Chapter 5: Parameter Estimation

Time Series Analysis
WISE, XMU

- Suppose that $\{Z_t\}$ follows an ARIMA(p, d, q) model, i.e. $W_t = \nabla^d Z_t = (1 - B)^d Z_t$, and

$$(1 - \phi_1 B - \cdots - \phi_p B^p) W_t = \theta_0 + (1 - \theta_1 B - \cdots - \theta_q B^q) a_t,$$

  where $\{a_t\} \sim WN(0, \sigma_a^2)$.

  1. Suppose that the orders p, d, and q are already known by the methods in the previous chapter.
  2. The model will be completely determined if the values of the parameters,
$$\phi_1, ..., \phi_p, \theta_1, ..., \theta_q,\ \theta_0 \text{ and } \sigma_a^2,$$
     are further known.
  3. There is an alternative form of the ARMA model,
$$(1 - \phi_1 B - \cdots - \phi_p B^p)(W_t - \mu) = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t,$$
     where
$$\mu = E(W_t) = \frac{\theta_0}{1 - \phi_1 - \cdots - \phi_p}.$$

- Without loss of generality, in this chapter, we will assume that $\{Z_1, ..., Z_n\}$ comes from a **stationary** and **invertible** ARMA(p, q) process with unknown parameters
    1. $(\phi_1, ..., \phi_p, \theta_1, ..., \theta_q)'$; (dependence structure)
    2. μ or θ_0; (constant mean)
    3. σ_a^2. (variation).

  The aim of this chapter is to estimate these unknown parameters based on the observed $\{Z_1, ..., Z_n\}$.

§5.1 Content

- The method of moments.
- Conditional least squares (CLS).
- Maximum likelihood (ML) and unconditional least square (ULS).
- Properties of the estimates.

§5.2 The method of moments

- The method of moments consists of equating sample moments to corresponding theoretical moments, and solving the resulting equations to obtain estimates of any unknown parameters.
  This method is frequently one of the easiest methods for obtaining parameter estimates.

- For example, suppose that $X_1, ..., X_n \sim i.i.d.\ N(\mu, \sigma^2)$. We want to estimate parameters μ and σ^2.
  1. Equating the sample moments,

     $$\widehat{M}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i,\ \widehat{M}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \text{ and } \widehat{M}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \text{ with } k \geq 3,$$

     to corresponding theoretical moments,

     $$M_1 = \mathrm{E}(X),\ M_2 = \mathrm{E}(X^2) \text{ and } M_k = \mathrm{E}(X^k) \text{ with } k \geq 3.$$

  2. Based on the first two order moments, it holds that $\mu = M_1$ and $\sigma^2 = M_2 - M_1^2$. Then we can get the estimates, $\widehat{\mu} = \widehat{M}_1$ and $\widehat{\sigma}^2 = \widehat{M}_2 - (\widehat{M}_1)^2$.
  3. Note that the third order moment $M_3 = E(X^3) = 3\mu\sigma^2 + \mu^3$. Hence, we can get another estimates, $\widehat{\mu} = \widehat{M}_1$ and $\widehat{\sigma}^2 = (\widehat{M}_3 - (\widehat{M}_1)^3)/(3\widehat{M}_1)$.

     The method of moment estimate (MME) is not unique!

- The method of moments for parameters in ARMA models:
  1. Equating the sample moments,

     $$\widehat{\mu} = \frac{1}{n}\sum_{t=1}^{n} Z_t, \ \widehat{\gamma}_0 = \frac{1}{n-1}\sum_{t=1}^{n}(Z_t - \bar{Z})^2$$

     and

     $$\hat{\rho}_k = \frac{\sum_{t=k+1}^{n}(Z_t - \overline{Z})(Z_{t-k} - \overline{Z})}{\sum_{t=1}^{n}(Z_t - \overline{Z})^2} \text{ with } k \geq 1,$$

     to corresponding theoretical moments,

     $$\mu = \mathrm{E}(Z_t), \ \gamma_0 = \mathrm{var}(Z_t) \text{ and } \rho_k = \mathrm{corr}(Z_t, Z_{t-k}) \text{ with } k \geq 1.$$

  2. Using some resulting equations to obtain estimates of any unknown parameters.

- Consider an AR(p) model,

$$(Z_t - \mu) = \phi_1(Z_{t-1} - \mu) + \cdots + \phi_p(Z_{t-p} - \mu) + a_t,$$

where $\mu = E(Z_t)$.

1. The MME for μ is

$$\widehat{\mu} = \overline{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t.$$

2. **Review of the calculation of ACFs and ACVFs for AR(p) models:**
   - Unique feature:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p}, \qquad k \neq 0.$$

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma_a^2.$$

   - When $k = 1, \cdots, p$,

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix}.$$

   - the variance

$$\gamma_0 = \frac{\sigma_a^2}{1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p}.$$

3. The MME for $\phi_1, ..., \phi_p$ are

$$\begin{pmatrix} \widehat{\phi}_1 \\ \widehat{\phi}_2 \\ \vdots \\ \widehat{\phi}_p \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{pmatrix}.$$

4. The MME for σ_a^2 is

$$\widehat{\sigma}_a^2 = \widehat{\gamma}_0(1 - \widehat{\phi}_1\hat{\rho}_1 - \cdots - \widehat{\phi}_p\hat{\rho}_p).$$

- **Exercise 1:** From a series of length 100, we have computed $\hat{\rho}_1 = 0.8$, $\hat{\rho}_2 = 0.5$, $\hat{\rho}_3 = 0.4$, $\overline{Z} = 2$, and a sample variance of 5. If we assume that an AR(2) model with a constant term is appropriate, how can we get the method of moment estimates of ϕ_1, ϕ_2, θ_0, and σ_a^2?

  Sol: Obtain $\hat{\phi}_1$ and $\hat{\phi}_2$ by Yule Walker equation (sample version), then obtain $\hat{\sigma}_a^2$. Finally, $\hat{\theta}_0 = \hat{\mu}(1 - \hat{\phi}_1 - \cdots - \hat{\phi}_p)$

- Consider an MA(1) process, $Z_t = \theta_0 + a_t - \theta a_{t-1}$, remember that

$$\rho_1 = -\frac{\theta}{1+\theta^2}.$$

1. Equating ρ_1 to $\hat{\rho}_1$, we need to solve a quadratic equation $\hat{\rho}_1\theta^2 + \theta + \hat{\rho}_1 = 0$. If $|\hat{\rho}_1| < 0.5$, then the two real roots are

$$\frac{-1 \pm \sqrt{1-4\hat{\rho}_1^2}}{2\hat{\rho}_1} = -\frac{1}{2\hat{\rho}_1} \pm \left(\frac{1}{4\hat{\rho}_1^2} - 1\right)^{1/2}.$$

   As can easily be checked, the product of the two solutions is always equal to 1; therefore, only one of the solutions satisfies the invertibility condition $|\theta| < 1$.

2. To guarantee invertible, $|\rho_1| < 0.5$, and

$$\widehat{\theta} = \frac{-1 + \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1}.$$

3. **Remarks:**
   - If $\hat{\rho}_1 = \pm 0.5$, unique, real solutions exist, namely ± 1, but neither is invertible.
   - If $|\hat{\rho}_1| > 0.5$, which is possible even though $|\rho_1| < 0.5$, no real solutions exist, and so the method of moments fails to yield a reasonable estimator of θ.
4. The MME for θ_0 is $\widehat{\theta}_0 = \overline{Z} = n^{-1} \sum_{t=1}^{n} Z_t$ since $\theta_0 = \mathrm{E}(Z_t)$. The MME for σ_a^2 is

$$\widehat{\sigma}_a^2 = \frac{\widehat{\gamma}_0}{1 + \widehat{\theta}^2},$$

since $\gamma_0 = (1 + \theta^2)\sigma_a^2$.

- For higher order MA models, the method of moments quickly gets **complicated**. The method of moments for MA models generally produces **poor** estimates.
- **Exercise 2:** The sample ACF for a series and its first difference are given in the following table. Here $n = 100$. The sample mean is $\bar{Z} = 0$ and the sample variance $S_Z^2 = 10$.

lag	1	2	3	4	5	6
ACF for Z_t	0.97	0.97	0.93	0.85	0.80	0.71
ACF for ∇Z_t	-0.42	0.18	-0.02	0.07	-0.10	-0.09

Based on this information alone, identify a suitable ARIMA model for Z_t, and find the method of moment estimates.
Solution: This should be ARIMA(0,1,1). However, if we treat it as ARIMA(0,1,2),then

$$
\begin{aligned}
\mathrm{r}_1 &= -\theta_1 {\sigma_a}^2 - \theta_1\theta_2 {\sigma_a}^2 \\
\mathrm{r}_2 &= -\theta_2 {\sigma_a}^2 \\
\sigma_Z^2 &= (1 + \theta_1^2 + \theta_2^2)\sigma_a^2
\end{aligned}
$$

- Consider the general ARMA(p, q) model,

$$\begin{aligned} Z_t = & \theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t \\ & - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}, \end{aligned}$$

where $\{a_t\} \sim WN(0, \sigma_a^2)$. The situation is more complicated!

1. However, for $k > q$, we have Yule-Walker type equations. So let $k = q+1, ..., q+p$, we get the MME for $\phi_1, ..., \phi_p$ by

$$\begin{pmatrix} \widehat{\phi}_1 \\ \widehat{\phi}_2 \\ \vdots \\ \widehat{\phi}_p \end{pmatrix} = \begin{pmatrix} \hat{\rho}_q & \hat{\rho}_{q-1} & \cdots & \hat{\rho}_{q-p+1} \\ \hat{\rho}_{q+1} & \hat{\rho}_q & \cdots & \hat{\rho}_{q-p+2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{q+p-1} & \hat{\rho}_{q+p-2} & \cdots & \hat{\rho}_q \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_{q+1} \\ \hat{\rho}_{q+2} \\ \vdots \\ \hat{\rho}_{q+p} \end{pmatrix}.$$

2. The MME for $\mu = \mathrm{E}(Z_t)$ is $\widehat{\mu} = \overline{Z} = n^{-1} \sum_{t=1}^{n} Z_t$.

§5.2 Conditional least squares (CLS)

- **The least squares estimate for linear regression:** For a regression model,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} + \varepsilon_i,$$

where $\{\varepsilon_i\}$ is an $i.i.d.$ sequence with mean zero and variance σ^2, the least squares estimate are

$$(\widehat{\beta}_0, \widehat{\beta}_1, ..., \widehat{\beta}_p) = \text{argmin} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{1i} - \cdots - \beta_p X_{pi})^2$$

and

$$\widehat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1i} - \cdots - \widehat{\beta}_p X_{pi})^2.$$

- Consider an AR(p) model,

$$Z_t = \theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t.$$

The **conditional least squares** (CLS) estimates are

$$(\widehat{\theta}_0, \widehat{\phi}_1, ..., \widehat{\phi}_p) = \text{argmin } S_C(\theta_0, \phi_1, ..., \phi_p)$$

and

$$\widehat{\sigma}_a^2 = \frac{1}{n-p} S_C(\widehat{\theta}_0, \widehat{\phi}_1, ..., \widehat{\phi}_p),$$

where

$$S_C(\theta_0, \phi_1, ..., \phi_p) = \sum_{t=p+1}^{n} (Z_t - \theta_0 - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p})^2$$

is called the **conditional sum-of-squares function**.

- Consider an MA(1) process, $Z_t = a_t - \theta a_{t-1}$, the conditional sum-of-squares function is

$$S_C(\theta) = \sum_{t=1}^{n} (a_t)^2.$$

1. By the invertibility, we have that

   $$a_t = Z_t + \theta Z_{t-1} + \theta^2 Z_{t-2} + \cdots .$$

   To continue the estimating, we need set the initial values $0 = Z_0 = Z_{-1} = Z_{-2} = \cdots$, and then, **conditional on** $0 = Z_0 = Z_{-1} = Z_{-2} = \cdots$,

   $$S_C(\theta) = \sum_{t=1}^{n} (Z_t + \theta Z_{t-1} + \theta^2 Z_{t-2} + \cdots + \theta^{t-1} Z_1)^2$$

   can be used to calculate the estimate of θ.

2. Note that, if $a_0 = 0$, then

   $$a_1 = Z_1, \quad a_2 = Z_2 + \theta a_1, \quad a_3 = Z_3 + \theta a_2, \quad \cdots \quad , a_n = Z_n + \theta a_{n-1}.$$

   Conditional on $a_0 = 0$, $S_C(\theta)$ can be used to calculate the estimate of θ.
3. The above two conditions reach the same result!
4. It is impossible to directly calculate the conditional LS estimate for MA and ARMA models. Some numerical optimization algorithm, such as Gauss-Newton and Nelder-Mead, are usually used to search the estimates.

- **Disadvantages:** For series of moderate length and also for stochastic seasonal models, these initial values will have a more pronounced effect on the final estimates for the parameters.

§5.2 Maximum likelihood & unconditional least square

- **Maximum likelihood (ML) estimate:**
  - For any set of observations $Z_1, Z_2, \cdots, Z_n$ (time series or otherwise), the **likelihood function** L is defined to be the probability (density) of obtaining the data actually observed; however, it is considered a function of the parameters in the model. For example, $L = \Pi_{t=1}^n f_\alpha(Z_t)$ if $\{Z_t\}$ is $i.i.d.$ with density $f_\alpha(z)$ and parameter(s) α.
  - The **ML estimates** are then defined as those values of the parameters for which the data actually observed are *most likely*, that is, the values that maximize the likelihood function.
  - **Advantages:** (1) All of the information in the data is used rather than just the first and second moments, as is the case with least squares. (2) Many large-sample results are known under very general conditions.
  - **Disadvantages:** The joint probability function might be complicated and it may be hard to find the optimizer.

- Example: Suppose X is a random variable from exponential distribution with parameter λ, i.e. $f_X(x) = \lambda e^{-\lambda x}$, then what's the MLE of λ?

  Sol: $L(\lambda) = \lambda^n e^{-\lambda(x_1 + \cdots x_n)}$, $\log L = n \log(\lambda) - \lambda \sum_{i=1}^n x_i$ and $\hat{\lambda} = n / \sum_{i=1}^n x_i$

- Consider the AR(1) model

$$Z_t - \mu = \phi(Z_{t-1} - \mu) + a_t,$$

where the white noise $\{a_t\} \sim i.i.d.N(0, \sigma_a^2)$, and the unknown parameters include μ, ϕ and σ_a^2.

1. The random variable Z_t, conditional on $\{Z_{t-1}, Z_{t-2}, ...\}$, will follow the normal distribution $N\{\mu + \phi(Z_{t-1} - \mu), \sigma_a^2\}$, and hence has the density

$$\begin{aligned} f(z|Z_{t-1}, Z_{t-2}, ...) &= f(z|Z_{t-1}) \\ &= (2\pi\sigma_a^2)^{-1/2} \exp\left\{-\frac{[z - \mu - \phi(Z_{t-1} - \mu)]^2}{2\sigma_a^2}\right\}. \end{aligned}$$

2. Note that the AR(1) process also has the MA representation $Z_t = \mu + \sum_{j=0}^{\infty} \phi^j a_{t-j}$, i.e. Z_t is normally distributed with mean μ and variance $\gamma_0 = \sigma_a^2/(1-\phi^2)$.

   Then the random variable Z_1 has the density

$$f(z) = [\frac{2\pi\sigma_a^2}{1-\phi^2}]^{-1/2} \exp\left\{-\frac{(z-\mu)^2}{2\sigma_a^2/(1-\phi^2)}\right\}.$$

3. The likelihood function for $Z_1, ..., Z_n$ is

$$\begin{aligned} L(\phi, \mu, \sigma_a^2) &= f(z_n, z_{n-1}, ..., z_1) = f(z_n|z_{n-1}, ..., z_1) f(z_{n-1}, ..., z_1) \\ &= f(z_n|z_{n-1}) f(z_{n-1}, ..., z_1) = \cdots \\ &= f(z_n|z_{n-1}) f(z_{n-1}|z_{n-2}) \cdot f(z_2|z_1) f(z_1) \\ &= (2\pi\sigma_a^2)^{-n/2}(1-\phi^2)^{1/2} \exp\left\{ -\frac{1}{2\sigma_a^2} S(\phi, \mu) \right\}, \end{aligned}$$

where

$$\begin{aligned} S(\phi, \mu) &= \sum_{t=2}^{n} [(Z_t - \mu) - \phi(Z_{t-1} - \mu)]^2 + (1-\phi^2)(Z_1 - \mu)^2 \\ &= S_C(\phi, \mu) + (1-\phi^2)(Z_1 - \mu)^2 \end{aligned}$$

is called the **unconditional sum-of-squares function**.

4. The ML estimates for μ, ϕ and σ_a^2 will minimize the log-likelihood function

$$-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_a^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma_a^2}S(\phi,\mu)$$

5. As a compromise between conditional least squares (CLS) estimates and full maximum likelihood (ML) estimates, the **unconditional least squares (ULS) estimates** are

$$(\widehat{\phi},\widehat{\mu}) = \operatorname{argmin} S(\phi,\mu) = \operatorname{argmin}\{S_C(\phi,\mu)+(1-\phi^2)(Z_1-\mu)^2\}$$

   and

$$\widehat{\sigma}_a^2 = \frac{1}{n-1}S(\widehat{\phi},\widehat{\mu}).$$

- For the general ARMA(p,q) model, the situation is complicated, and you can refer to Brockwell and Davis (1991) for details.

§5.5 Properties of the estimates

- The CLS, ULS and ML estimates have the same large-sample properties.
    1. When estimating an AR(1) process with the AR(2) model, the variance of the estimates will increase. It is the same for the MA models.
    2. For the ARMA(1,1) process with the AR coefficient very close to the MA coefficient, the variance may be very large.
- The method of moment estimates have the same large-sample properties as the CLS, ULS and ML estimates for the AR models, however, are worse for the MA and ARMA models.

Reference

Please read Chapter 7 of Cryer & Chan (2008).