Time Series Analysis

Maximum Likelihood Estimation

Binzhi Chen

Department of Economics
University of Birmingham

March 4, 2020

Introduction

The previous chapters assumed that the population parameters

$$(c, \phi_1, \phi_2, \ldots, \theta_1, \theta_2, \ldots, \sigma^2)$$

were known and showed how population moments such as $E(Y_t Y_{t-j})$ and linear forecasts $E(Y_{t+s}|Y_t, Y_{t-1} \ldots)$ could be calculated as functions of these population parameters. This chapter explores how to estimate the values of $(c, \phi_1, \phi_2, \ldots, \theta_1, \theta_2, \ldots, \sigma^2)$ on the basis of observations on Y

The primary principle on which estimation will be based is maximum likelihood.

Let $\boldsymbol{\Theta} = (c, \phi_1, \phi_2, \ldots, \theta_1, \theta_2, \ldots, \sigma^2)'$ denote the vector of population parameters. The approach will be to calculate the probability density:

$$f_{Y_T, Y_{T-1}, \ldots, Y_1}(y_T, y_{T-1}, \ldots, y_1; \boldsymbol{\Theta})$$

MLE for an AR(1) Process

Consider an AR(1) process:

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

with $\epsilon_t \sim i.i.d.\ N(0, \sigma^2)$
For this case, the vector of population parameters to be estimated consists of $\mathbf{\Theta} = (c, \phi, \sigma^2)'$

The density of the first observation takes the form:

$$\begin{aligned} f_{Y_1}(y_1; \mathbf{\Theta}) &= f_{Y_1}(y_1; c, \phi, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2/(1-\phi^2)}} exp[\frac{-\{y_1 - c/(1-\phi^2)\}^2}{2\sigma^2/(1-\phi^2)}] \end{aligned} \quad (1)$$

Where $E(Y_1) = c/(1-\phi)$, $Var(Y_1) = \sigma^2/(1-\phi^2)$

MLE for an AR(1) Process

Next consider the distribution of the second observation Y_2 conditional on Y_1:

$$Y_2 = c + \phi Y_1 + \epsilon_2$$

Conditioning on Y_1 means treating Y_1 as if it were a constant. Hence:

$$f_{Y_2|Y_1}(y_2|y_1;\boldsymbol{\Theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[\frac{-(y_2 - c - \phi y_1)^2}{2\sigma^2}]$$

The joint density is:

$$f_{Y_2,Y_1}(y_2, y_1;\boldsymbol{\Theta}) = f_{Y_2|Y_1}(y_2|y_1;\boldsymbol{\Theta}) \cdot f_{Y_1}(y_1;\boldsymbol{\Theta})$$

MLE for an AR(1) Process

Similarly, the distribution of the third observation conditional on the first two is:

$$f_{Y_3|Y_2,Y_1}(y_3|y_2,y_1;\boldsymbol{\Theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[\frac{-(y_3 - c - \phi y_2)^2}{2\sigma^2}]$$

The joint density is:

$$f_{Y_3,Y_2,Y_1}(y_3,y_2,y_1;\boldsymbol{\Theta}) = f_{Y_3|Y_2,Y_1}(y_3|y_2,y_1;\boldsymbol{\Theta}) \cdot f_{Y_2,Y_1}(y_2,y_1;\boldsymbol{\Theta})$$

Hence, the density of observation t conditional on $t-1$ is given by:

$$\begin{aligned} f_{Y_T|Y_{T-1},Y_{T-2},\ldots,Y_1}&(y_t|y_{t-1},y_{t-2}\ldots,y_1;\boldsymbol{\Theta}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} exp[\frac{-(y_t - c - \phi y_{t-1})^2}{2\sigma^2}] \end{aligned} \qquad (2)$$

MLE for an AR(1) Process

Hence, the joint density of observation t is given by:

$$
\begin{aligned}
& f_{Y_t, Y_{t-1}, Y_{t-1}, \ldots, Y_1}(y_t, y_{t-1}, y_{t-2}, \ldots, y_1; \boldsymbol{\Theta}) \\
&= f_{Y_T | Y_{T-1}(y_t | y_{t-1}; \boldsymbol{\Theta})} \cdot f_{Y_{T-1}, Y_{T-2}, \ldots, Y_1}(y_{t-1}, y_{t-2}, y_1; \boldsymbol{\Theta}) \\
&= f_{Y_1}(y_1; \boldsymbol{\Theta}) \cdot \prod_{t=2}^{T} f_{Y_T | Y_{T-1}}(y_t | y_{t-1}; \boldsymbol{\Theta})
\end{aligned}
\tag{3}
$$

In general, we use log likelihood function to produce the result of equation (2):

$$
\mathcal{L}(\boldsymbol{\Theta}) = log\, f_{Y_1}(y_1; \boldsymbol{\Theta}) + \sum_{t=2}^{T} log\, f_{Y_T | Y_{T-1}}(y_t | y_{t-1}; \boldsymbol{\Theta}) \tag{4}
$$

MLE for an AR(1) Process

Substituting equation(1) and equation(2) into equation(3) we get:

$$\mathcal{L}(\boldsymbol{\Theta}) = -\frac{1}{2}log(2\pi) - \frac{1}{2}log[\sigma^2/(1-\phi^2)] - \frac{\{y_1 - [c/(1-\phi)]\}^2}{2\sigma^2/(1-\phi^2)} - [(T-1)/2]log(2\pi) - [(T-1)/2]log(\sigma^2) - \sum_{t=2}^{T}[\frac{(y_t - c - \phi Y_{t-1})^2}{2\sigma^2}] \tag{5}$$

However, one can use, to an approximation, the conditional MLE. Here we treat Y_1 as fixed, and maximise the likelihood in this case.

$$f_{Y_t, Y_{t-1}, Y_{t-2}, \ldots, | Y_1}(y_t, y_{t-1}, y_{t-2}, \ldots, | y_1; \boldsymbol{\Theta}) = \prod_{t=2}^{T} f_{Y_T | Y_{T-1}}(y_t | y_{t-1}; \boldsymbol{\Theta})$$

MLE for an AR(1) Process

Then the conditional log likelihood function would be:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\Theta}) &= log\, f_{Y_t, Y_{t-1}, Y_{t-1}, \ldots, | Y_1}(y_t, y_{t-1}, y_{t-2}, \ldots, | y_1; \boldsymbol{\Theta}) \\
&= -[(T-1)/2] log(2\pi) - [(T-1)/2] log(\sigma^2) \\
&- \sum_{t=2}^{T} [\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}]
\end{aligned}
\tag{6}
$$

MLE for an AR(1) Process

Now, taking first-order differential equation with respect to c, ϕ, σ^2:

$$\frac{\partial \mathcal{L}(c, \phi, \sigma^2)}{\partial c} = \frac{2}{2\sigma^2} \sum_{t=2}^{T} (y_t - c - \phi y_{t-1}) = 0$$

$$c = \bar{y}_t - \phi \bar{y}_{t-1}$$

$$\frac{\partial \mathcal{L}(c, \phi, \sigma^2)}{\partial \phi} = \frac{2}{2\sigma^2} \sum_{t=2}^{T} (y_t - c - \phi y_{t-1})(y_{t-1}) = 0$$

$$\phi = \frac{\sum_{t=2}^{T} (y_t - \bar{y}_t)(y_{t-1} - \bar{y}_{t-1})}{\sum_{t=2}^{T} (y_{t-1} - \bar{y}_{t-1})^2}$$

$$\frac{\partial \mathcal{L}(c, \phi, \sigma^2)}{\partial \sigma^2} = -\frac{T-1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=2}^{T} (y_t - c - \phi y_{t-1}) = 0$$

$$\sigma^2 = \frac{\sum_{t=2}^{T} \epsilon_t^2}{T-1}$$

MLE for an MA(1) Process

Consider an MA(1) process:

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

with $\epsilon_t \sim i.i.d.N(0, \sigma^2)$. Let $\mathbf{\Theta} = (\mu, \theta, \sigma^2)'$ denote the population parameter to be estimated, then:

$$Y_t | \epsilon_{t-1} \sim N((\mu + \theta \epsilon_{t-1}, \sigma^2))$$

Remember that we set $\epsilon_0 = 0$:

$$Y_1 | \epsilon_0 \sim N(0, \sigma^2)$$

The conditional density of the tth observation is:

$$\begin{aligned} f_{Y_T | Y_{T-1}, Y_{T-2}, \ldots, Y_1, \epsilon_0 = 0}(y_t | y_{t-1}, y_{t-2} \ldots, y_1, \epsilon_0 = 0; \mathbf{\Theta}) \\ = f_{Y_t | \epsilon_{t-1}}(y_t | \epsilon_{t-1}; \mathbf{\Theta}) \\ = \frac{1}{\sqrt{2\pi\sigma^2}} exp[\frac{-\epsilon_t^2}{2\sigma^2}] \end{aligned}$$

MLE for an MA(1) Process

Similarly, the conditional likelihood would be:

$$f_{Y_T, Y_{T-1}, Y_{T-2}, \ldots, Y_1 | \epsilon_0 = 0}(y_t, y_{t-1}, y_{t-2} \ldots, y_1 | \epsilon_0 = 0; \mathbf{\Theta})$$
$$= \prod_{t=1}^{T} f_{Y_T | \epsilon_{T-1}}(y_t | \epsilon_{t-1}; \ )$$

The conditional log likelihood is:

$$\mathcal{L}(\mathbf{\Theta}) = -\frac{T}{2} log(2\pi) - \frac{T}{2} log(\sigma^2) - \sum_{t=1}^{T} \frac{\epsilon_t^2}{2\sigma^2}$$

MLE for an ARMA(p, q) Process

A Gaussian ARMA(p, q) process takes the form

$$\begin{aligned} Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} \\ + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \end{aligned} \tag{7}$$

The goal is to estimate the vector of population parameters $\boldsymbol{\Theta} = (c, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_1, \sigma^2)'$. A common approximation to tha likelihood function for an ARMA(p, q) process conditions on both y and ϵ

From the equation above we have:

$$\begin{aligned} \epsilon_t = y_t - c - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p} \\ - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} \end{aligned} \tag{8}$$

The conditional log likelihood is then:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Theta}) &= log\, f_{Y_t, Y_{t-1}, Y_{t-1}, \ldots, Y_1 | Y_0, \epsilon_0}(y_t, y_{t-1}, y_{t-2}, \ldots, y_1 | y_0, \epsilon_0; \boldsymbol{\Theta}) \\ &= -\frac{T}{2} log(2\pi) - \frac{T}{2} log(\sigma^2) - \sum_{t=1}^{T} \frac{\epsilon_t^2}{2\sigma^2} \end{aligned}$$