4. Asymptotic Normality of MLE:

Let $\tilde{\theta}$ be MLE of θ.

As n goes to infinity, we have the following result:

$$\sqrt{n}(\tilde{\theta} - \theta) \longrightarrow N\left(0, \lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)^{-1}\right),$$

where it is assumed that $\lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)$ converges.

That is, when n is large, $\tilde{\theta}$ is approximately distributed as follows:

$$\tilde{\theta} \sim N\left(\theta, (I(\theta))^{-1}\right).$$

Suppose that $s(X) = \tilde{\theta}$.

When n is large, $\mathrm{V}(s(X))$ is approximately equal to $(I(\theta))^{-1}$.

5. **Optimization (最適化):**

MLE of θ results in the following maximization problem:

$$\max_{\theta} \ \log L(\theta; x).$$

We often have the case where the solution of θ is not derived in closed form.

$\Longrightarrow$ Optimization procedure

$$0 = \frac{\partial \log L(\theta; x)}{\partial \theta} = \frac{\partial \log L(\theta^*; x)}{\partial \theta} + \frac{\partial^2 \log L(\theta^*; x)}{\partial \theta \partial \theta'}(\theta - \theta^*).$$

Solving the above equation with respect to θ, we obtain the following:

$$\theta = \theta^* - \left(\frac{\partial^2 \log L(\theta^*; x)}{\partial \theta \partial \theta'}\right)^{-1} \frac{\partial \log L(\theta^*; x)}{\partial \theta}.$$

Replace the variables as follows:

$$\theta \longrightarrow \theta^{(i+1)}$$

$$\theta^* \longrightarrow \theta^{(i)}$$

Then, we have:

$$\theta^{(i+1)} = \theta^{(i)} - \left(\frac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'}\right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial\theta}.$$

$\Longrightarrow$ **Newton-Raphson method (ニュートン・ラプソン法)**

Replacing $\dfrac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'}$ by $\mathrm{E}\left(\dfrac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'}\right)$, we obtain the following optimization algorithm:

$$\begin{aligned}
\theta^{(i+1)} &= \theta^{(i)} - \left(\mathrm{E}\left(\frac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'}\right)\right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial\theta} \\
&= \theta^{(i)} + \left(I(\theta^{(i)})\right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial\theta}
\end{aligned}$$

$\Longrightarrow$ **Method of Scoring (スコア法)**

9.1 MLE: The Case of Single Regression Model

The regression model:

$$y_i = \beta_1 + \beta_2 x_i + u_i,$$

1. $u_i \sim N(0, \sigma^2)$ is assumed.

2. The density function of u_i is:

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

Because $u_1, u_2, \cdots, u_n$ are mutually independently distributed, the joint density function of $u_1, u_2, \cdots, u_n$ is written as:

$$f(u_1, u_2, \cdots, u_n) = f(u_1) f(u_2) \cdots f(u_n)$$
$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} u_i^2\right)$$

3. Using the transformation of variable ($u_i = y_i - \beta_1 - \beta_2 x_i$), the joint density function of $y_1, y_2, \cdots, y_n$ is given by:

$$f(y_1, y_2, \cdots, y_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2\right)$$
$$\equiv L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n).$$

$L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)$ is called the likelihood function.

$\log L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)$ is called the log-likelihood function.

$$\begin{aligned}
\log L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n) \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_t - \beta_1 - \beta_2 x_i)^2
\end{aligned}$$

4. **Transformation of Variable (変数変換)**:

Suppose that the density function of a random variable X is $f_x(x)$.

Defining $X = g(Y)$, the density function of Y, $f_y(y)$, is given by:

$$f_y(y) = f_x(g(y)) \left| \frac{\mathrm{d}g(y)}{\mathrm{d}y} \right|.$$

In the case where X and $g(Y)$ are $n \times 1$ vectors, $\left| \frac{\mathrm{d}g(y)}{\mathrm{d}y} \right|$ should be replaced by $\left| \frac{\partial g(y)}{\partial y'} \right|$, which is an absolute value of a determinant of the matrix $\frac{\partial g(y)}{\partial y'}$.

Example: When $X \sim U(0, 1)$, derive the density function of $Y = -\log(X)$.

$$f_x(x) = 1$$

$X = \exp(-Y)$ is obtained.

Therefore, the density function of Y, $f_y(y)$, is given by:

$$f_y(y) = \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| f_x(g(y)) = |-\exp(-y)| = \exp(-y)$$

5. Given the observed data $y_1, y_2, \cdots, y_n$, the likelihood function $L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)$, or the log-likelihood function $\log L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)$ is maximized with respect to $(\alpha, \beta, \sigma^2)$.

Solve the following three simultaneous equations:

$$\frac{\partial \log L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)}{\partial \alpha} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i) = 0,$$

$$\frac{\partial \log L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)}{\partial \beta} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i) x_i = 0,$$

$$\frac{\partial \log L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)^2 = 0.$$

The solutions of $(\beta_1, \beta_2, \sigma^2)$ are called the maximum likelihood estimates, denoted by $(\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\sigma}^2)$.

The maximum likelihood estimates are:

$$\tilde{\beta}_2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \qquad \tilde{\beta}_1 = \overline{y} - \tilde{\beta}_2 \overline{x}, \qquad \tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i)^2.$$

The MLE of σ^2 is divided by n, not $n - 2$.

9.2 MLE: The Case of Multiple Regression Model I

1. Multivariate Normal Distribution: $X: n \times 1$ and $X \sim N(\mu, \Sigma)$

   The density function of X is:

$$f(x) = (2\pi)^{n/2}|\Sigma|^{-1/2}\exp\Big(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\Big).$$

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

Transformation of Variables from u to y:

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

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

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

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

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

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

We obtain MLE of β and σ^2:

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

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 .

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

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

   Transformation of Variables from u to y:

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

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

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

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

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

Then, we obtain MLE of β and σ^2:

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

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: $t = 2, 3, \cdots, n,$

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

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

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

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

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:

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

The unconditional expectation and variance of y_t is:

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

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:

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

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:

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

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

$$f_u(u_n, u_{n-1}, \cdots, u_1; \rho, \sigma_\epsilon^2) = f_u(u_1; \rho, \sigma_\epsilon^2) \prod_{t=2}^{n} f_u(u_t | u_{t-1}, \cdots, u_1; \rho, \sigma_\epsilon^2)$$

$$= (2\pi\sigma_\epsilon^2/(1-\rho^2))^{-1/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2/(1-\rho^2)} u_1^2\right)$$
$$\times (2\pi\sigma_\epsilon^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^{n} (u_t - \rho u_{t-1})^2\right).$$

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

$$f_y(y_n, y_{n-1}, \cdots, y_1; \rho, \sigma_\epsilon^2, \beta)$$
$$= f_u(y_n - x_n\beta, y_{n-1} - x_{n-1}\beta, \cdots, y_1 - x_1\beta; \rho, \sigma_\epsilon^2) \left|\frac{\partial u}{\partial y'}\right|$$

$$= (2\pi\sigma_\epsilon^2/(1-\rho^2))^{-1/2}\exp\left(-\frac{1}{2\sigma_\epsilon^2/(1-\rho^2)}(y_1 - x_1\beta)^2\right)$$

$$\times(2\pi\sigma_\epsilon^2)^{-(n-1)/2}\exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{n}\Big((y_t-\rho y_{t-1})-(x_t-\rho x_{t-1})\beta\Big)^2\right)$$

$$= (2\pi\sigma_\epsilon^2)^{-1/2}(1-\rho^2)^{1/2}\exp\left(-\frac{1}{2\sigma_\epsilon^2}(\sqrt{1-\rho^2}y_1 - \sqrt{1-\rho^2}x_1\beta)^2\right)$$

$$\times(2\pi\sigma_\epsilon^2)^{-(n-1)/2}\exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{n}\Big((y_t-\rho y_{t-1})-(x_t-\rho x_{t-1})\beta\Big)^2\right)$$

$$= (2\pi\sigma_\epsilon^2)^{-n/2}(1-\rho^2)^{1/2}\exp\left(-\frac{1}{2\sigma_\epsilon^2}(y_1^* - x_1^*\beta)^2\right)\times\exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{n}(y_t^* - x_t^*\beta)^2\right)$$

$$
= (2\pi)^{-n/2}(\sigma_\epsilon^2)^{-n/2}(1-\rho^2)^{1/2}\exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{n}(y_t^* - x_t^*\beta)^2\right)
$$

$$
= L(\rho, \sigma_\epsilon^2, \beta; y_n, y_{n-1}, \cdots, y_1),
$$

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

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

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

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

$$\tilde{\beta} = (\sum_{t=1}^{T} x_t^{*\prime} x_t^*)^{-1}(\sum_{t=1}^{T} x_t^{*\prime} y_t^*)$$
$$= (X^{*\prime} X^*)^{-1} X^{*\prime} y^*$$

$\Longrightarrow$ 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 $L(\rho, \sigma_\epsilon^2, \beta; y_n, y_{n-1}, \cdots, y_1)$ 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

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

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

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

$L(\rho, \tilde{\sigma}_\epsilon^2, \tilde{\beta}; y)$ 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:

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

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

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