EC402: MLE of Time Series Data

Christian Julliard

Department of Economics and FMG
London School of Economics

Outline

Outline

1. Maximum Likelihood for Time Series
2. Ergodic Theorem
3. Examples of MLE estimation
   - MLE of the AR(1) process
   - MLE of Nonlinear least squares models
   - MLE of the MA(1) process

MLE for Time Series Models

- The standard approach to MLE you have seen so far is to obtain the likelihood function by
  1. writing the density for each observation and then
  2. *since the observations are independent*, write the likelihood as the product of these densities.
- the standard approach will not work in our case since the *observations are dependent*.

But: a *joint density* can be always factored into a *conditional* times a *marginal*.

Example: if you have three observations

$$
\begin{aligned}
f(y_3, y_2, y_1) &= f(y_3 | y_2, y_1) \cdot f(y_2, y_1) \\
&= f(y_3 | y_2, y_1) \cdot f(y_2 | y_1) \cdot f(y_1).
\end{aligned}
$$

- Hence the likelihood for T observations is

$$L(y;\psi) = \left[\prod_{t=2}^{T} f(y_t | y_{t-1}, ..., y_1)\right] \cdot f(y_1) = \prod_{t=2}^{T} f(y_t | I_{t-1}) \cdot f(y_1)$$

where I_{t-1} denotes all the information available at time $t-1$.

- Taking logs then yields

$$\log L(y;\psi) = \sum_{t=2}^{T} \log f(y_t | I_{t-1}) + \log f(y_1).$$

Note: $f(y_1)$ can be either modeled directly or y_1 can be assumed to be a constant (more on this later)

Outline

(An) Ergodic Theorem

- If a stochastic process y_t, $t = 1, 2, ...$ is ergodic with mean $\mu < \infty$ then
$$p\lim \frac{1}{T}\sum_{t=1}^{T} y_t = \mu.$$
- Ergodicity is a sufficient condition for sample means to converge to their expectations.
- This definition extends to vector valued stochastic processes.
- Moreover, functions of vector valued ergodic processes are ergodic.

MLE Asympotics for Time Series Models

- Even if observations are dependent, for *ergodic* processes, the ML estimator of a vector of parameters ψ is generally consistent.
- Moreover, the asymptotic results derived for the MLE in the iid setting carry over for *ergodic* processes.
  That is, for a vector of parameters ψ and ergodic processes, we have the standard results

$$\sqrt{T}(\hat{\psi} - \psi) \xrightarrow{D} N\left(0, \left(\lim \frac{1}{T} I(\psi)\right)^{-1}\right) \tag{1}$$

where $I(\psi)$ is the information matrix.

Outline

MLE of the AR(1) process

- Consider the $AR(1)$

$$y_t = \phi y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim \text{iid } N(0, \sigma^2), \ \ |\phi| < 1.$$

- Then $y_t | \, y_{t-1}$ is $N(\phi y_{t-1}, \sigma^2)$, therefore

$$f\left(y_t | \, I_{t-1}\right) = f\left(y_t | \, y_{t-1}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \Big(\underbrace{y_t - \phi y_{t-1}^2}_{\varepsilon_t} \Big)^2 \right\}$$

- And the log likelihood is simply,

$$\begin{aligned} \log L(y; \phi, \sigma^2) \quad &= \quad -\frac{(T-1)}{2} \log 2\pi - \frac{(T-1)}{2} \log \sigma^2 \\ &\quad -\frac{1}{2\sigma^2} \sum_{t=2}^{T} (y_t - \phi y_{t-1})^2 + \log f(y_1). \end{aligned}$$

What do we do about the *initial condition*?

- One possibility is to condition on y_1, i.e. take it as fixed. In this case the final term can be dropped and the likelihood becomes the likelihood for the linear regression of y_t on y_{t-1} for observations $t = 2, ..., T$.
- Thus we have, at the maximum,

$$\begin{aligned}
\frac{\partial \log L}{\partial \phi} &= \frac{1}{\sigma^2} \sum \left(y_t - \phi y_{t-1}\right) y_{t-1} = 0 \\
&\Rightarrow \hat{\phi} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} \Rightarrow \hat{\phi} = \hat{\phi}_{OLS}
\end{aligned}$$

MLE of the AR(1) process

- Alternatively, you can use the *unconditional distribution* for y_1,

Recall: in the $AR(1)$, the unconditional mean, $E(y_t) = 0$, and the unconditional variance, $var(y_t) = \frac{\sigma^2}{(1-\phi^2)}$.

- so the unconditional distribution is $N\left(0, \frac{\sigma^2}{(1-\phi^2)}\right)$.
- This assumption for y_1 is sensible if the process has been going on for a long time at $t = 1$.
- Under this assumption

$$\log f(y_1) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2}(1-\phi^2)y_1^2$$

- And this gives the log likelihood

$$\log L(y;\phi,\sigma^2) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2 + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2}(1-\phi^2)y_1^2.$$

MLE of the AR(1) process

Note: these results can be extended to:

1. the stationary $AR(p)$ model
2. the regression model with both process independent regressors and lagged dependent variables.

MLE of Nonlinear least squares models

- An important sub class of MLE is that of nonlinear regression models,
$$y_t = g(x_t;\beta) + \varepsilon_t \qquad \varepsilon_t \text{ iid } N(0,\sigma^2), \quad t = 1, ..., T,$$
x_t process independent.
- Note that
$$\begin{aligned} \varepsilon_t(\beta) &= y_t - g(x_t;\beta) \\ f(\varepsilon_t(\beta)) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-\varepsilon_t(\beta)^2}{2\sigma^2}\right\}. \end{aligned}$$
- Hence,
$$\log L(\beta,\sigma^2) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t(\beta)^2,$$
- So, maximizing $\log L$ wrt β is equivalent to minimizing the residual sum of squares with respect to β.

MLE of Nonlinear least squares

- Differentiating the log likelihood,

$$
\begin{aligned}
\frac{\partial \log L}{\partial \beta} &= -\frac{1}{\sigma^2}\sum_t \frac{\partial \varepsilon_t(\beta)}{\partial \beta}\varepsilon_t(\beta) = \frac{1}{\sigma^2}\sum_t z_t\varepsilon_t = 0 \\
\frac{\partial \log L}{\partial \sigma^2} &= -\frac{T}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_t \varepsilon_t(\beta)^2 = 0
\end{aligned}
$$

  where

$$z_t = -\frac{\partial \varepsilon_t}{\partial \beta} = \frac{\partial g(x_t;\beta)}{\partial \beta}.$$

Note: the first order conditions with respect to β are nonlinear and the ML estimates of β have to be obtained by numerical maximization.

- The first order conditions with respect to σ^2 yield the usual ML estimator for σ^2,

$$\hat{\sigma}^2 = \frac{1}{T}\sum_t \varepsilon_t(\hat{\beta})^2.$$

- Recall that we constructed an estimate of the variance-covariance matrix of our estimates based on the empirical information matrix $I(\psi)$,

$$I(\psi) = -E\left[\frac{\partial^2 \log L(\psi)}{\partial\psi\partial\psi'}\right].$$

- In the present case $\psi = (\beta', \sigma^2)$.

- So, looking at the components of $I(\psi)$, we have

$$
\begin{aligned}
-E\left[\frac{\partial^2 \log L}{\partial\beta\partial\beta'}\right] &= \frac{1}{\sigma^2}\left[E\sum_t \frac{\partial^2\varepsilon_t}{\partial\beta\partial\beta'}\cdot\varepsilon_t + E\sum_t \frac{\partial\varepsilon_t}{\partial\beta}\frac{\partial\varepsilon_t}{\partial\beta'}\right] \\
&= \frac{1}{\sigma^2}\left[\sum_t E\frac{\partial^2\varepsilon_t}{\partial\beta\partial\beta'}\cdot E(\varepsilon_t) + E\sum_t \frac{\partial\varepsilon_t}{\partial\beta}\frac{\partial\varepsilon_t}{\partial\beta'}\right] \\
&= \frac{1}{\sigma^2}E\sum_t z_t z_t' \qquad \text{since } E(\varepsilon_t) = 0. \\
-E\left[\frac{\partial^2 L}{\partial(\sigma^2)^2}\right] &= -\frac{T}{2(\sigma^2)^2} + \frac{2}{2(\sigma^2)^3}\sum_t E(\varepsilon_t^2) \\
&= -\frac{T}{2(\sigma^2)^2} + \frac{2T}{2(\sigma^2)^3}\sigma^2 \\
&= \frac{T}{2(\sigma^2)^2} \\
-E\left[\frac{\partial^2 \log L}{\partial\beta\partial\sigma^2}\right] &= \frac{1}{(\sigma^2)^2}E\sum_t z_t\varepsilon_t = \frac{1}{(\sigma^2)^2}\sum_t E(z_t)E(\varepsilon_t) \\
&= 0 \text{ (since } x \text{ is indipendent of } \varepsilon)
\end{aligned}
$$

- Hence, the information matrix is

$$I(\psi) = \frac{1}{\sigma^2}\begin{bmatrix} E\sum_t z_t z_t' & 0 \\ 0 & \frac{T}{2\sigma^2} \end{bmatrix}.$$

- Inverting, and substituting the consistent ML estimates of β and σ^2 for unknown parameters, and the sample moment $\sum_t z_t z_t'$ for $E\sum_t z_t z_t'$, we approximate the distribution of $(\hat{\beta}', \hat{\sigma}^2)$ by

$$\begin{bmatrix} \hat{\beta}' - \beta' \\ \hat{\sigma}^2 - \sigma^2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}; \begin{bmatrix} \hat{\sigma}^2\left(\sum_t z_t z_t'\right)^{-1} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{T} \end{bmatrix}\right)$$

that is equivalent to (1)

MLE of the MA(1) process

- Consider the MA(1)

$$
\begin{aligned}
y_t &= \varepsilon_t + \theta\varepsilon_{t-1} \qquad \varepsilon_t \text{ iid } N(0,\sigma^2) \\
&\rightarrow \quad y_t | \, \varepsilon_{t-1} \sim N\left(\theta\varepsilon_{t-1}, \sigma^2\right)
\end{aligned}
$$

- Assume we start from $\varepsilon_0 = 0$, then we may define $\varepsilon_t(\theta)$ by using the recursive equation

$$
\varepsilon_t(\theta) = y_t - \theta\varepsilon_{t-1}(\theta), \qquad t = 1, 2, ..., T.
$$

- Since $\varepsilon_0 = 0$,

$$
\begin{aligned}
\varepsilon_1(\theta) &= y_1 \\
\varepsilon_2(\theta) &= y_2 - \theta y_1 \\
\varepsilon_3(\theta) &= y_3 - \theta y_2 + \theta^2 y_1 \\
\varepsilon_t(\theta) &= y_t - \theta y_{t-1} + \theta^2 y_{t-2} + ... + (-\theta)^{t-1} y_1 .
\end{aligned}
$$

MLE of the MA(1) process

- Since $y_t|\,\varepsilon_{t-1} \sim N\left(\theta\varepsilon_{t-1}, \sigma^2\right)$, then

$$f\left(y_t|\,I_{t-1}\right) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp -\frac{\left(y_t - \theta\varepsilon_{t-1}(\theta)\right)^2}{2\sigma^2}.$$

- So the log likelihood is

$$\begin{aligned} \log L(\theta, \sigma^2) &= -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - \theta\varepsilon_{t-1}(\theta)\right)^2 \\ &= -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t(\theta)^2. \end{aligned}$$

- As before we have

$$\frac{\partial \log L}{\partial\theta} = \frac{1}{\sigma^2}\sum_t z_t(\theta)\varepsilon_t(\theta) \text{ where } z_t(\theta) = -\frac{\partial\varepsilon_t(\theta)}{\partial\theta}.$$

- So the $\hat{\theta}_{MLE}$ satisfies

$$\sum_t z_t(\theta)\varepsilon_t(\theta) = 0$$

MLE of the MA(1) process

Furthermore, using the empirical $I(\psi)$ we can show as before that the variance of $\hat{\theta}$ is given by

$$var(\hat{\theta}) = \hat{\sigma}^2 \left(\sum_t z_t^2(\hat{\theta}) \right)^{-1}$$

where

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2(\hat{\theta}).$$