Econometrics I

TA Session 11

Jukina HATAKEYAMA *

June 25, 2024

Contents

*E-mail: u868710a@ecs.osaka-u.ac.jp

1 The MLE of a Single Regression Model

Now we consider the MLE of single regression model: $y_i = \alpha + \beta x_i + u_i$, where u_i. Assume the error term, u_i follows the Gaussian distribution: $u_i \overset{i.i.d}{\sim} N(0, \sigma^2)$. Let us denote the probability density function of the error term $f_u(\theta; u_i)$ for all i. In addition, we set $f_y(\theta; y_i)$ as the pdf for y_i for all i. By the Change of Variables, we have:

$$\begin{aligned} f_y(\theta; y_i) &= f_u(\theta; y_i) \left| \frac{\partial u_i}{\partial y_i} \right| \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2) \end{aligned}$$

The parameter vector is $\theta = (\alpha, \beta, \sigma^2)' \in \mathbb{R}^3$. The joint density function, represented as $f_y(\theta; y_1, \cdots, y_n)$ (or simply $f_y(\theta; y)$), is rewritten as:

$$\begin{aligned} f_y(\theta; y_1, \cdots, y_n) &= f(\theta; y_1) \cdots f(\theta; y_n) \\ &= \prod_{i=1}^{n} f(\theta; y_i) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2) \end{aligned}$$

by the i.i.d assumption. This is the likelihood function. Then, the log-likelihood function is defined as:

$$\begin{aligned} l_n(\theta; (y, x)) &:= l_n(\theta; (y_i, x_i), i = 1, 2, \cdots, n) \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2. \end{aligned}$$

Given the observed data $(y_i, x_i), (i = 1, \cdots, n,)$ we consider the maximisation problem of the log-likelihood function with respect to $(\beta, \alpha, \sigma^2)$ and obtain the following MLE:

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} l_n(\theta; y, x).$$

The first order condition of the maximisation problem is given as follows:

$$\begin{aligned} \partial_\alpha l_n(\hat{\theta}; (y, x)) &= \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0, \\ \partial_\beta l_n(\hat{\theta}; (y, x)) &= \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i) x_i = 0, \\ \partial_{\sigma^2} l_n(\hat{\theta}; (y, x)) &= -\frac{n}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{2(\hat{\sigma}^2)^2} \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 = 0. \end{aligned}$$

The solution is denoted as $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2)'$, called the MLE. These solutions are given by

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \\ \hat{\alpha} &= \bar{y} - \hat{\beta} \bar{x}, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2. \end{aligned}$$

Note that the estimator of the variance is not same as that of the OLS. (biased estimator)

2 The MLE of a Multiple Regression Model

2.1 Reminder on Change of Variables

Consider the case that we change the density function $f_X(x)$ of a random variable X into another density function of $f_Z(z)$ of another random variable Z. In this subsection, we learn the properties of the changed random variable Z.

Theorem 2.1. Let $f_X(x)$ the density function of the random variable X. Let $z = \phi(x)$ with $\phi(\cdot)$ continuous and strictly monotone real value function. When the inverse function of $z = \phi(x)$ is given as $x = \phi^{-1}(z) = h(z)$, the following relationship is established:

$$f_Z(z) = |h'(z)| f_X(h(z)). \tag{1}$$

Proof. Suppose that the distribution function of X as $F_X(x)$ and the distribution function of Z as $F_Z(z)$.

(i) In the case of $h'(x) > 0$, $F_Z(z)$ is rewritten as follows:

$$\begin{aligned} F_Z(z) &= \text{Prob}(Z \leq z) = \text{Prob}(\phi(X) \leq z) \\ &= \text{Prob}(X \leq h(z)) = F_X(h(z)) \end{aligned}$$

By differentiating both sides of the above equation, we can get $f_Z(z) = h(z)' f_X(h(z))$.

(ii) In the case of $h'(x) < 0$, $F_Z(z)$ is rewritten as

$$\begin{aligned} F_Z(z) &= \text{Prob}(Z \leq z) = \text{Prob}(\phi(X) \leq z) \\ &= \text{Prob}(X \geq h(z)) = 1 - \text{Prob}(X \leq h(z)) \\ &= 1 - F_X(h(z)), \end{aligned}$$

because $\phi(\cdot)$ is strictly monotone. By differentiating both sides of the above equation, we can get $f_Z(z) = -h'(z) f_X(h(z))$.

□

In the multivariate case, if $Z = H(X)$ with H a bijective and differentiable function, the density of Z is

$$f_Z(z) = f_X(x)|\det(\nabla_y x)|,$$

where the differential is the Jacobian of the inverse of H, evaluated at y.

2.2 Multiple Regression Model

Multivariate Normal Distribution

Let X a n-dimensional random vector. When X follows **a multivariate normal distribution**, denoted as $X \sim N_{\mathbb{R}^{dim(X)}}(\mu, \Sigma)$, its pdf is defined as:

$$f(X) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)] \tag{2}$$

.

Suppose the regression model such that $y = x\beta + u$, where $u \sim N(0, \sigma^2 I_n)$. Then, the density function of u is

$$f_u(u) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\frac{1}{2\sigma^2} u'u),$$

By the change of variables from u to y, we have:

$$\begin{aligned} f_Y(y) &= f_u(y - x\beta)\det(\nabla_y u) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\frac{1}{2\sigma^2}(y - x\beta)'(y - x\beta)), \end{aligned}$$

since we have $\nabla_y u = I_n$. Remind that we can calculate the joint density as the products of individual densities like the case of the single regression, because conditionally on x_i, $y_i|x_i$ are iid. Assume that the case of $\theta = (\beta', \sigma^2)' \in \mathbb{R}^{K+1}$. The statistical criterion is

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \mathbb{L}_n(\theta; y, x),$$

with the log-likelihood function

$$\mathbb{L}_n(\theta; y, x) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - x\beta)'(y - x\beta).$$

Then, by optimizing the above equation, we have MLEs as follows:

$$\hat{\beta} = (x'x)^{-1}(x'y), \quad \hat{\sigma}^2 = \frac{1}{n}(y - x\hat{\beta})'(y - x\hat{\beta}).$$

3 The Properties of AR(1) Model and its Estimation

The AR(1) process satisfies the following stochastic difference equation:

$$y_t = \phi_1 y_{t-1} + \epsilon_t, \quad |\phi| < 1,$$

where ϵ_t is the white noise such as:

$$\begin{aligned}
&\mathbb{E}(\epsilon_t) = 0, \\
&\gamma_k = \mathbb{E}(\epsilon_t \epsilon_{t-k}) = \begin{cases} \sigma^2, & k = 0 \\ 0, & k \neq 0. \end{cases}
\end{aligned}$$

In this class, we assume that $\epsilon_t \overset{i.i.d}{\sim} N(0, \sigma^2)$. The conditional mean and variance of y_t given $\{y_{t-1}, y_{t-2}, \cdots\}$ are given as:

$$\begin{aligned}
&\mathbb{E}(y_t | y_{t-1}, y_{t-2}, \cdots) = \phi_1 y_{t-1}; \\
&\mathrm{Var}(y_t | y_{t-1}, y_{t-2}, \cdots) = \sigma^2.
\end{aligned}$$

Thus, $\{y_t | y_{t-1}, y_{t-2}, \cdots\} \overset{i.i.d}{\sim} N(0, \sigma^2)$ and the conditional distribution of y_t is

$$f(y_t | y_{t-1}, y_{t-2}, \cdots) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(y_t - \phi_1 y_{t-1})^2).$$

3.1 Stationary Solution

Rewriting the AR(1) model, we have

$$\begin{aligned}
y_t &= \phi_1 y_{t-1} + \epsilon_t \\
&= \phi_1^2 y_{t-2} + \epsilon_t + \phi_1^2 \epsilon_{t-1} \\
&\vdots \\
&= \phi_1^s y_{t-s} + \epsilon_t + \phi_1 \epsilon_{t-1} + \cdots + \phi_1^{s-1} \epsilon_{t-s+1}.
\end{aligned}$$

As s goes to infinity, ϕ_1^s approaches to zero. Therefore, $\phi_1^s y_{t-s}$ also goes to zero. Thus, we have the following relationship such that:

$$y_t = \sum_{s=0}^{\infty} \phi_1^s \epsilon_{t-s}.$$

In this case, the mean of y_t is

$$\mathbb{E}(y_t) = \mathbb{E}(\epsilon_t + \phi_1 \epsilon_{t-1} + \cdots) = 0.$$

Also, the variance of y_t is

$$\begin{aligned}
\mathrm{Var}(y_t) &= \mathrm{Var}(\epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots) \\
&= \mathrm{Var}(\epsilon_t) + \mathrm{Var}(\phi_1 \epsilon_{t-1}) + \mathrm{Var}(\phi_1^2 \epsilon_{t-2}) + \cdots \\
&= \sigma^2 (1 + \phi_1^2 + \phi_1^4 + \cdots) \\
&= \frac{\sigma^2}{1 - \phi^2}.
\end{aligned}$$

As a consequence, the unconditional distribution of y_t is

$$f(y_t) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi_1^2)}}\exp(-\frac{1}{2\sigma^2/(1-\phi_1^2)}y_t^2).$$

The joint density of y_t is written as

$$\begin{aligned}
f(y_1,\cdots,y_n) &= f(y_1,\cdots,y_{n-1})f(y_n|y_1,\cdots,y_{n-1}) \\
&= f(y_1,\cdots,y_{n-2})f(y_{n-1}|y_1,\cdots,y_{n-2})f(y_n|y_1,\cdots,y_{n-1}) \\
&\vdots \\
&= f(y_1)\prod_{t=2}^{n} f(y_t|y_{t-1},\cdots,y_1) \\
&= \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi_1^2)}}\exp(-\frac{1}{2\sigma^2/(1-\phi_1^2)}y_1^2) \\
&\times \prod_{t=2}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}(y_t-\phi_1 y_{t-1})^2).
\end{aligned}$$

In the above equation, we use the Bayes Rule such as $f_{X,Y} = f_{X|Y}(x|y)f_Y(y)$. The log-likelihood function is given by

$$\begin{aligned}
\mathbb{L}(\theta; y_1,\cdots,y_n) = &-\frac{1}{2}\log(2\pi\sigma^2/(1-\phi_1^2)) - \frac{1}{2\sigma^2/(1-\phi_1^2)}y_1^2 \\
&-\frac{n-1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=2}^{n}(y_t-\phi_1 y_{t-1})^2.
\end{aligned}$$

The Newton-Raphson method can be applied to find the optimum.

4 Linear Regression Model with the Auto Correlation of the Error Term

4.1 GLS Method

The regression model with AR(1) error is defined as

$$y_t = \beta x_t + u_t,\ u_t = \rho u_{t-1} + \epsilon_t$$

where we assume $\epsilon_t \sim N(0,\sigma_\epsilon^2)$. Then, the stacked model is

$$Y = x\beta + u, \tag{3}$$

where we assume $u \overset{i.i.d.}{\sim} N(0,\sigma^2\Omega)$. Now we calculate the variance covariance matrix of the error term in (3). Assume that the variance of u_t is given as $\mathrm{Var}(u_t) = \sigma^2$ for all t and ϵ_t is independent every periods. Moreover, it is independent on the previous error terms. Because we know $\mathrm{Var}(u_t) = \sigma^2$ for all t and $u_t = \rho u_{t-1} + \epsilon_t$, we have

$$(1-\rho^2)\sigma^2 = \sigma_\epsilon^2. \tag{4}$$

In addition, the auto-covariance of the disturbance term u_t and u_{t-1} is

$$\begin{aligned}
\mathrm{Cov}(u_t, u_{t-1}) &= \mathrm{Cov}(\rho u_{t-1} + \epsilon_t,\ u_{t-1}) \\
&= \mathbb{E}((\rho u_{t-1} + \epsilon_t)u_{t-1}) - \mathbb{E}(\rho u_{t-1} + \epsilon_t)\mathbb{E}(u_{t-1}) \\
&= \mathbb{E}((\rho u_{t-1} + \epsilon_t)u_{t-1}) \\
&= \mathbb{E}(\rho u_{t-1}^2 + u_{t-1}\epsilon_t) \\
&= \rho\sigma^2.
\end{aligned}$$

Generally, the covariance between u_t and u_{t-s} is $\mathrm{Cov}(u_t, u_{t-s}) = \rho^s\sigma^2$. Therefore, we can represent the variance covariance matrix of the error term by using the above relationship and (4):

$$\mathrm{Var}(u) = \frac{\sigma_\epsilon^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & \rho^2 & \cdots & \cdots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & \rho & 1 \end{pmatrix} := \sigma^2\Omega.$$

There exists P that satisfies the (Cholesky) decomposition $\Omega = PP'$. Then, like in the case of the regression model with the heteroscedasticity error term, multiply P^{-1} on both sides from the left:

$$\begin{aligned}
P^{-1}Y &= P^{-1}x\beta + P^{-1}u \\
Y^* &= x^*\beta + u^*,
\end{aligned} \tag{5}$$

where we represent $Y^* = (y_1^*, \cdots, y_n^*)'$ and $x^* = ({x'}_1^*, \cdots, {x'}_n^*)'$ as follows:

$$P^{-1}Y := Y^* = \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_n - \rho y_{n-1} \end{pmatrix}, \quad P^{-1}x := x^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}x_1 \\ x_2 - \rho x_1 \\ \vdots \\ x_n - \rho x_{n-1} \end{pmatrix}, \tag{6}$$

because the result of the decomposition is

$$P^{-1} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$

By applying the OLS method into (5), we can derive the estimator $\hat{\beta}$.

4.2 ML Method

Suppose the case that we estimate a MLE of the (3). Let $\theta_u = (\rho, \sigma_\epsilon^2)'$ and joint distribution of $u_1, \cdots, u_n$ is

$$
\begin{aligned}
f_u(u_1, \cdots, u_n; \theta) &= f(u_1; \theta) \prod_{t=2}^{n} f(u_t | u_{t-1}, \cdots, u_1; \theta) \\
&= \frac{1}{\sqrt{2\pi\sigma_\epsilon^2/(1-\rho^2)}} \exp(-\frac{1}{2\sigma_\epsilon^2/(1-\rho^2)} u_1^2) \\
&\times \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}^{n-1}} \exp(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^{n} (u_t - \rho u_{t-1})^2), \qquad (7)
\end{aligned}
$$

because we have $u_t \sim N(0, \frac{\sigma^2{}_\epsilon}{1-\rho^2})$. Let $\theta_y = (\beta, \theta_u')'$. Applying a change of variables from $u_1, \cdots, u_n$ to $y_1, \cdots, y_n$, the joint distribution of $y_1, \cdots, y_n$ is

$$
\begin{aligned}
f_y(y_1, \cdots, y_n; \theta_y) &= f_u(y_1 - \beta x_1, \cdots, y_n - \beta x_n; \theta_u) |\nabla_y u| \\
&= \frac{1}{\sqrt{2\pi\sigma_\epsilon^2/(1-\rho^2)}} \exp(-\frac{1}{2\sigma_\epsilon^2/(1-\rho^2)} (y_1 - \beta x_1)^2) \\
&\times \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}^{n-1}} \exp(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^{n} ((y_t - \rho y_{t-1}) - \beta(x_t - \rho x_{t-1}))^2) \\
&= \frac{1}{\sqrt{2\pi\sigma_\epsilon^2/(1-\rho^2)}} \exp(-\frac{1}{2\sigma_\epsilon^2} (\sqrt{1-\rho^2} y_1 - \sqrt{(1-\rho^2)} \beta x_1)^2) \\
&\times \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}^{n-1}} \exp(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^{n} ((y_t - \rho y_{t-1}) - \beta(x_t - \rho x_{t-1}))^2). \qquad (8)
\end{aligned}
$$

We thus rewrite the joint density function by using (6).

$$
\begin{aligned}
f_y(y_n, \cdots, y_1; \theta_y) &= (2\pi\sigma_\epsilon^2)^{-n/2} (1-\rho^2)^{1/2} \exp(-\frac{1}{2\sigma_\epsilon^2} (y_1^* - \beta x_1^*)) \\
&\times \exp(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^{n} (y_t^* - \beta x_t^*)^2) \qquad (9)
\end{aligned}
$$

The log-likelihood function is

$$
\begin{aligned}
\mathbb{L}_n(\theta_y; (y_i, x_i), i = 1, \cdots, n) &= -\frac{n}{2} \log(2\pi\sigma_\epsilon^2) + \frac{1}{2} \log(1-\rho^2) \\
&- \frac{1}{2\pi\sigma_\epsilon^2} \sum_{t=1}^{n} (y_t^* - \beta x_t^*)^2.
\end{aligned}
$$

By maximizing this log-likelihood, we can obtain the MLE same as the case of the GLS.

$$
\tilde{\beta} = (\sum_{t=1}^{n} x_t^{*\prime} x_t^*)^{-1} (\sum_{t=1}^{n} x_t^{*\prime} y_t^*) = (x^{*\prime} x^*)^{-1} (x^{*\prime} y^*).
$$

In the same manner, we can get the MLE of σ_ϵ^2 as follows:

$$
\tilde{\sigma_\epsilon}^2 = \frac{1}{n} \sum_{t=1}^{n} (y_t{}^* - \beta x_t{}^*)^2 = \frac{1}{n} (Y^* - x^*\beta)'(Y^* - x^*\beta).
$$