j) - \sigma_w^2 \sum_{j=h}^{q} \theta_j \psi_{j-h} = 0, \quad 0 \leq h < \max\{p, q+1\}.$$

Yule-Walker equations

- We first consider the case of AR(p) models,

$$X_t = \phi_1 X_{t-1} + \ldots + \theta_p X_{t-p} + W_t.$$

- For $h = 1, 2, \ldots, p$, we can write

$$\gamma(h) - \sum_{j=1}^{p} \phi_j \gamma(h-j) = 0.$$

- For $h = 0$,

$$\gamma(0) - \sum_{j=1}^{p} \phi_j \gamma(j) - \sigma_w^2 = 0,$$

which implies $\sigma_w^2 = \gamma(0) - \sum_{j=1}^{p} \phi_j \gamma(j)$.

MoM estimator of ϕ

- $\boldsymbol{\Gamma}_p$ is a $p \times p$ matrix with its (i,j)th element $\gamma(i-j)$
- $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_p]'$
- $\boldsymbol{\gamma}_p = [\gamma(1), \ldots, \gamma(p)]'$
- In matrix notation, the Yule-Walker equations are

$$\boldsymbol{\Gamma}_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p, \quad \sigma_W^2 = \gamma(0) - \boldsymbol{\phi}' \boldsymbol{\gamma}_p.$$

- We replace $\gamma(h)$ by $\hat{\gamma}(h)$ and we obtain

$$\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p, \quad \hat{\sigma}_W^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\gamma}}_p' \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p.$$

- In practice, both sides are usually divided by $\hat{\gamma}(0)$ and the equation is rewritten in terms of sample ACF.

Large sample results

- For AR(p) models, if the sample size is large, the Yule-Walker estimators are approximately normally distributed, and $\hat{\sigma}^2_W$ is close to the true value of σ^2_W.
- The asymptotic ($T \to \infty$) behavior of the Yule-Walker estimators in the case of causal AR(p) processes is as follows:

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{D} \mathrm{MVN}_p(\mathbf{0}, \sigma^2_W \mathbf{\Gamma}_p^{-1}).$$

- The Durbin-Levinson algorithm can be used to calculate $\hat{\phi}$ without inverting $\hat{\mathbf{\Gamma}}_p$.
- For a causal AR(p) process, asymptotically ($T \to \infty$),

$$\sqrt{T}\hat{\phi}_{h,h} \xrightarrow{D} \mathrm{N}(0, 1), \quad \text{for} \quad h > p.$$

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.

Thank you!