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

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:

$$\log L(\rho, \tilde\sigma_\epsilon^2, \tilde\beta; y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\tilde\sigma_\epsilon^2) + \frac{1}{2}\log(1-\rho^2) - \frac{n}{2}$$
$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2} - \frac{n}{2}\log\big(\tilde\sigma_\epsilon^2(\rho)\big) + \frac{1}{2}\log(1-\rho^2)$$

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 $\tilde\beta = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$.

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

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

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

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

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

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

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

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

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

$$X^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}x_1 \\ x_2 - \rho x_1 \\ \vdots \\ x_n - \rho x_{n-1} \end{pmatrix} = P^{-1}X \qquad \Longrightarrow \qquad \text{Check } P^{-1}\Omega P^{-1\prime} = aI_n,$$
where a 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:

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

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

$$\log f_u(u_n, u_{n-1}, \cdots, u_1; \sigma_1^2, \cdots, \sigma_n^2) = \sum_{i=1}^n \log f_u(u_i; \sigma_i^2)$$
$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n \log(\sigma_i^2) - \frac{1}{2}\sum_{i=1}^n \left(\frac{u_i}{\sigma_i}\right)^2$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n \log(z_i\alpha)^2 - \frac{1}{2}\sum_{i=1}^n \left(\frac{u_i}{z_i\alpha}\right)^2$$

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

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

$\Longrightarrow$ 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, $\cdots$ have distribution functions F_1, F_2, $\cdots$, respectively.

   If
   $$\lim_{i\to\infty} F_i = F,$$
   then we say that a series of random variables X_1, X_2, $\cdots$ converges to F in distribution.

2. **Consistency (一致性):**

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

   Let $\{Z_i : i = 1, 2, \cdots\}$ be a series of random variables.

   If the following holds,
   $$\lim_{i\to\infty} \text{Prob}(|Z_i - \theta| < \epsilon) = 1,$$
   for any positive ϵ, then we say that Z_i converges to θ in probability.

   θ is called a **probability limit (確率極限)** of Z_i.
   $$\text{plim } Z_i = \theta.$$

(b) Let $\hat{\theta}_i$ be an estimator of parameter θ.

If $\hat{\theta}_i$ converges to θ in probability, we say that $\hat{\theta}_T$ is a consistent estimator of θ.

3. **Chebyshev's inequality:**

For $g(X) \geq 0$,

$$\text{Prob}(g(X) \geq k) \leq \frac{\text{E}(g(X))}{k},$$

where k is a positive constant.

4. **Example:** For a random variable X, set $g(X) = (X-\mu)'(X-\mu)$, $\text{E}(X) = \mu$

and $\text{Var}(X) = \Sigma$.

Then, we have the following inequality:

$$\text{Prob}((X-\mu)'(X-\mu) \geq k) \leq \frac{\text{tr}(\Sigma)}{k}.$$

Note as follows:

$$\text{E}((X-\mu)'(X-\mu)) = \text{E}\Big(\text{tr}((X-\mu)'(X-\mu))\Big) = \text{E}\Big(\text{tr}((x-\mu)(x-\mu)')\Big)$$
$$= \text{tr}\Big(\text{E}((x-\mu)(x-\mu)')\Big) = \text{tr}(\Sigma).$$

5. **Example 1:**

Suppose that $X_i \sim (\mu, \sigma^2)$, $i = 1, 2, \cdots, n$.

Then, the sample average $\overline{X}$ is a consistent estimator of μ.

Proof:

Note that $g(\overline{X}) = (\overline{X}-\mu)^2$, $\epsilon^2 = k$, $\text{E}(g(\overline{X})) = \text{V}(\overline{X}) = \dfrac{\sigma^2}{n}$.

Use Chebyshev's inequality.

If $n \longrightarrow \infty$,

$$P(|\overline{X}-\mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0, \qquad \text{for any } \epsilon.$$

That is. for any ϵ,

$$\lim_{n\to\infty} P(|\overline{X}-\mu| < \epsilon) = 1$$

6. **Example 2 (Multivariate Case):**

Suppose that $X_i \sim (\mu, \Sigma)$, $i = 1, 2, \cdots, n$.

Then, the sample average $\overline{X}$ is a consistent estimator of μ.

Proof:

Note that $g(\overline{X}) = (\overline{X}-\mu)'(\overline{X}-\mu)$, $\epsilon^2 = k$, $\text{E}(g(\overline{X})) = \text{V}(\overline{X}) = \dfrac{1}{n}\Sigma$.

Use Chebyshev's inequality.

If $n \longrightarrow \infty$,

$$P((\overline{X}-\mu)'(\overline{X}-\mu) \geq k) \leq \frac{\text{tr}(\Sigma)}{nk} \longrightarrow 0, \qquad \text{for any positive } k.$$

That is. for any positive k,

$$\lim_{n\to\infty} P((\overline{X}-\mu)'(\overline{X}-\mu) < k) = 1$$

7. **Some Formulas:**

Let X_n and Y_n be the random variables which satisfy $\text{plim}\, X_n = c$ and $\text{plim}\, Y_n = d$.

Then,

(a) $\text{plim}\,(X_n + Y_n) = c + d$

(b) $\text{plim}\, X_n Y_n = cd$

(c) $\text{plim}\, X_n/Y_n = c/d$ for $d \neq 0$

(d) $\text{plim}\, g(X_n) = g(c)$ for a function $g(\cdot)$

$\Longrightarrow$ **Slutsky's Theorem (スルツキー定理)**

8. **Central Limit Theorem (中心極限定理)**

X_1, X_2, $\cdots$, X_n are mutually independently, identically and normally distributed as $X_i \sim (\mu,\ \Sigma)$.

Then,
$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu) \ \longrightarrow\ N(0,\Sigma)$$

9. **Central Limit Theorem (Generalization)**

X_1, X_2, $\cdots$, X_n are mutually independently, identically and normally distributed as $X_i \sim (\mu,\ \Sigma_i)$.

Then,
$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu) \ \longrightarrow\ N(0,\Sigma),$$
where
$$\Sigma = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\Sigma_t\right).$$

10. **Definition:** Let $\hat{\theta}_n$ be a consistent estimator of θ.

Suppose that $\sqrt{n}(\hat{\theta}_n-\theta)$ converges to $N(0,\Sigma)$ in distribution.

Then, we say that $\hat{\theta}_n$ has an **asymptotic distribution (漸近分布)**: $N(\theta,\Sigma/n)$.

11. **Definition:** We say that $\hat{\theta}_n$ is consistent uniformly asymptotically normal, when the following three conditions are satisfied:

  (a) $\hat{\theta}_n$ is consistent,

  (b) $\sqrt{n}(\hat{\theta}_n-\theta)$ converges to $N(0,\Sigma)$ in distribution,

  (c) Uniform convergence.

12. **Definition:** Suppose that $\hat{\theta}_n$ and $\tilde{\theta}_n$ are consistent, uniformly, asymptotically normal, and that the asymptotic variances are given by Σ/n and Ω/n.

If $\Omega-\Sigma$ is positive semidefinite, $\hat{\theta}_n$ is **asymptotically more efficient (漸近的**

に有効) than $\tilde{\theta}_n$.

13. **Definition:** If a consistent, uniformly, asymptotically normal estimator is asymptotically more efficient than any other consistent, uniformly, asymptotically normal estimators, we say that the consistent, uniformly, asymptotically normal estimator is asymptotically efficient (漸近的有効).

14. The sufficient condition for an asymptotically efficient and consistent, uniformly, asymptotically normal estimator is that the asymptotic variance is equivalent to Cramer-Rao lower bound.

15. $X_1, X_2, \cdots, X_n$ are random variables with density function $f(x;\theta)$.

Let $\hat{\theta}_n$ be a maximum likelihood estimator of θ.

Then, under some regularity conditions. $\hat{\theta}_n$ is a consistent estimator of θ and the asymptotic distribution of $\sqrt{n}(\hat{\theta}-\theta)$ is given by: $N\left(0, \lim\left(\frac{I(\theta)}{n}\right)^{-1}\right)$.

16. Regularity Conditions:

  (a) The domain of X_i does not depend on θ.

  (b) There exists at least third-order derivative of $f(x;\theta)$ with respect to θ, and their derivatives are finite.

17. Thus, MLE is

  (i) consistent,

  (ii) asymptotically normal, and

  (iii) asymptotically efficient.

18. **Slutsky's Theorem**

Let $\hat{\theta}$ be a consistent estimator of θ.

Then, $g(\hat{\theta})$ is also a consistent estimator of $g(\theta)$, where $g(\cdot)$ is a well-defined continuous function.

19. **Invariance of Maximum Likelihood Estimation (最尤法の不変性)**

Let $\hat{\theta}_1$, $\hat{\theta}_2$, $\cdots$, $\hat{\theta}_k$ be maximum likelihood estimators of θ_1, θ_2, $\cdots$, θ_k.

Consider the following one-to-one transformation:

$$\alpha_1 = \alpha_1(\theta_1, \theta_2, \cdots, \theta_k),\ \ \alpha_2 = \alpha_2(\theta_1, \theta_2, \cdots, \theta_k),\ \ \cdots,\ \ \alpha_k = \alpha_k(\theta_1, \theta_2, \cdots, \theta_k)$$

Then, MLEs of α_1, α_2, $\cdots$, α_k are given by:

$$\hat{\alpha}_1 = \alpha_1(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k),\ \ \hat{\alpha}_2 = \alpha_2(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k),\ \ \cdots,\ \ \hat{\alpha}_k = \alpha_k(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k).$$