

The unconditional expectation and variance of y_t is:

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

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:

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

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)

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

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

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

Mean:

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

Variance:

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

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

Covariance:

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

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

$\text{E}(u_t u_s) = 0$ for $t > s$.

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

Thus,

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

$$V(y) = E(yy') = E \left(\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} (y_1, y_2, \dots, y_n) \right)$$
$$= \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) \\ & \gamma(1) & \gamma(0) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \gamma(1) \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(1) & \gamma(0) \end{pmatrix}$$

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

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

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

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

which is the same as the innovation form.

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.

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

Note that both $\tilde{\sigma}_\epsilon^2$ and $\tilde{\beta}$ depend only on ρ .

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

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

Remark: The regression model with AR(1) error is:

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

$$V(u) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{n-3} \\ \rho^3 & \rho^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \rho \\ \rho^{n-1} & \rho^{n-2} & \cdots & \rho^2 & \rho & 1 \end{pmatrix} = \sigma^2 \Omega, \quad \text{where } \sigma^2 = \frac{\sigma_\epsilon^2}{1 - \rho^2}.$$

where $\text{Cov}(u_i, u_j) = E(u_i u_j) = \sigma^2 \rho^{|i-j|}$, i.e., the i th row and j th column of Ω is $\rho^{|i-j|}$.

The regression model with AR(1) error is: $y = X\beta + u, \quad u \sim N(0, \sigma^2\Omega)$.

There exists P which satisfies that $\Omega = PP'$, because Ω is a positive definite matrix.

Multiply P^{-1} on both sides from the left.

$$\begin{aligned} P^{-1}y = P^{-1}X\beta + P^{-1}u &\implies y^* = X^*\beta + u^* \text{ and } u^* \sim N(0, \sigma^2I_n) \\ &\implies \text{Apply OLS.} \end{aligned}$$

$$y^* = \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_n - \rho y_{n-1} \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \cdots & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\rho & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = P^{-1}y$$

$$X^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}x_1 \\ x_2 - \rho x_1 \\ \vdots \\ x_n - \rho x_{n-1} \end{pmatrix} = P^{-1}X \quad \Rightarrow \quad \text{Check } P^{-1}\Omega P^{-1'} = aI_n, \\ \text{where } a \text{ is constant.}$$

9.6 MLE: Regression Model with Heteroscedastic Errors

In the case where the error term depends on the other exogenous variables, the regression model is written as follows:

$$y_i = x_i\beta + u_i, \quad u_i \sim \text{id } N(0, \sigma_i^2), \quad \sigma_i^2 = (z_i\alpha)^2.$$

The joint distribution of u_n, u_{n-1}, \dots, u_1 , denoted by $f_u(\cdot; \cdot)$, is given by:

$$\begin{aligned} \log f_u(u_n, u_{n-1}, \dots, u_1; \sigma_1^2, \dots, \sigma_n^2) &= \sum_{i=1}^n \log f_u(u_i; \sigma_i^2) \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log(\sigma_i^2) - \frac{1}{2} \sum_{i=1}^n \left(\frac{u_i}{\sigma_i} \right)^2 \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log(z_i\alpha)^2 - \frac{1}{2} \sum_{i=1}^n \left(\frac{u_i}{z_i\alpha} \right)^2 \end{aligned}$$

By the transformation of variables from u_n, u_{n-1}, \dots, u_1 to y_n, y_{n-1}, \dots, y_1 , the log-

likelihood function is:

$$\begin{aligned}L(\alpha, \beta; y_n, y_{n-1}, \dots, y_1) &= \log f_y(y_n, y_{n-1}, \dots, y_1; \alpha, \beta) \\ &= \log f_u(y_n - x_n\beta, y_{n-1} - x_{n-1}\beta, \dots, y_1 - x_1\beta; \sigma_i^2) \left| \frac{\partial u}{\partial y'} \right| \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log(z_i\alpha)^2 - \frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - x_i\beta}{z_i\alpha} \right)^2\end{aligned}$$

\Rightarrow Maximize the above log-likelihood function with respect to β and α .

10 Asymptotic Theory

1. **Definition: Convergence in Distribution** (分布収束)

A series of random variables $X_1, X_2, \dots, X_n, \dots$ have distribution functions F_1, F_2, \dots , respectively.

If

$$\lim_{n \rightarrow \infty} F_n = F,$$

then we say that a series of random variables X_1, X_2, \dots converges to F in distribution.

2. **Consistency** (一致性):

(a) **Definition: Convergence in Probability** (確率収束)

Let $\{Z_n : n = 1, 2, \dots\}$ be a series of random variables.