Chapter 3: Maximum Likelihood Estimation.

Jesús Bueren

EUI

Introduction

- The previous chapter assumed that the population parameters were known and showed how the population moments could be calculated.
- This chapter explores how to estimate the parameter values on the basis of observations on Y.
- This chapter follows chapter 5 in Hamilton.

Likelihood Function of an AR(1)

- A Gaussian AR(1) process takes the form:

$$Y_t = \phi Y_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2) \tag{1}$$

- the vector of population parameters to be estimated consists of $\boldsymbol{\theta} \equiv (\phi, \sigma^2)$

- The approach that we follow in this chapter will be to calculate the probability density:

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

  which can be viewed as the probability of observing the this particular data given a value of $\boldsymbol{\theta}$

- The maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ is the value of $\boldsymbol{\theta}$ that maximizes the probability of observing this particular data.

Likelihood Function of an AR(1)

- Consider the distribution of Y_1, the first observation in the sample.
- Since no previous observation is on the data, and assuming covariance-stationarity ($|\phi| < 1$), Y_1 comes from the unconditional distribution of Y which is given by:

$$Y_1 \sim N(0, \sigma^2/(1-\phi^2))$$

- Hence the density of the first observation takes the form:

$$\begin{aligned} f_{Y_1}(y_1; \boldsymbol{\theta}) &= f_{Y_1}(y_1; \phi, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}} \exp\left[\frac{1}{2}\frac{y_1^2}{\sigma^2/(1-\phi^2)}\right] \end{aligned}$$

Likelihood Function of an AR(1)

- Next, consider the distribution of the second observation Y_t conditional on observing $Y_{t-1} = y_{t-1}$.

$$Y_t | y_{t-1} \sim N(y_{t-1}, \sigma^2)$$

- Hence the density of the second observation takes the form:

$$\begin{aligned} f_{Y_t}(y_{t-1}; \boldsymbol{\theta}) &= f_{Y_t}(y_{t-1}; \phi, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{1}{2}\frac{(y_t - \phi y_{t-1})^2}{\sigma^2}\right] \end{aligned}$$

Likelihood Function of an AR(1)

- The joint likelihood of the full sample can be written as:

$$
\begin{aligned}
f_{Y_T,Y_{T-1},\dots,Y_2,Y_1}(y_T, y_{T-1}, \dots, y_2, y_1; \boldsymbol{\theta}) = \\
& f_{Y_T|Y_{T-1},\dots,Y_2,Y_1}(y_T|y_{T-1}, \dots, y_2, y_1; \boldsymbol{\theta}). \\
& f_{Y_{T-1}|Y_{T-2},\dots,Y_2,Y_1}(y_{T-1}|y_{T-2}, \dots, y_2, y_1; \boldsymbol{\theta}). \\
& \dots \\
& f_{Y_2|Y_1}(y_2|y_1; \boldsymbol{\theta}).f_{Y_1}(y_1; \boldsymbol{\theta})
\end{aligned}
$$

- Since the process is AR(1):

$$
\begin{aligned}
f_{Y_T,Y_{T-1},\dots,Y_2,Y_1}(y_T, y_{T-1}, \dots, y_2, y_1; \boldsymbol{\theta}) = \\
& f_{Y_1}(y_1; \boldsymbol{\theta}). \prod_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \boldsymbol{\theta}).
\end{aligned}
$$

Likelihood Function of an AR(1)

- The log likelihood function of the full sample can be found by taking logs on the previous equation:

$$\mathcal{L}(\boldsymbol{\theta}) = \log f_{Y_1}(y_1; \boldsymbol{\theta}) + \sum_{t=2}^{T} \log f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \boldsymbol{\theta}).$$

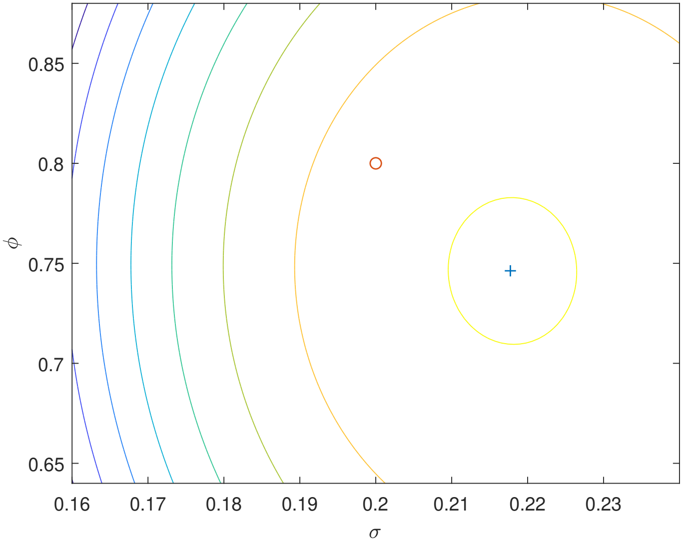
- Clearly, the value of $\boldsymbol{\theta}$, that maximizes the likelihood is identical to the value that maximizes the log-likelihood (computationally convenient).

Likelihood Function of an AR(1)

- To show how this works in practice I simulate an AR(1) process with $\phi = 0.8$ and $\sigma = 1$, and $T = 200$.
- I computed the log-likelihood on a grid of ϕ's and σ's and the MLE using a non-linear equation solver.

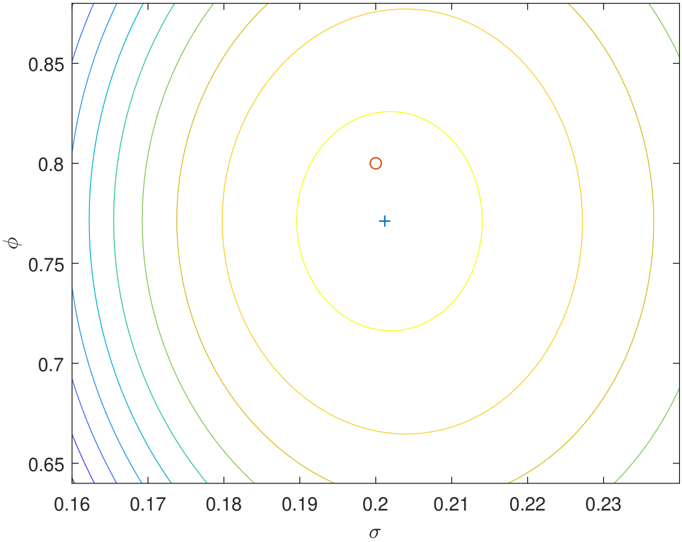
Likelihood Function of an AR(1)

T=100



Likelihood Function of an AR(1)

T=500



Conditional Likelihood Function of an AR(1)

- An alternative to numerical maximization of the exact likelihood function is to regard the value of y_1 as deterministic and maximize the likelihood conditioned on the first observation,

$$f_{Y_T,Y_{T-1},\ldots,Y_2|Y_1}(y_T, y_{T-1}, \ldots, y_2|y_1;\boldsymbol{\theta}) = \prod_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\boldsymbol{\theta}),$$

the objective being to maximize:

$$\log f_{Y_T,Y_{T-1},\ldots,Y_2|Y_1}(y_T, y_{T-1}, \ldots, y_2|y_1;\boldsymbol{\theta}) = -[(T-1)/2]\log\sigma^2 - [(T-1)/2]\log(2\pi) - \sum_{t=2}^{T}\frac{(y_t - \phi y_{t-1})^2}{2\sigma^2}, \quad (2)$$

Conditional Likelihood Function of an AR(1)

- Maximization of equation (2) wrt ϕ is equivalent to minimization of:

$$\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2,$$

which is achieved by an OLS regression of y_t on y_{t-1}:

$$\hat{\phi} = \frac{\sum_{t=2}^{T} y_{t-1} y_t}{\sum_{t=2}^{T} y_{t-1}^2}$$

- Differentiating equation (2) wrt σ^2 we get:

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=2}^{T}(y_t - \hat{\phi} y_{t-1})^2$$

- The conditional MLE is trivial to compute: if sample is large $\simeq$ MLE and doesn't require $|\phi| < 1$

Likelihood Function of an AR(p)

- A Gaussian AR(p) process takes the form:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2)$$

- For write the Likelihood function we are going to break the joint density in two parts:

$$f_{Y_T,\dots,Y_1}(y_T, \dots, y_1; \theta) = f_{Y_T,\dots,Y_1|Y_p,\dots,Y_1}(y_T, \dots, y_{p+1}|y_p, \dots, y_p; \theta) f_{Y_p,\dots,Y_1}(y_p, \dots, y_1; \theta)$$

the first block is trivial (just like the ar1 case).

- The second block is a bit more tedious.

Likelihood Function of an AR(p)

- The first p observations in the sample $\mathbf{y}_p = (y_1, \dots, y_p)$ are the realization of a p-dimensional Gaussian variable: $\mathbf{Y}_p \sim N(\mathbf{0}, \mathbf{V})$, where,

$$\mathbf{V} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix}$$

  we saw how to derive γ's in the previous chapter.

- The rest is a piece of cake.

Conditional Likelihood Function of an AR(p)

- As in the AR(1) case, maximization of the full likelihood must be accomplished numerically.
- In contrast, the the conditional MLE (taking as given the first p observations) of ϕ_p coincides with an OLS regression of y_t on $y_{t-1}, \dots, y_{t-p}$.
- The conditional MLE coincides with a sample average of square residuals.
- The MLE and the conditional MLE estimates have the same large-sample distribution.

Conditional Likelihood on an ARMA(p,q)

- The simplest approach to calculating the exact likelihood function for a Gaussian ARMA process is to use the Kalman filter that we will cover in two chapters.
- A Gaussian $ARMA(p,q)$ process take the form:

$$\begin{aligned} Y_t = &\phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} \\ &+ \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}, \epsilon_t \sim N(0, \sigma^2) \end{aligned}$$

Conditional Likelihood on an ARMA(p,q)

- Taking as given $\boldsymbol{y}_0 \equiv (y_0, y_{-1}, \ldots, y_{-p+1})$ and $\boldsymbol{\epsilon}_0 \equiv (\epsilon_0, \epsilon_{-1}, \ldots, \epsilon_{-p+1})$, the sequence $\{\epsilon_1, \ldots, \epsilon_T\}$ can be recovered from:

$$\epsilon_t = y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}$$

- The conditional log-likelihood is then:

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

- Initial conditions:
  - Set $\boldsymbol{\epsilon}_0$ and $\boldsymbol{y}_0$ to zero.
  - Use the first p observations of y as initial conditions and set $\epsilon_1, \ldots, \epsilon_q$ to zero

Statistical Inference with Maximum Likelihood

- If the sample size T is sufficiently large the MLE $\hat{\theta}$ can be approximated by:

$$\hat{\theta} \sim N(\theta_0, T^{-1}\mathcal{J}^{-1}),$$

where $\mathcal{J}$ is the information matrix that can be estimated with:

$$\mathcal{J} = -\frac{1}{T}\frac{\partial^2 \mathcal{L}}{\partial\theta\partial\theta'}$$

- Therefore, one could use the estimated variance covariance matrix of $\hat{\theta}$ for testing hypotheses.

Likelihood Ratio Test

- Another popular approach to testing hypotheses about parameters that are estimated by maximum likelihood is the likelihood ratio test.
- Suppose a null hypothesis implies a set of m different restrictions on the value of the $(a \times 1)$ parameter vector $\boldsymbol{\theta}$.
- Road map:
  1. Estimate the restricted $\mathcal{L}(\tilde{\boldsymbol{\theta}})$ and the unrestricted model $\mathcal{L}(\hat{\boldsymbol{\theta}})$
  2. Under the null that these restriction are true: $2[\mathcal{L}(\hat{\boldsymbol{\theta}}) - \mathcal{L}(\tilde{\boldsymbol{\theta}})] \simeq \chi^2(m)$

Model Selection Criteria

- Inspection of the sample autocorrelation function and sample partial autocorrelation function to identify ARMA models is somewhat of an art rather than a science.
- A more rigorous procedure to identify an ARMA model is to use formal model selection criteria.
- The two most widely used criteria are the Akaike information criterion (AIC) and the Bayesian criterion (BIC or SIC):

$$\begin{aligned} \text{AIC} &= 2k - 2\mathcal{L} \\ \text{BIC} &= \log(T)k - 2\mathcal{L} \end{aligned}$$