2022 Fall

IE 313 Time Series Analysis

7. Parameter Estimation

Yongjae Lee
Department of Industrial Engineering
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Chapter 7. Parameter Estimation

- **7.1 The Method of Moments**
- **7.2 Least Squares Estimation**
- **7.3 Maximum Likelihood and Unconditional Least Squares**
- **7.4 Properties of the Estimates**
- 7.5 Illustrations of Parameter Estimation
- 7.6 Bootstrapping ARIMA Models

Parameter estimation

- Once we have specified the orders of ARIMA(p,d,q) model for our observed time series $Y_1, Y_2, \dots, Y_n$, we need to estimate its parameters $\phi_1, \phi_2, \dots, \phi_p$ and $\theta_1, \theta_2, \dots, \theta_q$

- That is, we are going to estimate the parameters of ARMA(p,q) model from the d th difference of the observed series

- Therefore, we will simply assume in this chapter that $Y_1, Y_2, \dots, Y_n$ denote our *stationary* process
  - It may be an appropriate difference of the original series

Chapter 7.1

The Method of Moments

Method of moments

- The method of moments is frequently one of the easiest methods for obtaining parameter estimates
  - It consists of
    - Equating sample moments to corresponding theoretical moments
    - Solving the resulting equations to obtain estimates of any unknown parameters

Autoregressive models

- AR(1): $Y_t = \phi Y_{t-1} + e_t$
  - For this, we know from Chapter 4 that $\rho_k = \phi^k$
  - Therefore,
$$\hat{\phi} = \hat{\rho}_1 = r_1$$
    - r_1: lag 1 sample ACF

- AR(2): $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$
  - From the Yule-Walker equations (see page 72),
$$\rho_1 = \phi_1 + \rho_1\phi_2 \text{ and } \rho_2 = \rho_1\phi_1 + \phi_2$$
  - Therefore,
$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2} \quad \text{and} \quad \hat{\phi}_2 = \frac{r_2-r_1^2}{1-r_1^2}$$

Autoregressive models

- AR(p): $Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t$
  - Similarly, we can use the Yule-Walker equation (see page 114)

$$\left.\begin{array}{rrrrrcl}
\phi_1 + & r_1\phi_2 + & r_2\phi_3 & + \cdots + & r_{p-1}\phi_p & = & r_1 \\
r_1\phi_1 + & \phi_2 + & r_1\phi_3 & + \cdots + & r_{p-2}\phi_p & = & r_2 \\
 & & & & & \vdots & \\
r_{p-1}\phi_1 + & r_{p-2}\phi_2 + & r_{p-3}\phi_3 & + \cdots + & \phi_p & = & r_p
\end{array}\right\}$$

  - The above linear equation can be solved for $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$

- The estimates obtained in this way are also called Yule-Walker estimates

Moving average models

- Surprisingly, the method of moments is not nearly as convenient when applied to MA models

- Consider MA(1) case: $Y_t = e_t - \theta e_{t-1}$
  - We know from Chapter 4 (see page 57) that
$$\rho_1 = -\frac{\theta}{1+\theta^2}$$
  - Equating ρ_1 to r_1, we need to solve a quadratic equation in θ

Chapter 7.2

Least Squares Estimation

Autoregressive models

- AR(1) with nonzero mean: $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$
  - We can view this as a regression model with predictor (or independent) variable Y_{t-1} and response (or dependent) variable Y_t
  - Then, least squares estimation would proceed by minimizing the sum of squares of the differences

$$(Y_t - \mu) - \phi(Y_{t-1} - \mu)$$

  - Since only $Y_1, Y_2, \ldots, Y_n$ are observed, we can only sum from $t = 2$ to $t = n$

$$S_c(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2$$

    - This is usually called the **conditional sum-of-squares function**

Autoregressive models

- AR(1) with nonzero mean: $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$
  - Then, we want to find ϕ and μ that minimize $S_c(\phi, \mu)$
  - First, consider the equation
    $\frac{\partial S_c}{\partial \mu} = \sum_{t=2}^{n} 2[(Y_t - \mu) - \phi(Y_{t-1} - \mu)](-1 + \phi) = 0$
    - If we solve it for μ,
      $\mu = \frac{1}{(n-1)(1-\phi)}[\sum_{t=2}^{n} Y_t - \phi \sum_{t=2}^{n} Y_{t-1}]$
    - For large n,
      $\frac{1}{n-1}\sum_{t=2}^{n} Y_t \approx \frac{1}{n-1}\sum_{t=2}^{n} Y_{t-1} \approx \bar{Y}$
    - Thus,
      $\hat{\mu} \approx \frac{1}{1-\phi}(\bar{Y} - \phi\bar{Y}) = \bar{Y}$

Autoregressive models

- AR(1) with nonzero mean: $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$
  - Then, we want to find ϕ and μ that minimize $S_c(\phi, \mu)$
  - Similarly, consider the equation
    $\frac{\partial S_c(\phi,\bar{Y})}{\partial \phi} = \sum_{t=2}^{n} 2[(Y_t - \bar{Y}) - \phi(Y_{t-1} - \bar{Y})](Y_{t-1} - \bar{Y}) = 0$
    - If we solve it for ϕ,
      $\hat{\phi} = \frac{\sum_{t=2}^{n}[(Y_t-\bar{Y})(Y_{t-1}-\bar{Y})]}{\sum_{t=2}^{n}(Y_{t-1}-\bar{Y})^2}$
    - We can see that this is almost the same as sample ACF at lag 1 (r_1) except that we are missing $(Y_t - \bar{Y})^2$ in the denominator
    - This is negligible for stationary processes, and thus, the least equares and method-of-moments estimators are nearly identical, especially for large samples

Moving average models

- MA(1): $Y_t = e_t - \theta e_{t-1}$
  - At first, it is not apparent how a least squares (or regression) method can be applied to such models
  - However, recall from Chapter 4 (page 77) that invertible MA(1) models can be rewritten as

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

  - Then, least squares can be carried out by choosing θ that minimizes

$$S_c(\theta) = \sum e_t^2 = \sum[Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \theta^3 Y_{t-3} + \cdots]^2$$

    - But we will not be able to minimize $S_c(\theta)$ by taking a derivative with respect to θ, setting it to zero, and solving
    - Because calculating $S_c(\theta)$ is not obvious

Moving average models

- MA(1): $Y_t = e_t - \theta e_{t-1}$
  - Then, consider evaluating $S_c(\theta)$ for a single given value of θ
    - Noting that the only Y's we have available are $Y_1, Y_2, \dots, Y_n$, use the equation $e_t = Y_t + \theta e_{t-1}$ conditional on $e_0 = 0$

$$\begin{aligned} e_1 &= Y_1 \\ e_2 &= Y_2 + \theta e_1 \\ e_3 &= Y_3 + \theta e_2 \\ &\vdots \\ e_n &= Y_n + \theta e_{n-1} \end{aligned}$$

  - Now we can minimize $S_c(\theta)$ by trying many different values of θ within the invertible rage $(-1, +1)$
    - For more general MA(q) models, a numerical optimization algorithm, such as Gauss-Newton or Nelter-Mead, will be needed

Mixed models

- ARMA(1,1): $Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}$
  - As in pure MA case, we use $e_t = Y_t - \phi Y_{t-1} + \theta e_{t-1}$ to calculate $S_c(\phi, \theta) = \sum e_t^2$
  - However, even if we set $e_0 = 0$, we still need an additional "startup" problem, namely Y_0
    - One approach is to set $Y_0 = 0$ or to $\bar{Y}$
    - A better approach is to begin recursion at $t = 2$ to minimize

$$S_c(\phi, \theta) = \sum_{t=2}^{n} e_t^2$$

      - For general ARMA(p,q) model, we need to compute $e_t = Y_t - \phi_1 Y_{t-1} - \cdots - -\phi_p Y_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}$ with $e_p = e_{p-1} = \cdots = e_{p+1-q} = 0$ and minimize $S_c\left(\phi_1, \dots, \phi_1, \theta_1, \dots, \theta_p\right)$

Chapter 7.3

Maximum Likelihood and Unconditional Least Squares

Maximum likelihood

- For series of moderate length, the startup values $e_p = e_{p-1} = \cdots = e_{p+1-q} = 0$ will have a more pronounced effect on the final estimates for the parameters
  - Hence, we need to consider the more difficult problem of **maximum likelihood estimation**
  - The advantage of maximum likelihood method is that all of the information in the data is used
    - Rather than just the first and second moments as in the case with least squares
  - Another advantage is that many large-sample results are known under very general conditions
  - One disadvantage is that we must for the first time work specifically with the **joint probability density function** of the process

Maximum likelihood estimation

- For any set of observations, $Y_1, Y_2, \ldots, Y_n$, time series or not, **the likelihood function L** is defined to be the joint probability density of obtaining the data actually observed

- However, it is considered as a function of the unknown parameters in the model with the observed data held fixed
  - For ARIMA models, L will be a function of the ϕ's, θ's, μ, and σ_e^2 given the observations $Y_1, Y_2, \ldots, Y_n$

- **The maximum likelihood estmators** are 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

Autoregressive models

- AR(1) with nonzero mean: $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$
  - The most common assumption is the white noise terms are independent, normally distributed random variables with zero means and common standard deviation σ_e

    - The probability density function (pdf) of each e_t is

$$(2\pi\sigma_e^2)^{-\frac{1}{2}} \exp\left(-\frac{e_t^2}{2\sigma_e^2}\right) \text{ for } -\infty < e_t < \infty$$

Normal density
with mean μ and standard deviation σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- And by independence, the joint pdf for $e_2, e_3, \dots, e_n$ is

$$(2\pi\sigma_e^2)^{-\frac{n-1}{2}} \exp\left(-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n} e_t^2\right)$$

Autoregressive models

Joint pdf for $e_2, e_3, \dots, e_n$

$$(2\pi\sigma_e^2)^{-\frac{n-1}{2}} \exp\left(-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n} e_t^2\right)$$

- AR(1) with nonzero mean: $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$
  - Now consider

$$Y_2 - \mu = \phi(Y_1 - \mu) + e_2$$
$$Y_3 - \mu = \phi(Y_2 - \mu) + e_3$$
$$\vdots$$
$$Y_n - \mu = \phi(Y_{n-1} - \mu) + e_n$$

  - Then, the joint pdf of $Y_2, Y_3, \dots, Y_n$ conditioning on $Y_1 = y_1$ as

$$f(y_2, y_3, \dots, y_n | y_1)$$
$$= (2\pi\sigma_e^2)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n}[(y_t - \mu) - \phi(y_{t-1} - \mu)]^2\right\}$$

  - From the linear process representation of AR(1) (see page 70),

$$Y_1 \sim N(\mu, \frac{\sigma_e^2}{1-\phi^2})$$

Autoregressive models

Axiom of probability
$P(A \cap B) = P(B \mid A)\, P(A)$

- AR(1) with nonzero mean: $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$
  - Then, by multiplying the marginal pdf of Y_1 ($\mathrm{P}(Y_1)$)to the joint pdf of $Y_2, Y_3, \dots, Y_n$ given Y_1 ($\mathrm{P}(Y_2, \dots, Y_n | Y_1)$) gives the joing pdf of $Y_1, Y_2, \dots, Y_n$.
  - Interpreted as a function of the parameters ϕ, μ, and σ_e^2, the likelihood function for an AR(1) model is given by

$$L(\phi, \mu, \sigma_e^2) = (2\pi\sigma_e^2)^{-\frac{n}{2}}(1 - \phi^2)^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_e^2} S(\phi, \mu)\right\}$$

    - Where $S(\phi, \mu) = \sum_{t=2}^{n}[(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1 - \phi^2)(Y_1 - \mu)$, which is called the **unconditional sum-of-squares function**

Autoregressive models

Axiom of probability
$P(A \cap B) = P(B \mid A)\, P(A)$

- AR(1) with nonzero mean: $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$
  - In general, the log of the likelihood function (**log-likelihood function**) is more convenient to work with

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

  - Then, find ϕ, μ, σ_e^2 that maximize the log-likelihood

Chapter 7.4

Properties of the Estimates

Properties of estimates

- We will just briefly see the large-sample properties of the maximum likelihood and least squares estimators

$$\text{AR}(1):\ Var(\hat{\phi}) \approx \frac{1-\phi^2}{n}$$

$$\text{AR}(2): \begin{cases} Var(\hat{\phi}_1) \approx Var(\hat{\phi}_2) \approx \dfrac{1-\phi_2^2}{n} \\ Corr(\hat{\phi}_1, \hat{\phi}_2) \approx -\dfrac{\phi_1}{1-\phi_2} = -\rho_1 \end{cases}$$

$$\text{MA}(1):\ Var(\hat{\theta}) \approx \frac{1-\theta^2}{n}$$

$$\text{MA}(2): \begin{cases} Var(\hat{\theta}_1) \approx Var(\hat{\theta}_2) \approx \dfrac{1-\theta_2^2}{n} \\ Corr(\hat{\theta}_1, \hat{\theta}_2) \approx -\dfrac{\theta_1}{1-\theta_2} \end{cases}$$

$$\text{ARMA}(1,1): \begin{cases} Var(\hat{\phi}) \approx \left[\dfrac{1-\phi^2}{n}\right]\left[\dfrac{1-\phi\theta}{\phi-\theta}\right]^2 \\ Var(\hat{\theta}) \approx \left[\dfrac{1-\theta^2}{n}\right]\left[\dfrac{1-\phi\theta}{\phi-\theta}\right]^2 \\ Corr(\hat{\phi}, \hat{\theta}) \approx \dfrac{\sqrt{(1-\phi^2)(1-\theta^2)}}{1-\phi\theta} \end{cases}$$

  - AR(1): $Var(\hat{\phi})$ decreases as ϕ approaches ± 1
  - AR(2): estimation of ϕ_1 will suffer if we erroneously fit ϕ_2
    - Similar arguments can be made for MA(2) and ARMA(1,1)