

where $\theta = (\beta, \sigma^2)$, because of $\frac{\partial u}{\partial y'} = I_n$.

Therefore, the log-likelihood function is:

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

Note that $|\Sigma|^{-1/2} = |\sigma^2 I_n|^{-1/2} = \sigma^{-n/2}$.

$$3. \quad \max_{\theta} \log L(\theta; y, X)$$

$$\text{(FOC)} \quad \frac{\partial \log L(\theta; y, X)}{\partial \theta} = 0$$

$$\text{(SOC)} \quad \frac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'} \text{ is a negative definite matrix.}$$

We obtain MLE of β and σ^2 :

$$\tilde{\beta} = (X'X)^{-1}X'y, \quad \tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})'(y - X\tilde{\beta})}{n},$$

where $\tilde{\sigma}^2$ is divided by n , not $n - k$.

4. Fisher's information matrix is:

$$I(\theta) = -E\left(\frac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'}\right)$$

The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound of the variance - covariance matrix for unbiased estimators of θ .

$$I(\theta)^{-1} = \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}$$

For large n , we approximately obtain: $\begin{pmatrix} \tilde{\beta} \\ \tilde{\sigma}^2 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}\right).$

9.3 MLE: The Case of Multiple Regression Model II

1. Regression model: $y = X\beta + u$, $u \sim N(0, \sigma^2\Omega)$

Transformation of Variables from u to y :

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

$$\begin{aligned} f_y(y) &= f_u(y - X\beta) \left| \frac{\partial u}{\partial y'} \right| \\ &= (2\pi\sigma^2)^{-n/2} |\Omega|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} (y - X\beta)' \Omega^{-1} (y - X\beta)\right) \\ &= L(\theta; y, X), \end{aligned}$$

where $\theta = (\beta, \sigma^2)$, because of $\frac{\partial u}{\partial y'} = I_n$.

The log-likelihood function is:

$$\log L(\theta; y, X) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log |\Omega| - \frac{1}{2\sigma^2} (y - X\beta)' \Omega^{-1} (y - X\beta),$$

where $\theta = (\beta, \sigma^2)$.

$$2. \quad \max_{\theta} \log L(\theta; y, X)$$

$$\text{(FOC)} \quad \frac{\partial \log L(\theta; y, X)}{\partial \theta} = 0$$

$$\text{(SOC)} \quad \frac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'} \text{ is a negative definite matrix.}$$

Then, we obtain MLE of β and σ^2 :

$$\tilde{\beta} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y, \quad \tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})' \Omega^{-1} (y - X\tilde{\beta})}{n}$$

3. Fisher's information matrix is defined as:

$$I(\theta) = -E\left(\frac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'}\right)$$

The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound of the

variance - covariance matrix for unbiased estimators of θ , which is given by:

$$I(\theta)^{-1} = \begin{pmatrix} \sigma^2(X'\Omega^{-1}X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}$$

9.4 MLE: AR(1) Model

The p th-order Autoregressive Model, i.e., AR(p) Model (p 次の自己回帰モデル):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t$$

AR(1) Model: $t = 2, 3, \dots, n$,

$$y_t = \phi_1 y_{t-1} + u_t, \quad u_t \sim N(0, \sigma^2)$$

where $|\phi_1| < 1$ is assumed for now.

To obtain the joint density function of $y_1, y_2, \dots, y_n$, $f(y_n, y_{n-1}, \dots, y_1)$ is decomposed as follows:

$$f(y_n, y_{n-1}, \dots, y_1) = f(y_1) \prod_{t=2}^n f(y_t | y_{t-1}, \dots, y_1).$$

From $y_t = \phi_1 y_{t-1} + u_t$, we can obtain:

$$E(y_t|y_{t-1}, \dots, y_1) = \phi_1 y_{t-1}, \quad \text{and} \quad V(y_t|y_{t-1}, \dots, y_1) = \sigma^2.$$

Therefore, the conditional distribution $f(y_t|y_{t-1}, \dots, y_1)$ is:

$$f(y_t|y_{t-1}, \dots, y_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_t - \phi_1 y_{t-1})^2\right).$$

To obtain the unconditional distribution $f(y_t)$, y_t is rewritten as follows:

$$\begin{aligned}y_t &= \phi_1 y_{t-1} + u_t \\&= \phi_1^2 y_{t-2} + u_t + \phi_1 u_{t-1} \\&\quad \vdots \\&= \phi_1^j y_{t-j} + u_t + \phi_1 u_{t-1} + \cdots + \phi_1^j u_{t-j} \\&\quad \vdots \\&= u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots, \quad \text{when } j \text{ goes to infinity.}\end{aligned}$$

The unconditional expectation and variance of y_t is:

$$E(y_t) = 0, \quad \text{and} \quad V(y_t) = \sigma^2(1 + \phi_1^2 + \phi_1^4 + \cdots) = \frac{\sigma^2}{1 - \phi_1^2}.$$

Therefore, the unconditional distribution of y_t is given by:

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

Finally, the joint distribution of $y_1, y_2, \dots, y_n$ is given by:

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

The log-likelihood function is:

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

Maximize $\log L$ with respect to ϕ_1 and σ^2 .

Maximization Procedure:

- Newton-Raphson Method, or Method of Scoring
- Simple Grid Search (search maximization within the range $-1 < \phi_1 < 1$, changing the value of ϕ_1 by 0.01)

9.5 MLE: Regression Model with AR(1) Error

When the error term is autocorrelated, the regression model is written as:

$$y_t = x_t\beta + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{iid } N(0, \sigma_\epsilon^2).$$

The joint distribution of $u_n, u_{n-1}, \dots, u_1$ is:

$$\begin{aligned} f_u(u_n, u_{n-1}, \dots, u_1; \rho, \sigma_\epsilon^2) &= f_u(u_1; \rho, \sigma_\epsilon^2) \prod_{t=2}^n f_u(u_t | u_{t-1}, \dots, u_1; \rho, \sigma_\epsilon^2) \\ &= (2\pi\sigma_\epsilon^2/(1-\rho^2))^{-1/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2/(1-\rho^2)}u_1^2\right) \\ &\quad \times (2\pi\sigma_\epsilon^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^n (u_t - \rho u_{t-1})^2\right). \end{aligned}$$

By transformation of variables from $u_n, u_{n-1}, \dots, u_1$ to $y_n, y_{n-1}, \dots, y_1$, the joint distribution of $y_n, y_{n-1}, \dots, y_1$ is:

$$\begin{aligned}
& f_y(y_n, y_{n-1}, \dots, y_1; \rho, \sigma_\epsilon^2, \beta) \\
&= f_u(y_n - x_n\beta, y_{n-1} - x_{n-1}\beta, \dots, y_1 - x_1\beta; \rho, \sigma_\epsilon^2) \left| \frac{\partial u}{\partial y} \right| \\
&= (2\pi\sigma_\epsilon^2/(1 - \rho^2))^{-1/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2/(1 - \rho^2)}(y_1 - x_1\beta)^2\right) \\
&\quad \times (2\pi\sigma_\epsilon^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^n ((y_t - \rho y_{t-1}) - (x_t - \rho x_{t-1})\beta)^2\right) \\
&= (2\pi\sigma_\epsilon^2)^{-1/2} (1 - \rho^2)^{1/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2} (\sqrt{1 - \rho^2} y_1 - \sqrt{1 - \rho^2} x_1\beta)^2\right) \\
&\quad \times (2\pi\sigma_\epsilon^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^n ((y_t - \rho y_{t-1}) - (x_t - \rho x_{t-1})\beta)^2\right) \\
&= (2\pi\sigma_\epsilon^2)^{-n/2} (1 - \rho^2)^{1/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2} (y_1^* - x_1^*\beta)^2\right) \times \exp\left(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^n (y_t^* - x_t^*\beta)^2\right)
\end{aligned}$$

$$\begin{aligned}
&= (2\pi)^{-n/2}(\sigma_\epsilon^2)^{-n/2}(1-\rho^2)^{1/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^n (y_t^* - x_t^*\beta)^2\right) \\
&= L(\rho, \sigma_\epsilon^2, \beta; y_n, y_{n-1}, \dots, y_1),
\end{aligned}$$

where y_t^* and x_t^* are given by:

$$\begin{aligned}
y_t^* &= \begin{cases} \sqrt{1-\rho^2}y_t, & \text{for } t = 1, \\ y_t - \rho y_{t-1}, & \text{for } t = 2, 3, \dots, n, \end{cases} \\
x_t^* &= \begin{cases} \sqrt{1-\rho^2}x_t, & \text{for } t = 1, \\ x_t - \rho x_{t-1}, & \text{for } t = 2, 3, \dots, n, \end{cases}
\end{aligned}$$

© For maximization, the first derivative of $L(\rho, \sigma_\epsilon^2, \beta; y_n, y_{n-1}, \dots, y_1)$ with respect to β should be zero.

$$\begin{aligned}\tilde{\beta} &= \left(\sum_{t=1}^T x_t^*{}' x_t^* \right)^{-1} \left(\sum_{t=1}^T x_t^*{}' y_t^* \right) \\ &= (X^{*'} X^*)^{-1} X^{*'} y^*\end{aligned}$$

$\Rightarrow$ This is equivalent to OLS from the regression model: $y^* = X^* \beta + \epsilon$ and $\epsilon \sim N(0, \sigma^2 I_n)$, where $\sigma^2 = \sigma_\epsilon^2 / (1 - \rho^2)$.

© For maximization, the first derivative of $L(\rho, \sigma_\epsilon^2, \beta; y_n, y_{n-1}, \dots, y_1)$ with respect to σ_ϵ^2 should be zero.

$$\tilde{\sigma}_\epsilon^2 = \frac{1}{n} \sum_{t=1}^n (y_t^* - x_t^* \beta)^2 = \frac{1}{n} (y^* - X^* \beta)' (y^* - X^* \beta),$$

where

$$y^* = \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{pmatrix} = \begin{pmatrix} \sqrt{1 - \rho^2} y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_n - \rho y_{n-1} \end{pmatrix}, \quad X^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix} = \begin{pmatrix} \sqrt{1 - \rho^2} x_1 \\ x_2 - \rho x_1 \\ \vdots \\ x_n - \rho x_{n-1} \end{pmatrix}.$$

© For maximization, the first derivative of $L(\rho, \sigma_\epsilon^2, \beta; y_n, y_{n-1}, \dots, y_1)$ with respect to ρ should be zero.

$$\max_{\beta, \sigma_\epsilon^2, \rho} L(\rho, \sigma_\epsilon^2, \beta; y) \text{ is equivalent to } \max_{\rho} L(\rho, \tilde{\sigma}_\epsilon^2, \tilde{\beta}; y).$$

$L(\rho, \tilde{\sigma}_\epsilon^2, \tilde{\beta}; y)$ is called the **concentrated log-likelihood function** (集約対数尤度関数), which is a function of ρ , i.e., both $\tilde{\sigma}_\epsilon^2$ and $\tilde{\beta}$ depend only on ρ .

The log-likelihood function is written as:

$$\begin{aligned}\log L(\rho, \tilde{\sigma}_\epsilon^2, \tilde{\beta}; y) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\tilde{\sigma}_\epsilon^2) + \frac{1}{2} \log(1 - \rho^2) - \frac{n}{2} \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} - \frac{n}{2} \log(\tilde{\sigma}_\epsilon^2(\rho)) + \frac{1}{2} \log(1 - \rho^2)\end{aligned}$$

For maximization of $\log L$, use Newton-Raphson method, method of scoring or simple grid search

Note that $\tilde{\sigma}_\epsilon^2 = \tilde{\sigma}_\epsilon^2(\rho) = \frac{1}{n}(y^* - X^*\tilde{\beta})'(y^* - X^*\tilde{\beta})$ for $\tilde{\beta} = (X^{*'}X^*)^{-1}X^{*'}y^*$.