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

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

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

We obtain MLE of β and σ^2:

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

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

4. Fisher's information matrix is:

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

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

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

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

9.3 MLE: The Case of Multiple Regression Model II

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

   Transformation of Variables from u to y:

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

$$f_y(y) = f_u(y - X\beta)\left|\frac{\partial u}{\partial y'}\right|$$

$$= (2\pi\sigma^2)^{-n/2}|\Omega|^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta)\Big)$$

$$= L(\theta; y, X),$$

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

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. $\max\limits_{\theta} \ \log L(\theta; y, X)$

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

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

Then, we obtain MLE of β and σ^2:

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

3. Fisher's information matrix is defined as:

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

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:

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

9.4 MLE: AR(1) Model

The pth-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: $\quad t = 2, 3, \cdots, n,$

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

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

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

$$
\begin{aligned}
&f(y_n, y_{n-1}, \cdots, y_1) \\
&\qquad = f(y_n|y_{n-1}, \cdots, y_1) f(y_{n-1}, y_{n-2}, \cdots, y_1) \\
&\qquad = f(y_n|y_{n-1}, \cdots, y_1) f(y_{n-1}|y_{n-2}, \cdots, y_1) f(y_{n-2}, y_{n-3}, \cdots, y_1) \\
&\qquad\qquad \cdots \\
&\qquad = f(y_n|y_{n-1}, \cdots, y_1) f(y_{n-1}|y_{n-2}, \cdots, y_1) f(y_{n-2}, y_{n-3}, \cdots, y_1) \ \cdots \ f(y_2|y_1) f(y_1) \\
&\qquad = f(y_1) \prod_{t=2}^{n} f(y_t|y_{t-1}, \cdots, y_1).
\end{aligned}
$$

Note that Bayes theorem is applied and repeated.

That is, $P(A \cap B) = P(A|B)P(B)$ for two events A and B.

We say that the joint distribution (or the likelihood function) is represented in the **innovation form**.

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

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

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

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

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

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

The unconditional expectation and variance of y_t is:

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

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 $y_1, y_2, \cdots, y_n$ is given by:

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

The log-likelihood function is:

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

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)

Simple Grid Search: From $\dfrac{\partial L(\phi_1, \sigma^2; y_n, y_{n-1}, \cdots, y_1)}{\partial \sigma^2} = 0$, we obtain:

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\Big((\sqrt{1-\phi_1^2}y_1)^2 + \sum_{t=2}^{n}(y_t - \phi_1 y_{t-1})^2 \Big) \\ &= \sigma^2(\phi_1)\end{aligned}$$

Substituting $\sigma^2(\phi_1)$ into σ^2, $\log L(\phi_1, \sigma^2; y_n, y_{n-1}, \cdots, y_1)$ is given by:

$$\log L(\phi_1, \sigma^2(\phi_1); y_n, y_{n-1}, \cdots, y_1) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2(\phi_1) + \frac{1}{2}\log(1-\phi_1^2) - \frac{n}{2}$$

which is a function of ϕ_1.

This likelihood function is called the **concentrated log-likelihood function (集約対数尤度関数)**

Find a maximum value of $\log L(\phi_1, \sigma^2(\phi_1); y_n, y_{n-1}, \cdots, y_1)$ for $\phi_1 = -0.99,\ -0.98,\ \cdots,\ 0.99$.

Another representation of the joint distribution: Mean and variance of $y = (y_1, y_2, \cdots, y_n$

Remember that for $|\tau| < 1$ we have the following:

$$\begin{aligned} y_t &= \phi_1^{\tau} y_{t-\tau} + u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots \phi_1^{\tau-1} u_{t-\tau+1} \\ &= u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots \end{aligned}$$

Mean:

$$\begin{aligned} \mathrm{E}(y_t) &= \mathrm{E}(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots) \\ &= \mathrm{E}(u_t) + \phi_1 \mathrm{E}(u_{t-1}) + \phi_1^2 \mathrm{E}(u_{t-2}) + \cdots \\ &= 0 \end{aligned}$$

Variance:

$$\begin{aligned} \mathrm{V}(y_t) &= \mathrm{V}(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots) \\ &= \mathrm{V}(u_t) + \phi_1^2 \mathrm{V}(u_{t-1}) + \phi_1^4 \mathrm{V}(u_{t-2}) + \cdots \end{aligned}$$

$$
\begin{aligned}
&= \sigma^2(1 + \phi_1^2 + \phi_1^4 + \cdots) \\
&= \frac{\sigma^2}{1 - \phi_1^2}.
\end{aligned}
$$

Covariance:

$$
\begin{aligned}
\gamma(\tau) &= \mathrm{Cov}(y_t, y_{t-\tau}) \\
&= \mathrm{E}(y_t y_{t-\tau}) = \mathrm{E}\Big((\phi_1^\tau y_{t-\tau} + u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots \phi_1^{\tau-1} u_{t-\tau+1}) y_{t-\tau}\Big) \\
&= \phi_1^\tau \mathrm{E}(y_{t-\tau}^2) + \mathrm{E}(u_t y_{t-\tau}) + \phi_1 \mathrm{E}(u_{t-1} y_{t-\tau}) + \phi_1^2 \mathrm{E}(u_{t-2} y_{t-\tau}) + \cdots \phi_1^{\tau-1} \mathrm{E}(u_{t-\tau+1} y_{t-\tau}) \\
&= \phi_1^\tau \mathrm{E}(y_{t-\tau}^2) \\
&= \phi_1^\tau \gamma(0)
\end{aligned}
$$

Note that $\mathrm{E}(u_t y_s) = 0$ for $t > s$, because y_s is a linear function of u_s, u_{s-1}, $\cdots$ and $\mathrm{E}(u_t u_s) = 0$ for $t > s$.

Moreover, note that $\mathrm{V}(y_t) = \gamma(0)$.

Thus,

$$\mathrm{E}(y) = \mathrm{E}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\mathrm{V}(y) = \mathrm{E}(yy') = \mathrm{E}\left(\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}(y_1, y_2, \cdots, y_n)\right)$$

$$= \begin{pmatrix} \gamma(0) & \gamma(1) & & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) \\ & \gamma(1) & \gamma(0) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \gamma(1) \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(1) & \gamma(0) \end{pmatrix}$$

$$
= \begin{pmatrix}
\gamma(0) & \phi_1\gamma(0) & & \cdots & \phi_1^{n-1}\gamma(0) \\
\phi_1\gamma(0) & \gamma(0) & \phi_1\gamma(0) & \cdots & \phi_1^{n-2}\gamma(0) \\
 & \phi_1\gamma(0) & \gamma(0) & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \phi_1\gamma(0) \\
\phi_1^{n-1}\gamma(0) & \phi_1^{n-2}\gamma(0) & \cdots & \phi_1\gamma(0) & \gamma(0)
\end{pmatrix}
$$

$$
= \gamma(0) \begin{pmatrix}
1 & \phi_1 & & \cdots & \phi_1^{n-1} \\
\phi_1 & 1 & \phi_1 & \cdots & \phi_1^{n-2} \\
 & \phi_1 & 1 & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \phi_1 \\
\phi_1^{n-1} & \phi_1^{n-2} & \cdots & \phi_1 & 1
\end{pmatrix}
$$

$$= \frac{\sigma^2}{1-\phi_1^2} \begin{pmatrix} 1 & \phi_1 & & \cdots & \phi_1^{n-1} \\ \phi_1 & 1 & \phi_1 & \cdots & \phi_1^{n-2} \\ & \phi_1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \phi_1 \\ \phi_1^{n-1} & \phi_1^{n-2} & \cdots & \phi_1 & 1 \end{pmatrix} = \Omega$$

Thus, the joint distribution of $y = (y_1, y_2, \cdots, y_n)$ is:

$$f(y) = (2\pi)^{-n/2}|\Omega|^{-1/2}\exp(-\frac{1}{2}y'\Omega^{-1}y),$$

which is the same as the innovation form.