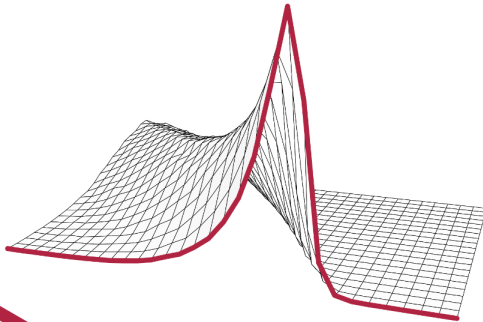


mcxs



Macroeconometrics

Lecture 2 Maximum Likelihood Estimation

Tomasz Woźniak

Department of Economics
University of Melbourne

Likelihood function

Estimation: analytical solution

Properties of the maximum likelihood estimator

Likelihood function

A simple model

Univariate linear regression model.

$$y_t = \beta x_t + \epsilon_t$$
$$\epsilon_t | x_t \sim iid \mathcal{N}(0, \sigma^2)$$

y_t – dependent variable

$\theta = (\beta, \sigma^2)'$ – a 2×1 vector of unknown parameters

x_t – explanatory variable

ϵ_t – error term

T – sample size and $t \in \{1, \dots, T\}$

A simple model

The model in matrix notation.

$$Y = \beta X + E$$
$$E|X \sim \mathcal{N}(\mathbf{0}_T, \sigma^2 I_T)$$

Data matrices.

$$\underset{(T \times 1)}{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} \quad \underset{(T \times 1)}{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_T \end{bmatrix} \quad \underset{(T \times 1)}{E} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{bmatrix}$$

Predictive density

Assumptions about the model and the conditional distribution of the error term determine the predictive distribution of data given the parameters and explanatory variables:

$$\begin{array}{l} Y = \beta X + E \\ E|X \sim \mathcal{N}(\mathbf{0}_T, \sigma^2 I_T) \end{array} \Rightarrow \begin{array}{l} Y = \beta X + E \\ Y|X \sim \mathcal{N}(\beta X, \sigma^2 I_T) \end{array}$$

Predictive density

Linear transformation of a normal vector.

Let a random vector Y follow an N -variate normal distribution with the mean vector μ and the covariance matrix Σ :

$$Y \sim \mathcal{N}_N(\mu, \Sigma)$$

Let $Z = AY + b$. Then:

$$Z \sim \mathcal{N}_N(A\mu + b, A\Sigma A')$$

Likelihood function

A likelihood function is equivalent to the conditional distribution of the data, given the parameters of the model.

However, for the purpose of the estimation and after plugging in data Y and X we treat it as a function of unknown parameters θ .

$$\begin{aligned} L(\theta|Y, X) &= L(\beta, \sigma^2|Y, X) = p(Y|X, \beta, \sigma^2) = \mathcal{N}_T(\beta X, \sigma^2 I_T) \\ &= (2\pi)^{-\frac{T}{2}} \det(\sigma^2 I_T)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(Y - \beta X)'(\sigma^2 I_T)^{-1}(Y - \beta X)\right\} \\ &= (2\pi)^{-\frac{T}{2}} (\sigma^2)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}(Y - \beta X)'(Y - \beta X)\right\} \end{aligned}$$

Useful operations.

Let c be a scalar and X an $N \times N$ matrix. Then $\det(cX) = c^N \det(X)$.

The likelihood principle

All of the information about the parameters of the model θ that is embedded in the dataset Y is captured by the likelihood function.

log-likelihood function

To derive the analytical solution and to be able to evaluate the likelihood function for any values of $\theta \in \Theta$, the logarithmic transformation is applied through which the log-likelihood function is obtained. Θ denotes the parameter space, that is, a set of all admissible values of the parameters.

$$\begin{aligned} l(\theta|Y, X) &= \ln L(\theta|Y, X) \\ &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2} \frac{1}{\sigma^2} (Y - \beta X)' (Y - \beta X) \\ &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2} \frac{1}{\sigma^2} (Y'Y - \beta' 2X'Y + \beta' X'X) \end{aligned}$$

The maximum likelihood estimator

The maximum likelihood estimator (MLE) of θ , denoted by $\hat{\theta}$, is found where the log-likelihood function is at its maximum:

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} l(\theta|Y, X)$$

Finding the maximum of the log-likelihood function is equivalent to finding the maximum of the likelihood function as the logarithm is a monotonic transformation that preserves local optima.

The derivations and properties are feasible under regularity conditions.

Regularity conditions

Let θ_0 denote the true values of the parameters θ .

A1 Existence

The following expectation exists:

$$\mathbb{E}[l(\theta|Y, X)] = \int_{-\infty}^{\infty} l(\theta|Y, X)L(\theta_0|Y, X)dY$$

A2 Convergence

$l(\theta|Y, X)$ converges in probability to its expectation uniformly in θ .

$$\text{plim } l(\theta|Y, X) = \mathbb{E}[l(\theta|Y, X)]$$

A3 Continuity

$l(\theta|Y, X)$ is continuous in θ .

A4 Differentiability

$l(\theta|Y, X)$ is at least twice differentiable in an open interval around θ_0 .

A5 Interchangeability

The differentiation and integration order of $l(\theta|Y, X)$ is interchangeable.

Estimation: analytical solution

Estimation: analytical solution

To derive the analytical solution of MLE use calculus.

Gradient vector.

$$\underset{(2 \times 1)}{G(\theta)} = \frac{\partial l(\theta|Y, X)}{\partial \theta} = \begin{bmatrix} \frac{\partial l(\theta|Y, X)}{\partial \beta} \\ \frac{\partial l(\theta|Y, X)}{\partial \sigma^2} \end{bmatrix}$$

The MLE occurs where all of the gradients are equal to zero:

$$G(\hat{\theta}) = \left. \frac{\partial l(\theta|Y, X)}{\partial \theta} \right|_{\theta=\hat{\theta}} = \mathbf{0}_2$$

Estimation: analytical solution

Hessian matrix.

$$H(\theta) = \frac{\partial^2 l(\theta|Y, X)}{\partial \theta \partial \theta'} = \begin{bmatrix} \frac{\partial^2 l(\theta|Y, X)}{\partial^2 \beta} & \frac{\partial^2 l(\theta|Y, X)}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 l(\theta|Y, X)}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 l(\theta|Y, X)}{\partial^2 \sigma^2} \end{bmatrix}$$

(2x2)

The MLE maximizes the log-likelihood function when the Hessian matrix ordinate at the MLE:

$$H(\hat{\theta}) = \frac{\partial^2 l(\theta|Y, X)}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}}$$

is negative definite.

Estimation: analytical solution

The gradient.

$$G(\theta) = \begin{bmatrix} \frac{\partial l(\theta|Y,X)}{\partial \beta} \\ \frac{\partial l(\theta|Y,X)}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \frac{1}{\sigma^2} (-2X'Y + \beta 2X'X) \\ -\frac{T}{2} \frac{1}{\sigma^2} + \frac{1}{2} \frac{1}{(\sigma^2)^2} (Y - \beta X)'(Y - \beta X) \end{bmatrix}$$

Necessary condition.

$$G(\hat{\theta}) = \begin{bmatrix} \frac{\partial l(\theta|Y,X)}{\partial \beta} \\ \frac{\partial l(\theta|Y,X)}{\partial \sigma^2} \end{bmatrix} \bigg|_{\theta=\hat{\theta}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{2} \frac{1}{\hat{\sigma}^2} (-2X'Y + \hat{\beta} 2X'X) \\ -\frac{T}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{2} \frac{1}{(\hat{\sigma}^2)^2} (Y - \hat{\beta} X)'(Y - \hat{\beta} X) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Estimation: analytical solution

The first equation.

$$\begin{aligned}0 &= -\frac{1}{2} \frac{1}{\hat{\sigma}^2} (-2X'Y + \hat{\beta}2X'X) \\ \hat{\beta}X'X &= X'Y \\ \hat{\beta} &= (X'X)^{-1}X'Y\end{aligned}$$

The second equation.

$$\begin{aligned}0 &= -\frac{T}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{2} \frac{1}{(\hat{\sigma}^2)^2} (Y - \hat{\beta}X)'(Y - \hat{\beta}X) \Bigg/ \cdot \frac{2(\hat{\sigma}^2)^2}{T} \\ \hat{\sigma}^2 &= \frac{1}{T} (Y - \hat{\beta}X)'(Y - \hat{\beta}X)\end{aligned}$$

Estimation: analytical solution

The Hessian matrix.

$$H(\theta) = \begin{bmatrix} \frac{\partial^2 l(\theta|Y,X)}{\partial^2 \beta} & \frac{\partial^2 l(\theta|Y,X)}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 l(\theta|Y,X)}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 l(\theta|Y,X)}{\partial^2 \sigma^2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sigma^2} X'X & -\frac{1}{(\sigma^2)^2} X'(Y - \beta X) \\ \frac{T}{2} \frac{1}{(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} (Y - \beta X)'(Y - \beta X) & \end{bmatrix}$$

Sufficient condition.

$H(\hat{\theta})$ must be negative definite.

$$H(\hat{\theta}) = \begin{bmatrix} -\frac{1}{\hat{\sigma}^2} X'X & -\frac{1}{(\hat{\sigma}^2)^2} X'\hat{E} \\ \frac{1}{2} \frac{T}{(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3} \hat{E}'\hat{E} & \end{bmatrix} = \begin{bmatrix} -\frac{1}{\hat{\sigma}^2} X'X & 0 \\ 0 & -\frac{1}{2} \frac{T}{(\hat{\sigma}^2)^2} \end{bmatrix}$$

where $\frac{1}{T} X'\hat{E} = 0$ (exogeneity condition), and $\hat{\sigma}^2 = \frac{1}{T} \hat{E}'\hat{E}$.

Estimation: analytical solution

Negative definite matrix.

A symmetric $N \times N$ matrix Z is negative definite if $z'Zz < 0$ for all $N \times 1$ vectors $z \neq \mathbf{0}_N$.

A symmetric 2×2 matrix Z is negative definite if $Z_{11} < 0$ and $\det(Z) > 0$.

Since $-\frac{1}{\hat{\sigma}^2}X'X < 0$ and $\det(H(\hat{\theta})) = \frac{1}{2} \frac{1}{(\hat{\sigma}^2)^3} X'X > 0$ the Hessian matrix is negative definite.

The MLE:

$$\hat{\theta} = \begin{bmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} (X'X)^{-1}X'Y \\ \frac{1}{T}(Y - \hat{\beta}X)'(Y - \hat{\beta}X) \end{bmatrix}$$

is a point at which the log-likelihood achieves the global maximum.

Properties of the MLE

MLE properties: consistency

Consistency.

The probability limit of the MLE when the sample size increases is the vector of the true parameter values.

$$\text{plim } \hat{\theta} = \theta_0$$

Definition of plim.

$$\lim_{T \rightarrow \infty} \Pr \left[|\hat{\theta} - \theta_0| < c \right] = 1, \text{ for any } c > 0$$

MLE properties: asymptotic normality

Normality.

The MLE converges in distribution to the following normal distribution when the sample size goes to infinity.

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Omega(\theta_0))$$

where $\Omega(\theta_0)$ is the inverse of the Fisher information matrix:

$$\Omega(\theta_0) = T I^{-1}(\theta_0) = T [-\mathbb{E}[H(\theta_0)]]^{-1}$$

Asymptotic distribution.

$$\hat{\theta} \overset{a}{\sim} \mathcal{N}\left(\theta_0, \frac{1}{T}\Omega(\theta_0)\right)$$

MLE properties: asymptotic normality

Estimator of covariance of $\hat{\theta}$.

$$\begin{aligned}\widehat{Var}(\hat{\theta}) &= \frac{1}{T} \Omega(\theta) |_{\theta=\hat{\theta}} = [-H(\theta)]^{-1} |_{\theta=\hat{\theta}} \\ &= \begin{bmatrix} \hat{\sigma}^2 (X'X)^{-1} & 0 \\ 0 & \frac{2(\hat{\sigma}^2)^2}{T} \end{bmatrix}\end{aligned}$$

Estimation standard erros.

$$\begin{aligned}\hat{se}(\hat{\beta}) &= \hat{\sigma} (X'X)^{-\frac{1}{2}} \\ \hat{se}(\hat{\sigma}^2) &= \sqrt{\frac{2}{T}} \hat{\sigma}^2\end{aligned}$$

MLE properties: efficiency and invariance

Efficiency.

The covariance of the MLE hits the Rao-Cramer lower bound:

$$\frac{1}{T}\Omega(\theta_0) = I^{-1}(\theta_0)$$

No other estimator has lower standard errors than the MLE.

Invariance.

The MLE of a continuous and differentiable function of parameters $g(\theta)$ is given by:

$$\widehat{g(\theta)} = g(\theta)|_{\theta=\hat{\theta}} = g(\hat{\theta})$$

Example: $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$

Maximum Likelihood Estimation

Maximum likelihood estimation and inference is a powerful tool for data analysis.

It is still one of the most frequently used methods in macroeconometrics as long as its application is **numerically feasible**.

Some specialised techniques, such as appropriately set **numerical optimization** and **concentration** of the likelihood function that are presented later during this subject, make its application simpler for some models.