Time Series Analysis

Parameter Estimation

Hossein Moradi Rekabdarkolaee

South Dakota State University

email: *hossein.moradirekabdarkolaee@sdstate.edu*.

Office: CAME Building, *Room: 240*.

Introduction

- So far, we have learned
  - Different components of time series
    - Trend,
    - Seasonality,
    - Error.
  - Different type of time series
    - Stationary (weak and strong)
    - Non-stationary.
  - Different models for time series
    - Autoregressive,
    - Moving Average,
    - ARMA,
    - ARIMA.

Introduction

- Different way to find the order of time series
  - ACF,
  - PACF,
  - EACF,
  - Information Criteria based method (AIC, BIC, AICc, etc.)
- In this section, we are going to investigate various approach to estimate the parameters of the model we choose to fit the time series.
- These approaches are
  - Method of Moment,
  - Least Squares,
  - Maximum Likelihood.

The Method of Moments

- The method consists of equating sample moments to corresponding theoretical moments and solving the resulting equations to obtain estimates of any unknown parameters.
- The simplest example of the method is to estimate a stationary process mean by a sample mean.

$$\mu_1 = E[X], \qquad m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \Rightarrow \quad \widehat{\mu}_1 = \bar{X}$$

- Another example

$$\mu_1 = E[X] = \mu, \qquad m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\mu_2 = E[X^2] = \sigma^2 + \mu^2, \qquad m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2,$$

$$\Rightarrow \mu_1 = m_1 \quad \Rightarrow \quad \widehat{\mu} = \bar{X}$$

$$\Rightarrow \mu_2 = m_2 \ \Rightarrow \sigma^2 + \mu^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

The Method of Moments for AR

- Consider AR(1)
$$Y_t = \phi Y_{t-1} + \epsilon_t$$
- For this process, $\rho_1 = \phi$.
- The lag 1 sample autocorrelation r_1 is a good estimation for ρ_1 which means we can use it as an estimator for ϕ i.e., $\hat{\phi} = r_1$.
- For AR(2), using the Yule-Walker equation from chapter 4, we have
$$\rho_1 = \phi_1 + \rho_1\phi_2, \quad \text{and} \quad \rho_2 = \rho_1\phi_1 + \phi_2$$
- Replacing ρ_1 and ρ_2 with r_1 and r_2 and solving the system of equation, we have
$$\begin{aligned} \hat{\phi}_1 &= \frac{r_1(1 - r_2)}{1 - r_1^2}, \\ \hat{\phi}_1 &= \frac{r_2 - r_1^2}{1 - r_1^2} \end{aligned}$$

The Method of Moments for AR

- For AR(p) case, we can again use the Yule-Walker equations

$$\begin{cases} \phi_1 + r_1\phi_2 + r_2\phi_3 + \ldots + r_{p-1}\phi_p = r_1 \\ r_1\phi_1 + \phi_2 + r_1\phi_3 + \ldots + r_{p-2}\phi_p = r_2 \\ \vdots \\ r_{p-1}\phi_1 + r_{p-2}\phi_2 + r_{p-3}\phi_3 + \ldots + \phi_p = r_p \end{cases}$$

- These linear equations are then solved for $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$.
- We will use Durbin-Levinson recursion algorithm to get estimates for these parameters.
- The DL algorithm is sensitive if solutions are close to the boundary of the stationarity region.
- The estimates obtained from this process are called Yule-Walker estimates.

The Method of Moments for MA

- Unlike the AR process, the method of moment estimation is not as convenient for the MA process.
- For MA(1), we know that

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

- Replacing ρ_1 with r_1, then we can solve the quadratic equation.
- If we have $|r_1|$ larger than 0.5, then the real solution does not exist.
- For $|r_1|$ less than 0.5, the estimate for the invertible MA parameter is

$$\hat{\theta} = \frac{-1+\sqrt{1-4r_1^2}}{2r_1}$$

The Method of Moments for MA

- For $|r_1|$ equal to 0.5, real solusion is $|1|$but the MA process is not invertible.
- This process is messy for higher orders of MA.

The Method of Moments for ARMA

- For ARMA(1,1), we have

$$\rho_k = \frac{(1-\theta\phi)(\phi-\theta)}{1-2\theta\phi+\theta^2}\phi^{k-1}, \quad \text{for } k \geq 1.$$

- Note that $\phi = \frac{\rho_2}{\rho_1}$, we can estimate ϕ as

$$\hat{\phi} = \frac{r_2}{r_1},$$

- Therefore, we can use

$$r_1 = \frac{(1-\theta\hat{\phi})(\hat{\phi}-\theta)}{1-2\theta\hat{\phi}+\theta^2}$$

and solve to obtain estimation for θ.

The Method of Moments for ARMA

- The last parameter for the model is the variance of the noise.
- The sample variance is

$$s^2 = \frac{1}{n-1}\sum_{t=1}^{n}(Y_t - \bar{Y})^2.$$

- For AR(1) process, we have

$$\hat{\sigma}^2 = (1 - r_1^2)s^2$$

- For the AR(p) process

$$\hat{\sigma}^2 = (1 - \hat{\phi}_1 r_1^2 - \hat{\phi}_2 r_2^2 - \ldots - \hat{\phi}_p r_p^2)s^2$$

The Method of Moments for ARMA

- For the MA(q) process, we have

$$\hat{\sigma}^2 = \frac{s^2}{1 + \hat{\theta}_1^2 + \hat{\theta}_2^2 \ldots + \hat{\theta}_q^2}.$$

- For the ARMA(1,1) process

$$\hat{\sigma}^2 = \frac{1 - \hat{\phi}^2}{1 - 2\hat{\phi}\hat{\theta} + \hat{\theta}^2} s^2.$$

Least Squares Estimation

- Method of Moment is unsatisfactory for many models.
- Thus, we need another method.
- Consider AR(1) with constant mean

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

- We can use regression with regressing Y_t on Y_{t-1}.
- Thus, we want to minimize the sum of squares of the difference $(Y_t - \mu) - \phi(Y_{t-1} - \mu)$ i.e.

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

- This is often called “conditional sum-of-squares function”.

Least Squares Estimation

- Now, we have to take the derivative of S_c and set it equal to 0.
- We have

$$\frac{\partial S_c}{\partial \mu} = 2\sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]\,(-1 + \phi) = 0$$

- Thus,

$$\mu = \frac{1}{(n-1)(1-\phi)}\left[\sum_{t=2}^{n} Y_t - \phi \sum_{t=2}^{n} Y_{t-1}\right].$$

- Thus, for large n, the estimate for the mean is

$$\hat{\mu} \approx \frac{1}{1-\phi}(\bar{Y} - \phi\bar{Y}) = \bar{Y}.$$

Least Squares Estimation

- For ϕ, we do the same thing i.e.

$$\frac{\partial S_c}{\partial \phi} = 2\sum_{t=2}^{n}\left[(Y_t - \bar{Y}) - \phi(Y_{t-1} - \bar{Y})\right](Y_{t-1} - \bar{Y}) = 0$$

- Then, solving ϕ, we have

$$\hat{\phi} = \frac{\sum_{t=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=2}^{n}(Y_{t-1} - \bar{Y})^2}$$

- This means the estimate for the method of moment and least squares are nearly identical.

Least Squares Estimation for AR(p)

- Now consider the general AR(2) process.
- Using the least squares method, we have

$$\hat{\mu} = \bar{Y}$$

- Taking the partial derivative for other parameters, we have

$$\begin{aligned} r_1 &= \phi_1 + r_1\phi_2 \\ r_2 &= r_1\phi_1 + \phi_2 \end{aligned}$$

- This leads to the sample Yule-Walker equations for an AR(2) model.
- The same thing can be extended to the AR(p) process meaning that we can use sample Yule-Walker equations to obtain estimates for the parameter of the AR(p) process.

Least Squares Estimation for MA

- Consider MA(1)

$$Y_t = \epsilon_t - \theta \epsilon_{t-1}$$

- If this is an invertible MA(1), then we can rewrite this model as $AR(\infty)$ i.e.

$$Y_t = -\theta Y_{t-1} - \theta^2 Y_{t-2} - \ldots - \theta^p Y_{p-1} - \ldots + \epsilon_t$$

- Then, the sum of squares can be written as

$$S_c(\theta) = \sum \epsilon_t^2 = \sum \left(Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \ldots + \theta^p Y_{p-1} + \ldots \right)^2$$

- Taking a derivative with respect to θ and setting it to 0 might not be possible.

- Therefore, we use numerical techniques to get an estimate for the θ.

Least Squares Estimation for MA

- For simple cases, we can use a grid search over (-1, +1) to find the minimum sum of squares.
- We can also use numerical techniques such as Gauss-Newton or Nelder-Mead to obtain an estimate for the parameters.
- Using grid search for higher-order MA processes might be computationally challenging.
- Finding the solution for higher-order MA processes even using other numerical techniques could be a computationally challenging task.

Least Squares Estimation for ARMA

- Consider an ARMA(1,1)

$$Y_t = \phi Y_{t-1} + \epsilon_t - \theta \epsilon_{t-1}.$$

- Looking at it as a MA process, let $\epsilon_t = \epsilon_t(\phi, \theta)$ and we want to minimize $S_c(\phi, \theta) = \sum \epsilon_t^2$ where

$$\epsilon_t = Y_t - \phi Y_{t-1} + \theta \epsilon_{t-1}.$$

- We can set Y_0 to be 0 if it is a mean 0 process or the sample mean if it is a non-zero constant mean process.

- Another approach is to begin recursion at $t = 2$ and avoid Y_0 and minimize $S_c(\phi, \theta) = \sum_{t=2}^{n} \epsilon_t^2$.

Least Squares Estimation for ARMA

- For a general ARMA(p,q), we have

$$\epsilon_t = Y_t - \phi_1 Y_{t-1} - \ldots \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \ldots \theta_q \epsilon_{t-q}$$

- Then we minimize $S_c(\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q)$ using numerical techniques.

- If the model is invertible, this minimization is not sensitive to the initial value of $\epsilon_p, \epsilon_{p-1}, \ldots, \epsilon_{p+1-q}$.

Maximum Likelihood and Unconditional Least Squares

- For any set of observations $Y_1, Y_2, \ldots, Y_n$, the likelihood function L is defined to be the joint probability density of obtaining the data actually observed.

- For ARIMA models, L will be a function of the ϕs, θs, μ, and σ^2 given the observed values.

- The MLEs are those values for the parameters that the data actually observed are most likely, that is, the values that maximize the likelihood function.

MLE for AR

- For AR(1)

$$Y_t = \phi Y_{t-1} + \epsilon_t$$

  where $\epsilon_t \sim N(0, \sigma^2)$.

- The pdf for each ϵ_t is

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\epsilon_t^2}{2\sigma^2}\right\} \quad \text{for} \quad -\infty < \epsilon_t < \infty,$$

- Since ϵ_ts are independent, the joint pdf of $\epsilon_2, \epsilon_3, \ldots, \epsilon_n$ is

$$\left(2\pi\sigma^2\right)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=2}^{n} \epsilon_t^2\right\}.$$

MLE for AR

- Define

$$\begin{cases} Y_2 - \mu = \phi(Y_1 - \mu) + \epsilon_2 \\ Y_3 - \mu = \phi(Y_2 - \mu) + \epsilon_2 \\ \vdots \\ Y_n - \mu = \phi(Y_{n-1} - \mu) + \epsilon_n \end{cases}$$

- This is a linear transformation between $\epsilon_2, \epsilon_3, \ldots, \epsilon_n$ and $Y_2, Y_3, \ldots, Y_n$.
- With Jacobian equal to 1, the joint pdf of the $Y_2, Y_3, \ldots, Y_n$ given $Y_1 = y_1$ can be written as

$$\begin{aligned} f(y_2, y_3, \ldots, y_n | y_1) &= \left(2\pi\sigma^2\right)^{-\frac{n-1}{2}} \\ &\times \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=2}^{n}[(y_t - \mu) - \phi(y_{t-1} - \mu)]^2\right\}. \end{aligned}$$

MLE for AR

- Now consider the (marginal) distribution of Y_1.
- It follows from the linear process representation of the AR(1) process that $Y_1 \sim N\left(\mu, \frac{\sigma^2}{1-\phi^2}\right)$.
- Now, multiplying the conditional pdf by this marginal gives us the following joint pdf:

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

where

$$S(\phi,\mu)=\sum_{t=2}^{n}[(Y_t-\mu)-\phi(Y_{t-1}-\mu)]^2+(1-\phi^2)(Y_1-\mu),$$

is called the unconditional sum-of-squares function.

MLE for AR

- The log-likelihood function can be written as

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

- For a given μ and ϕ, we can take the derivative $\ell(\mu, \phi, \sigma^2)$ with respect to σ^2 and set that equal to 0 to maximize the log-likelihood function.
- This estimate is

$$\hat{\sigma}^2 = \frac{S(\hat{\mu}, \hat{\phi})}{n}.$$

- We usually divide by $n - 2$ rather than n to obtain an estimator with less bias.

MLE for AR

- Recall
$$S(\phi, \mu) = S_c(\phi, \mu) + (1 - \phi^2)(Y_1 - \mu)^2.$$
where $S_c(\phi, \mu)$ involes $n - 1$ component and does not involves $(1 - \phi^2)(Y_1 - \mu)^2$.

- Since $S(\phi, \mu) \approx S_c(\phi, \mu)$, we can obtain estimation for μ and ϕ that minimizes either one of these quantities.

- Using either quantity often provides similar results and the effect of $(1 - \phi^2)(Y_1 - \mu)^2$ is substantial near the stationarity boundary of ± 1.

MLE for AR

- The term $(1 - \phi^2)(Y_1 - \mu)^2$ makes the partial derivative of $S(\mu, \phi)$ with respect to μ and ϕ to be nonlinear in μ and ϕ and minimization should be done numerically.
- This estimate is called the unconditional least squares estimate.
- The derivation of the likelihood function for more general ARMA models is considerably more involved.
- One derivation can be found in Appendix H of the Cryer and Chan (2008),
- More detailed derivation can be found in Brockwell and Davis (1991) and Shumway and Stoffer (2006).

Properties of MLE

- For large n, the estimators for different parameters of the ARMA model are approximately unbiased and follow the normal distribution.

- For AR(1), the variance of $\hat{\phi}$ is approximately $\frac{1-\phi^2}{n}$.

- For AR(2):

$$\begin{cases} var(\hat{\phi}_1) = var(\hat{\phi}_2) \approx \frac{1-\phi_2^2}{n} \\ corr(\hat{\phi}_1, \hat{\phi}_2) \approx \frac{\phi_1}{1-\phi_2} = -\rho_1 \end{cases}$$

Properties of MLE

- For MA(1):

$$var(\hat{\theta}) \approx \frac{1-\theta^2}{n}$$

- For MA(2):

$$\begin{cases} var(\hat{\theta}_1) \approx var(\hat{\theta}_2) \approx \frac{1-\theta_2^2}{n} \\ corr(\hat{\theta}_1, \hat{\theta}_2) \approx -\frac{\theta_1}{1-\theta_2} \end{cases}$$

- For ARMA(1,1):

$$\begin{cases} var(\hat{\phi}) \approx \left[\frac{1-\phi^2}{n}\right] \left[\frac{1-\phi\theta}{\phi-\theta}\right]^2 \\ var(\hat{\theta}) \approx \left[\frac{1-\theta^2}{n}\right] \left[\frac{1-\phi\theta}{\phi-\theta}\right]^2 \\ corr(\hat{\phi}, \hat{\theta}) \approx \frac{\sqrt{(1-\phi^2)(1-\theta^2)}}{1-\phi\theta} \end{cases}$$

Properties of MLE

- It can be seen for the AR(1), the variance $\hat{\phi}$ decreases as ϕ approaches ± 1.
- Although AR(1) is a special case of AR(2), the variance of $\hat{\phi}_1$ will suffers if we fit AR(2) to a true AR(1) model since true $\phi_2 = 0$.
- A similar thing happens for the MA process or if we fit the ARMA model to a true AR or true MA process.
- For the ARMA model if ϕ and θ are close to each other, the variance of their estimates is extremely large.
- Furthermore, the estimate of ϕ and θ are highly correlated.