Econometrics II

(Wed., 8:50-10:20)

Room # 509 (法経大学院総合研究棟)

- The prerequisite of this class is knowledge about **Econometrics I** (last semester) and **Econometrics** (undergraduate level).

TA Session (by Mr. Kinoshita):

From	**Oct. 9, 2013**
Wed.,	**13:00 - 14:30**
Room	**# 505 (法経大学院総合研究棟)**
Content:	**Matrix Algebra**

Econometrics (Undergraduate Course)

Mon., 8:50-10:20 (基礎工 B401)

Fri., 8:50-10:20 (基礎工 B401)

- If you have not taken Econometrics in undergraduate level, attend the above class.
- Textbook: 『計量経済学』(山本 拓 著，新世社)

1 Regression Analysis (回帰分析)

1.1 Setup of the Model

When (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) are available, suppose that there is a linear relationship between y and x, i.e.,

$$y_i = \beta_1 + \beta_2 x_i + u_i, \tag{1}$$

for $i = 1, 2, \cdots, n$. x_i and y_i denote the ith observations.

$\longrightarrow$ **Single (or simple) regression model (単回帰モデル)**

y_i is called the **dependent variable (従属変数)** or the **explained variable (被説明変数)**, while x_i is known as the **independent variable (独立変数)** or the **explanatory (or explaining) variable (説明変数)**.

$$\beta_1 = \textbf{Intercept (切片)}, \qquad \beta_2 = \textbf{Slope (傾き)}$$

β_1 and β_2 are unknown **parameters (パラメータ，母数)** to be estimated.

β_1 and β_2 are called the **regression coefficients (回帰係数)**.

u_i is the unobserved **error term (誤差項)** assumed to be a random variable with mean zero and variance σ^2.

σ^2 is also a parameter to be estimated.

x_i is assumed to be **nonstochastic (非確率的)**, but y_i is **stochastic (確率的)** because y_i depends on the error u_i.

The error terms $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed, which is called ***iid***.

It is assumed that u_i has a distribution with mean zero, i.e., $\mathrm{E}(u_i) = 0$ is assumed.

Taking the expectation on both sides of (1), the expectation of y_i is represented as:

$$\begin{aligned} \mathrm{E}(y_i) &= \mathrm{E}(\beta_1 + \beta_2 x_i + u_i) = \beta_1 + \beta_2 x_i + \mathrm{E}(u_i) \\ &= \beta_1 + \beta_2 x_i, \end{aligned} \tag{2}$$

for $i = 1, 2, \cdots, n$.

Using $\mathrm{E}(y_i)$ we can rewrite (1) as $y_i = \mathrm{E}(y_i) + u_i$.

(2) represents the true regression line.

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be estimates of β_1 and β_2.

Replacing β_1 and β_2 by $\hat{\beta}_1$ and $\hat{\beta}_2$, (1) turns out to be:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + e_i, \tag{3}$$

for $i = 1, 2, \cdots, n$, where e_i is called the **residual (残差)**.

The residual e_i is taken as the experimental value (or realization) of u_i.

We define $\hat{y}_i$ as follows:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i, \tag{4}$$

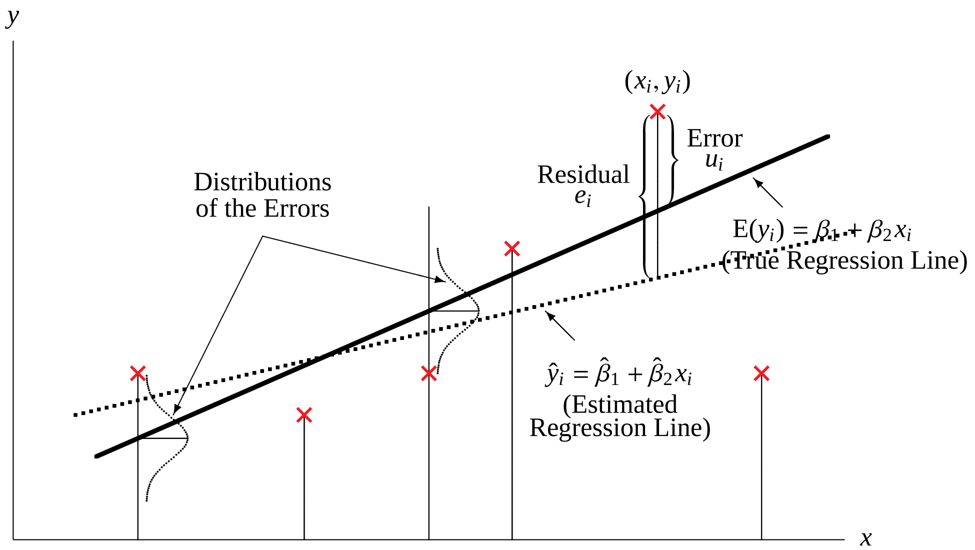
for $i = 1, 2, \cdots, n$, which is interpreted as the **predicted value (予測値)** of y_i.

(4) indicates the estimated regression line, which is different from (2).

Moreover, using $\hat{y}_i$ we can rewrite (3) as $y_i = \hat{y}_i + e_i$.

(2) and (4) are displayed in Figure 1.

Figure 1. True and Estimated Regression Lines (回帰直線)



Consider the case of $n = 6$ for simplicity.

$\times$ indicates the observed data series.

The true regression line (2) is represented by the solid line, while the estimated regression line (4) is drawn with the dotted line.

Based on the observed data, β_1 and β_2 are estimated as: $\hat{\beta}_1$ and $\hat{\beta}_2$.

In the next section, we consider how to obtain the estimates of β_1 and β_2, i.e., $\hat{\beta}_1$ and $\hat{\beta}_2$.

1.2 Ordinary Least Squares Estimation

Suppose that (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) are available.

For the regression model (1), we consider estimating β_1 and β_2.

Replacing β_1 and β_2 by their estimates $\hat{\beta}_1$ and $\hat{\beta}_2$, remember that the residual e_i is given by:

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i.$$

The sum of squared residuals is defined as follows:

$$S(\hat{\beta}_1, \hat{\beta}_2) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2.$$

It might be plausible to choose the $\hat{\beta}_1$ and $\hat{\beta}_2$ which minimize the sum of squared residuals, i.e., $S(\hat{\beta}_1, \hat{\beta}_2)$.

This method is called the **ordinary least squares estimation (最小二乘法，OLS)**.

To minimize $S(\hat{\beta}_1, \hat{\beta}_2)$ with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$, we set the partial derivatives equal to zero:

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0,$$

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0.$$

The second order condition for minimization is:

$$\begin{pmatrix} \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial\hat{\beta}_1^2} & \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial\hat{\beta}_1\partial\hat{\beta}_2} \\ \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial\hat{\beta}_2\partial\hat{\beta}_1} & \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial\hat{\beta}_2^2} \end{pmatrix} = \begin{pmatrix} 2n & 2\sum_{i=1}^n x_i \\ 2\sum_{i=1}^n x_i & 2\sum_{i=1}^n x_i^2 \end{pmatrix}$$

should be a positive definite matrix.

The diagonal elements $2n$ and $2\sum_{i=1}^n x_i^2$ are positive.

The determinant:

$$\begin{vmatrix} 2n & 2\sum_{i=1}^n x_i \\ 2\sum_{i=1}^n x_i & 2\sum_{i=1}^n x_i^2 \end{vmatrix} = 4n\sum_{i=1}^n x_i^2 - 4(\sum_{i=1}^n x_i)^2 = 4n\sum_{i=1}^n (x_i - \overline{x})^2$$

is positive. $\Longrightarrow$ The second-order condition is satisfied.

The first two equations yield the following two equations:

$$\overline{y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{x}, \tag{5}$$

$$\sum_{i=1}^{n} x_i y_i = n\overline{x}\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^{n} x_i^2, \tag{6}$$

where $\overline{y} = \dfrac{1}{n} \sum_{i=1}^{n} y_i$ and $\overline{x} = \dfrac{1}{n} \sum_{i=1}^{n} x_i$.

Multiplying (5) by $n\overline{x}$ and subtracting (6), we can derive $\hat{\beta}_2$ as follows:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}. \tag{7}$$

From (5), $\hat{\beta}_1$ is directly obtained as follows:

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}. \tag{8}$$

When the observed values are taken for y_i and x_i for $i = 1, 2, \cdots, n$, we say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are called the **ordinary least squares estimate**s (or simply the **least squares estimate**s, 最小二乗推定値) of β_1 and β_2.

When y_i for $i = 1, 2, \cdots, n$ are regarded as the random sample, we say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are called the **ordinary least squares estimator**s (or the **least squares estimator**s, 最小二乗推定量) of β_1 and β_2.

1.3 Properties of Least Squares Estimator

Equation (7) is rewritten as:

$$
\begin{aligned}
\hat{\beta}_2 &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} - \frac{\overline{y}\sum_{i=1}^{n}(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \\
&= \sum_{i=1}^{n} \frac{x_i - \overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2} y_i = \sum_{i=1}^{n} \omega_i y_i. \qquad (9)
\end{aligned}
$$

In the third equality, $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$ is utilized because of $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

In the fourth equality, ω_i is defined as: $\omega_i = \frac{x_i - \overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$.

ω_i is nonstochastic because x_i is assumed to be nonstochastic.

ω_i has the following properties:

$$\sum_{i=1}^{n} \omega_i = \sum_{i=1}^{n} \frac{x_i - \overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0, \tag{10}$$

$$\sum_{i=1}^{n} \omega_i x_i = \sum_{i=1}^{n} \omega_i (x_i - \overline{x}) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 1, \tag{11}$$

$$\sum_{i=1}^{n} \omega_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \right)^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2} = \frac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2}. \tag{12}$$

The first equality of (11) comes from (10).

From now on, we focus only on $\hat{\beta}_2$, because usually β_2 is more important than β_1 in the regression model (1).

In order to obtain the properties of the least squares estimator $\hat{\beta}_2$, we rewrite (9) as:

$$\begin{aligned}\hat{\beta}_2 &= \sum_{i=1}^{n} \omega_i y_i = \sum_{i=1}^{n} \omega_i(\beta_1 + \beta_2 x_i + u_i) \\ &= \beta_1 \sum_{i=1}^{n} \omega_i + \beta_2 \sum_{i=1}^{n} \omega_i x_i + \sum_{i=1}^{n} \omega_i u_i = \beta_2 + \sum_{i=1}^{n} \omega_i u_i. \end{aligned} \tag{13}$$

In the fourth equality of (13), (10) and (11) are utilized.

Mean and Variance of $\hat{\beta}_2$: $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2, but they are not necessarily normal.

Remember that we do not need normality assumption to obtain mean and variance but the normality assumption is required to test a hypothesis.

From (13), the expectation of $\hat{\beta}_2$ is derived as follows:

$$\begin{aligned} \mathrm{E}(\hat{\beta}_2) &= \mathrm{E}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \beta_2 + \mathrm{E}(\sum_{i=1}^{n} \omega_i u_i) \\ &= \beta_2 + \sum_{i=1}^{n} \omega_i \mathrm{E}(u_i) = \beta_2. \end{aligned} \tag{14}$$

It is shown from (14) that the ordinary least squares estimator $\hat{\beta}_2$ is an unbiased estimator of β_2.

From (13), the variance of $\hat{\beta}_2$ is computed as:

$$\begin{aligned} \mathrm{V}(\hat{\beta}_2) &= \mathrm{V}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \mathrm{V}(\sum_{i=1}^{n} \omega_i u_i) = \sum_{i=1}^{n} \mathrm{V}(\omega_i u_i) = \sum_{i=1}^{n} \omega_i^2 \mathrm{V}(u_i) \\ &= \sigma^2 \sum_{i=1}^{n} \omega_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}. \end{aligned} \tag{15}$$

The third equality holds because $u_1, u_2, \cdots, u_n$ are mutually independent.

The last equality comes from (12).

Thus, $\mathrm{E}(\hat{\beta}_2)$ and $\mathrm{V}(\hat{\beta}_2)$ are given by (14) and (15).

Gauss-Markov Theorem (ガウス・マルコフ定理): It has been discussed above that $\hat{\beta}_2$ is represented as (9), which implies that $\hat{\beta}_2$ is a linear estimator, i.e., linear in y_i.

In addition, (14) indicates that $\hat{\beta}_2$ is an unbiased estimator.

Therefore, summarizing these two facts, it is shown that $\hat{\beta}_2$ is a **linear unbiased estimator (線形不偏推定量)**.

Furthermore, here we show that $\hat{\beta}_2$ has minimum variance within a class of the linear unbiased estimators.

Consider the alternative linear unbiased estimator $\tilde{\beta}_2$ as follows:

$$\tilde{\beta}_2 = \sum_{i=1}^{n} c_i y_i = \sum_{i=1}^{n} (\omega_i + d_i) y_i,$$

where $c_i = \omega_i + d_i$ is defined and d_i is nonstochastic.

Then, $\tilde{\beta}_2$ is transformed into:

$$
\begin{aligned}
\tilde{\beta}_2 &= \sum_{i=1}^{n} c_i y_i = \sum_{i=1}^{n} (\omega_i + d_i)(\beta_1 + \beta_2 x_i + u_i) \\
&= \beta_1 \sum_{i=1}^{n} \omega_i + \beta_2 \sum_{i=1}^{n} \omega_i x_i + \sum_{i=1}^{n} \omega_i u_i + \beta_1 \sum_{i=1}^{n} d_i + \beta_2 \sum_{i=1}^{n} d_i x_i + \sum_{i=1}^{n} d_i u_i \\
&= \beta_2 + \beta_1 \sum_{i=1}^{n} d_i + \beta_2 \sum_{i=1}^{n} d_i x_i + \sum_{i=1}^{n} \omega_i u_i + \sum_{i=1}^{n} d_i u_i.
\end{aligned}
$$

Equations (10) and (11) are used in the forth equality.

Taking the expectation on both sides of the above equation, we obtain:

$$\begin{aligned}
\mathrm{E}(\tilde{\beta}_2) &= \beta_2 + \beta_1 \sum_{i=1}^{n} d_i + \beta_2 \sum_{i=1}^{n} d_i x_i + \sum_{i=1}^{n} \omega_i \mathrm{E}(u_i) + \sum_{i=1}^{n} d_i \mathrm{E}(u_i) \\
&= \beta_2 + \beta_1 \sum_{i=1}^{n} d_i + \beta_2 \sum_{i=1}^{n} d_i x_i.
\end{aligned}$$

Note that d_i is not a random variable and that $\mathrm{E}(u_i) = 0$.

Since $\tilde{\beta}_2$ is assumed to be unbiased, we need the following conditions:

$$\sum_{i=1}^{n} d_i = 0, \qquad \sum_{i=1}^{n} d_i x_i = 0.$$

When these conditions hold, we can rewrite $\tilde{\beta}_2$ as:

$$\tilde{\beta}_2 = \beta_2 + \sum_{i=1}^{n}(\omega_i + d_i)u_i.$$

The variance of $\tilde{\beta}_2$ is derived as:

$$\begin{aligned}
\mathrm{V}(\tilde{\beta}_2) &= \mathrm{V}\Big(\beta_2 + \sum_{i=1}^{n}(\omega_i + d_i)u_i\Big) = \mathrm{V}\Big(\sum_{i=1}^{n}(\omega_i + d_i)u_i\Big) = \sum_{i=1}^{n}\mathrm{V}\Big((\omega_i + d_i)u_i\Big) \\
&= \sum_{i=1}^{n}(\omega_i + d_i)^2\mathrm{V}(u_i) = \sigma^2(\sum_{i=1}^{n}\omega_i^2 + 2\sum_{i=1}^{n}\omega_i d_i + \sum_{i=1}^{n}d_i^2) \\
&= \sigma^2(\sum_{i=1}^{n}\omega_i^2 + \sum_{i=1}^{n}d_i^2).
\end{aligned}$$

From unbiasedness of $\tilde{\beta}_2$, using $\sum_{i=1}^{n} d_i = 0$ and $\sum_{i=1}^{n} d_i x_i = 0$, we obtain:

$$\sum_{i=1}^{n} \omega_i d_i = \frac{\sum_{i=1}^{n}(x_i - \overline{x})d_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i d_i - \overline{x} \sum_{i=1}^{n} d_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0,$$

which is utilized to obtain the variance of $\tilde{\beta}_2$ in the third line of the above equation. From (15), the variance of $\hat{\beta}_2$ is given by: $\mathrm{V}(\hat{\beta}_2) = \sigma^2 \sum_{i=1}^{n} \omega_i^2$.

Therefore, we have:

$$\mathrm{V}(\tilde{\beta}_2) \geq \mathrm{V}(\hat{\beta}_2),$$

because of $\sum_{i=1}^{n} d_i^2 \geq 0$.

When $\sum_{i=1}^{n} d_i^2 = 0$, i.e., when $d_1 = d_2 = \cdots = d_n = 0$, we have the equality: $\mathrm{V}(\tilde{\beta}_2) = \mathrm{V}(\hat{\beta}_2)$.

Thus, in the case of $d_1 = d_2 = \cdots = d_n = 0$, $\hat{\beta}_2$ is equivalent to $\tilde{\beta}_2$.

As shown above, the least squares estimator $\hat{\beta}_2$ gives us the **minimum variance linear unbiased estimator (最小分散線形不偏推定量)**, or equivalently the **best linear unbiased estimator (最良線形不偏推定量，BLUE)**, which is called the **Gauss-Markov theorem (ガウス・マルコフ定理)**.

Asymptotic Properties (漸近的性質) of $\hat{\beta}_2$: We assume that as n goes to infinity we have the following:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \longrightarrow m < \infty,$$

where m is a constant value. From (12), we obtain:

$$n\sum_{i=1}^{n}\omega_i^2 = \frac{1}{(1/n)\sum_{i=1}^{n}(x_i - \overline{x})} \longrightarrow \frac{1}{m}.$$

Note that $f(x_n) \longrightarrow f(m)$ when $x_n \longrightarrow m$, called **Slutsky's theorem (スルツキー定理)**, where m is a constant value and $f(\cdot)$ is a function.

We show both **consistency (一致性)** of $\hat{\beta}_2$ and **asymptotic normality (漸近正規性)** of $\sqrt{n}(\hat{\beta}_2 - \beta_2)$.

● First, we prove that $\hat{\beta}_2$ is a consistent estimator of β_2.

[Review] Chebyshev's inequality (チェビシェフの不等式) is given by:

$$P(|X-\mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}, \qquad \text{where } \mu = \mathrm{E}(X) \text{ and } \sigma^2 = \mathrm{V}(X).$$

[End of Review]

Replace X, E(X) and V(X) by:

$$\hat{\beta}_2, \qquad \mathrm{E}(\hat{\beta}_2) = \beta_2, \quad \text{and} \quad \mathrm{V}(\hat{\beta}_2) = \sigma^2 \sum_{i=1}^{n} \omega_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})}.$$

Then, when $n \longrightarrow \infty$, we obtain the following result:

$$P(|\hat{\beta}_2 - \beta_2| > \epsilon) \leq \frac{\sigma^2 \sum_{i=1}^n \omega_i^2}{\epsilon^2} = \frac{\sigma^2 n \sum_{i=1}^n \omega_i^2}{n\epsilon^2} \longrightarrow 0,$$

where $\sum_{i=1}^n \omega_i^2 \longrightarrow 0$ because $n \sum_{i=1}^n \omega_i^2 \longrightarrow \dfrac{1}{m}$ from the assumption.

Thus, we obtain the result that $\hat{\beta}_2 \longrightarrow \beta_2$ as $n \longrightarrow \infty$.

Therefore, we can conclude that $\hat{\beta}_2$ is a **consistent estimator (一致推定量)** of β_2.

● Next, we want to show that $\sqrt{n}(\hat{\beta}_2 - \beta_2)$ is asymptotically normal.

[Review] The **Central Limit Theorem (中心極限定理, CLT)** is: for random variables $X_1, X_2, \cdots, X_n$,

$$\frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \frac{\sum_{i=1}^n X_i - \mathrm{E}(\sum_{i=1}^n X_i)}{\sqrt{\mathrm{V}(\sum_{i=1}^n X_i)}} \longrightarrow N(0, 1), \qquad \text{as} \qquad n \longrightarrow \infty,$$

where $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i$.

$X_1, X_2, \cdots, X_n$ are not necesarily iid, if $\mathrm{V}(\overline{X})$ is finite as n goes to infinity.

[End of Review]

Note that $\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i$ as in (13), and X_i is replaced by $\omega_i u_i$.

From the central limit theorem, asymptotic normality is shown as follows:

$$\frac{\sum_{i=1}^{n} \omega_i u_i - \mathrm{E}(\sum_{i=1}^{n} \omega_i u_i)}{\sqrt{\mathrm{V}(\sum_{i=1}^{n} \omega_i u_i)}} = \frac{\sum_{i=1}^{n} \omega_i u_i}{\sigma \sqrt{\sum_{i=1}^{n} \omega_i^2}} = \frac{\hat{\beta}_2 - \beta_2}{\sigma / \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \longrightarrow N(0, 1),$$

where

- $\mathrm{E}(\sum_{i=1}^{n} \omega_i u_i) = 0$,
- $\mathrm{V}(\sum_{i=1}^{n} \omega_i u_i) = \sigma^2 \sum_{i=1}^{n} \omega_i^2$, and
- $\sum_{i=1}^{n} \omega_i u_i = \hat{\beta}_2 - \beta_2$

are substituted in the first and second equalities.

Moreover, we can rewrite as follows:

$$\frac{\hat{\beta}_2 - \beta_2}{\sigma / \sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = \frac{\sqrt{n}(\hat{\beta}_2 - \beta_2)}{\sigma / \sqrt{(1/n)\sum_{i=1}^{n}(x_i - \overline{x})^2}} \longrightarrow \frac{\sqrt{n}(\hat{\beta}_2 - \beta_2)}{\sigma / \sqrt{m}} \longrightarrow N(0, 1).$$

Or equivalently,

$$\sqrt{n}(\hat{\beta}_2 - \beta_2) \longrightarrow N(0, \frac{\sigma^2}{m}).$$

Thus, the asymptotic normality of $\sqrt{n}(\hat{\beta}_2 - \beta_2)$ is shown.

Finally, replacing σ^2 by its consistent estimator s^2, it is known as follows:

$$\frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} \longrightarrow N(0,1), \tag{16}$$

where s^2 is defined as:

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2, \tag{17}$$

which is a consistent and unbiased estimator of σ^2. $\longrightarrow$ Proved later.

Thus, using (16), in large sample we can construct the confidence interval and test the hypothesis.

Exact Distribution of $\hat{\beta}_2$: We have shown asymptotic normality of $\sqrt{n}(\hat{\beta}_2 - \beta_2)$, which is one of the large sample properties.

Now, we discuss the small sample properties of $\hat{\beta}_2$.

In order to obtain the distribution of $\hat{\beta}_2$ in small sample, the distribution of the error term has to be assumed.

Therefore, the extra assumption is that $u_i \sim N(0, \sigma^2)$.

Writing (13), again, $\hat{\beta}_2$ is represented as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i.$$

First, we obtain the distribution of the second term in the above equation.

[Review] Note that the **moment-generating function (積率母関数, MGF)** is given by $M(\theta) \equiv \mathrm{E}(\exp(\theta X)) = \exp(\mu\theta + \frac{1}{2}\sigma^2\theta^2)$ when $X \sim N(\mu, \sigma^2)$.

X_1, X_2, $\cdots$, X_n are mutually independently distributed as $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, \cdots, n$.

MGF of X_i is $M_i(\theta) \equiv \mathrm{E}(\exp(\theta X_i)) = \exp(\mu_i\theta + \frac{1}{2}\sigma_i^2\theta^2)$.

Consider the distribution of $Y = \sum_{i=1}^n (a_i + b_i X_i)$, where a_i and b_i are constant.

$$
\begin{aligned}
M_y(\theta) &\equiv \mathrm{E}(\exp(\theta Y)) = \mathrm{E}(\exp(\theta \textstyle\sum_{i=1}^n (a_i + b_i X_i))) \\
&= \textstyle\prod_{i=1}^n \exp(\theta a_i)\mathrm{E}(\exp(\theta b_i X_i)) = \prod_{i=1}^n \exp(\theta a_i) M_i(\theta b_i) \\
&= \textstyle\prod_{i=1}^n \exp(\theta a_i)\exp(\mu_i\theta b_i + \frac{1}{2}\sigma_i^2(\theta b_i)^2) = \exp(\theta \sum_{i=1}^n (a_i + b_i\mu_i) + \frac{1}{2}\theta^2 \sum_{i=1}^n b_i^2\sigma_i^2),
\end{aligned}
$$

which implies that $Y \sim N(\sum_{i=1}^{n}(a_i + b_i\mu_i), \sum_{i=1}^{n} b_i^2\sigma_i^2)$.

[End of Review]

Substitute $a_i = 0$, $\mu_i = 0$, $b_i = \omega_i$ and $\sigma_i^2 = \sigma^2$.

Then, using the moment-generating function, $\sum_{i=1}^{n} \omega_i u_i$ is distributed as:

$$\sum_{i=1}^{n} \omega_i u_i \sim N(0,\ \sigma^2 \sum_{i=1}^{n} \omega_i^2).$$

Therefore, $\hat{\beta}_2$ is distributed as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i \sim N(\beta_2,\ \sigma^2 \sum_{i=1}^{n} \omega_i^2),$$

or equivalently,

$$\frac{\hat{\beta}_2 - \beta_2}{\sigma \sqrt{\sum_{i=1}^{n} \omega_i^2}} = \frac{\hat{\beta}_2 - \beta_2}{\sigma / \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \sim N(0, 1),$$

for any n.

Moreover, replacing σ^2 by its estimator s^2 defined in (17), it is known that we have:

$$\frac{\hat{\beta}_2 - \beta_2}{s / \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \sim t(n-2),$$

where $t(n-2)$ denotes t distribution with $n-2$ degrees of freedom.

Thus, under normality assumption on the error term u_i, the $t(n-2)$ distribution is used for the confidence interval and the testing hypothesis in small sample.

Or, taking the square on both sides,

$$\Big(\frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}\Big)^2 \sim F(1, n-2),$$

which will be proved later.

Before going to **multiple regression model (重回帰モデル)**,

2 Some Formulas of Matrix Algebra

1. Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lk} \end{pmatrix} = [a_{ij}]$,

   which is a $l \times k$ matrix, where a_{ij} denotes ith row and jth column of A.

The **transposed matrix (転置行列)** of A, denoted by A', is defined as:

$$A' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{l1} \\ a_{12} & a_{22} & \cdots & a_{l2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{lk} \end{pmatrix} = [a_{ji}],$$

where the ith row of A' is the ith column of A.

2. $(Ax)' = x'A'$,

   where A and x are a $l \times k$ matrix and a $k \times 1$ vector, respectively.

3. $a' = a$,

where a denotes a scalar.

4. $$\frac{\partial a'x}{\partial x} = a,$$

where a and x are $k \times 1$ vectors.

5. $$\frac{\partial x'Ax}{\partial x} = (A + A')x,$$

where A and x are a $k \times k$ matrix and a $k \times 1$ vector, respectively.

Especially, when A is symmetric,

$$\frac{\partial x'Ax}{\partial x} = 2Ax.$$

6. Let A and B be $k \times k$ matrices, and I_k be a $k \times k$ **identity matrix (単位行列)** (one in the diagonal elements and zero in the other elements).

   When $AB = I_k$, B is called the **inverse matrix (逆行列)** of A, denoted by $B = A^{-1}$.

   That is, $AA^{-1} = A^{-1}A = I_k$.

7. Let A be a $k \times k$ matrix and x be a $k \times 1$ vector.

If A is a **positive definite matrix (正値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax > 0.$$

If A is a **positive semidefinite matrix (非負値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax \geq 0.$$

If A is a **negative definite matrix (負値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax < 0.$$

If A is a **negative semidefinite matrix (非正値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax \leq 0.$$

Trace, Rank and etc.: $A : k \times k, \quad B : n \times k, \quad C : k \times n.$

1. The **trace (トレース)** of A is: $\mathrm{tr}(A) = \sum_{i=1}^{k} a_{ii}$, where $A = [a_{ij}]$.

2. The **rank (ランク，階数)** of A is the maximum number of linearly independent column (or row) vectors of A, which is denoted by $\mathrm{rank}(A)$.

3. If A is an **idempotent matrix (べき等行列)**, $A = A^2$.

4. If A is an idempotent and symmetric matrix, $A = A^2 = A'A$.

5. A is idempotent if and only if the eigen values of A consist of 1 and 0.

6. If A is idempotent, $\text{rank}(A) = \text{tr}(A)$.

7. $\text{tr}(BC) = \text{tr}(CB)$

Distributions in Matrix Form:

1. Let X, μ and Σ be $k \times 1$, $k \times 1$ and $k \times k$ matrices.

   When $X \sim N(\mu, \Sigma)$, the density function of X is given by:

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

   $\mathrm{E}(X) = \mu$ and $\mathrm{V}(X) = \mathrm{E}\Big((X-\mu)(X-\mu)'\Big) = \Sigma$

   The moment-generating function: $\phi(\theta) = \mathrm{E}\Big(\exp(\theta' X)\Big) = \exp(\theta'\mu + \frac{1}{2}\theta'\Sigma\theta)$

2. If $X \sim N(\mu, \Sigma)$, then $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2(k)$.

   Note that $X'X \sim \chi^2(k)$ when $X \sim N(0, I_k)$.

3. X: $n \times 1$, Y: $m \times 1$, $X \sim N(\mu_x, \Sigma_x)$, $Y \sim N(\mu_y, \Sigma_y)$

   X is independent of Y, i.e., $\mathrm{E}\Big((X-\mu_x)(Y-\mu_y)'\Big) = 0$ in the case of normal random variables.

$$\frac{(X-\mu_x)'\Sigma_x^{-1}(X-\mu_x)/n}{(Y-\mu_y)'\Sigma_y^{-1}(Y-\mu_y)/m} \sim F(n, m)$$

4. If $X \sim N(0, \sigma^2 I_n)$ and A is a symmetric idempotent $n \times n$ matrix of rank G, then $X'AX/\sigma^2 \sim \chi^2(G)$.

   Note that $X'AX = (AX)'(AX)$ and $\text{rank}(A) = \text{tr}(A)$ because A is idempotent.

5. If $X \sim N(0, \sigma^2 I_n)$, A and B are symmetric idempotent $n \times n$ matrices of rank G and K, and $AB = 0$, then

$$\frac{X'AX}{G\sigma^2} \Big/ \frac{X'BX}{K\sigma^2} = \frac{X'AX/G}{X'BX/K} \sim F(G, K).$$

3 Multiple Regression Model (重回帰モデル)

Up to now, only one independent variable, i.e., x_i, is taken into the regression model.

In this section, we extend it to more independent variables, which is called the **multiple regression model (重回帰モデル)**.

We consider the following regression model:

$$
\begin{aligned}
y_i &= \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_k x_{i,k} + u_i \\
&= (x_{i,1}, x_{i,2}, \cdots, x_{i,k}) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + u_i \\
&= x_i \beta + u_i,
\end{aligned}
$$

for $i = 1, 2, \cdots, n$,

where x_i and β denote a $1 \times k$ vector of the independent variables and a $k \times 1$ vector of the unknown parameters to be estimated, which are represented as:

$$x_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,k}), \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}.$$

$x_{i,j}$ denotes the ith observation of the jth independent variable.

The case of $k = 2$ and $x_{i,1} = 1$ for all i is exactly equivalent to (1).

Therefore, the matrix form above is a generalization of (1).

Writing all the equations for $i = 1, 2, \cdots, n$, we have:

$$
\begin{aligned}
y_1 &= \beta_1 x_{1,1} + \beta_2 x_{1,2} + \cdots + \beta_k x_{1,k} + u_1 = x_1\beta + u_1, \\
y_2 &= \beta_1 x_{2,1} + \beta_2 x_{2,2} + \cdots + \beta_k x_{2,k} + u_2 = x_2\beta + u_2, \\
&\vdots \\
y_n &= \beta_1 x_{n,1} + \beta_2 x_{n,2} + \cdots + \beta_k x_{n,k} + u_n = x_n\beta + u_n,
\end{aligned}
$$

which is rewritten as:

$$\begin{aligned}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} &= \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \\ &= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.\end{aligned}$$

Again, the above equation is compactly rewritten as:

$$y = X\beta + u, \tag{18}$$

where y, X and u are denoted by:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad X = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Utilizing the matrix form (18), we derive the ordinary least squares estimator of β, denoted by $\hat{\beta}$.

In (18), replacing β by $\hat{\beta}$, we have the following equation:

$$y = X\hat{\beta} + e,$$

where e denotes a $n \times 1$ vector of the residuals.

The ith element of e is given by e_i.

The sum of squared residuals is written as follows:

$$\begin{aligned} S(\hat{\beta}) &= \sum_{i=1}^{n} e_i^2 = e'e = (y - X\hat{\beta})'(y - X\hat{\beta}) = (y' - \hat{\beta}'X')(y - X\hat{\beta}) \\ &= y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} = y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}. \end{aligned}$$

In the last equality, note that $\hat{\beta}'X'y = y'X\hat{\beta}$ because both are scalars.

To minimize $S(\hat{\beta})$ with respect to $\hat{\beta}$, we set the first derivative of $S(\hat{\beta})$ equal to zero, i.e.,

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0.$$

Solving the equation above with respect to $\hat{\beta}$, the **ordinary least squares estimator (OLS, 最小自乗推定量)** of β is given by:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{19}$$

Thus, the ordinary least squares estimator is derived in the matrix form.

(*) Remark

The second order condition for minimization:

$$\frac{\partial^2 S(\hat{\beta})}{\partial\hat{\beta}\partial\hat{\beta}'} = 2X'X$$

is a positive definite matrix.

Set $c = Xd$.

For any $d \neq 0$, we have $c'c = d'X'Xd > 0$.

Now, in order to obtain the properties of $\hat{\beta}$ such as mean, variance, distribution and so on, (19) is rewritten as follows:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u. \end{aligned} \tag{20}$$

Taking the expectation on both sides of (20), we have the following:

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}(\beta + (X'X)^{-1}X'u) = \beta + (X'X)^{-1}X'\mathrm{E}(u) = \beta,$$

because of $\mathrm{E}(u) = 0$ by the assumption of the error term u_i.

Thus, unbiasedness of $\hat{\beta}$ is shown.

The variance of $\hat{\beta}$ is obtained as:

$$\begin{aligned}
\mathrm{V}(\hat{\beta}) &= \mathrm{E}((\hat{\beta}-\beta)(\hat{\beta}-\beta)') = \mathrm{E}\Big((X'X)^{-1}X'u((X'X)^{-1}X'u)'\Big) \\
&= \mathrm{E}((X'X)^{-1}X'uu'X(X'X)^{-1}) = (X'X)^{-1}X'\mathrm{E}(uu')X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}.
\end{aligned}$$

The first equality is the definition of variance in the case of vector.

In the fifth equality, $\mathrm{E}(uu') = \sigma^2 I_n$ is used, which implies that $\mathrm{E}(u_i^2) = \sigma^2$ for all i and $\mathrm{E}(u_i u_j) = 0$ for $i \neq j$.

Remember that $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2.

Under normality assumption on the error term u, it is known that the distribution of $\hat{\beta}$ is given by:

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1}).$$

Proof:

First, when $X \sim N(\mu, \Sigma)$, the moment-generating function, i.e., $\phi(\theta)$, is given by:

$$\phi(\theta) \equiv \mathrm{E}\Big(\exp(\theta' X)\Big) = \exp\Big(\theta'\mu + \frac{1}{2}\theta'\Sigma\theta\Big)$$

θ_u: $n \times 1$, u: $n \times 1$, θ_β: $k \times 1$, $\hat{\beta}$: $k \times 1$

The moment-generating function of u, i.e., $\phi_u(\theta_u)$, is:

$$\phi_u(\theta_u) \equiv \mathrm{E}\Big(\exp(\theta_u' u)\Big) = \exp\Big(\frac{\sigma^2}{2}\theta_u'\theta_u\Big),$$

which is $N(0, \sigma^2 I_n)$.

The moment-generating function of $\hat{\beta}$, i.e., $\phi_\beta(\theta_\beta)$, is:

$$\begin{aligned}
\phi_\beta(\theta_\beta) &\equiv \mathrm{E}\Big(\exp(\theta_\beta'\hat{\beta})\Big) = \mathrm{E}\Big(\exp(\theta_\beta'\beta + \theta_\beta'(X'X)^{-1}X'u)\Big) \\
&= \exp(\theta_\beta'\beta)\mathrm{E}\Big(\exp(\theta_\beta'(X'X)^{-1}X'u)\Big) = \exp(\theta_\beta'\beta)\phi_u\Big(\theta_\beta'(X'X)^{-1}X'\Big) \\
&= \exp(\theta_\beta'\beta)\exp\Big(\frac{\sigma^2}{2}\theta_\beta'(X'X)^{-1}\theta_\beta\Big) = \exp\Big(\theta_\beta'\beta + \frac{\sigma^2}{2}\theta_\beta'(X'X)^{-1}\theta_\beta\Big),
\end{aligned}$$

which is equivalent to the normal distribution with mean β and variance $\sigma^2(X'X)^{-1}$.

Note that $\theta_u = X(X'X)^{-1}\theta_\beta$. QED

Taking the jth element of $\hat{\beta}$, its distribution is given by:

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 a_{jj}), \qquad \text{i.e.,} \qquad \frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{a_{jj}}} \sim N(0,1),$$

where a_{jj} denotes the jth diagonal element of $(X'X)^{-1}$.

Replacing σ^2 by its estimator s^2, we have the following t distribution:

$$\frac{\hat{\beta}_j - \beta_j}{s\sqrt{a_{jj}}} \sim t(n-k),$$

where $t(n-k)$ denotes the t distribution with $n-k$ degrees of freedom.

s^2 is taken as follows:

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-k}e'e = \frac{1}{n-k}(y - X\hat{\beta})'(y - X\hat{\beta}),$$

which leads to an unbiased estimator of σ^2.

Proof:

Substitute $y = X\beta + u$ and $\hat{\beta} = \beta + (X'X)^{-1}X'u$ into $e = y - X\hat{\beta}$.

$$\begin{aligned} e = y - X\hat{\beta} &= X\beta + u - X(\beta + (X'X)^{-1}X'u) \\ &= u - X(X'X)^{-1}X'u = (I_n - X(X'X)^{-1}X')u \end{aligned}$$

$I_n - X(X'X)^{-1}X'$ is idempotent and symmetric, because we have:

$$(I_n - X(X'X)^{-1}X')(I_n - X(X'X)^{-1}X') = I_n - X(X'X)^{-1}X,'$$
$$(I_n - X(X'X)^{-1}X')' = I_n - X(X'X)^{-1}X'.$$

s^2 is rewritten as follows:

$$\begin{aligned} s^2 &= \frac{1}{n-k}e'e = \frac{1}{n-k}((I_n - X(X'X)^{-1}X')u)'(I_n - X(X'X)^{-1}X')u \\ &= \frac{1}{n-k}u'(I_n - X(X'X)^{-1}X')'(I_n - X(X'X)^{-1}X')u \\ &= \frac{1}{n-k}u'(I_n - X(X'X)^{-1}X')u \end{aligned}$$

Take the expectation of $u'(I_n - X(X'X)^{-1}X')u$ and note that $\mathrm{tr}(a) = a$ for a scalar a.

$$\begin{aligned}
\mathrm{E}(s^2) &= \frac{1}{n-k}\mathrm{E}\Big(\mathrm{tr}\Big(u'(I_n - X(X'X)^{-1}X')u\Big)\Big) = \frac{1}{n-k}\mathrm{E}\Big(\mathrm{tr}\Big((I_n - X(X'X)^{-1}X')uu'\Big)\Big) \\
&= \frac{1}{n-k}\mathrm{tr}\Big((I_n - X(X'X)^{-1}X')\mathrm{E}(uu')\Big) = \frac{1}{n-k}\sigma^2\mathrm{tr}\Big((I_n - X(X'X)^{-1}X')I_n\Big) \\
&= \frac{1}{n-k}\sigma^2\mathrm{tr}(I_n - X(X'X)^{-1}X') = \frac{1}{n-k}\sigma^2(\mathrm{tr}(I_n) - \mathrm{tr}(X(X'X)^{-1}X')) \\
&= \frac{1}{n-k}\sigma^2(\mathrm{tr}(I_n) - \mathrm{tr}((X'X)^{-1}X'X)) = \frac{1}{n-k}\sigma^2(\mathrm{tr}(I_n) - \mathrm{tr}(I_k)) \\
&= \frac{1}{n-k}\sigma^2(n-k) = \sigma^2
\end{aligned}$$

$\longrightarrow$ s^2 is an unbiased estimator of σ^2.

Note that we do not need normality assumption for unbiasedness of s^2.

Trace (トレース):

1. A: $n \times n$, $\quad \mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$, where a_{ij} denotes an element in the ith row and the jth column of a matrix A.

2. a: scalar (1×1), $\quad \mathrm{tr}(a) = a$

3. A: $n \times k$, B: $k \times n$, $\quad \mathrm{tr}(AB) = \mathrm{tr}(BA)$

4. $\mathrm{tr}(X(X'X)^{-1}X') = \mathrm{tr}((X'X)^{-1}X'X) = \mathrm{tr}(I_k) = k$

5. When X is a vector of random variables, $\mathrm{E}(\mathrm{tr}(X)) = \mathrm{tr}(\mathrm{E}(X))$

Under normality assumption for u, the distribution of s^2 is:

$$\frac{(n-k)s^2}{\sigma^2} = \frac{u'(I_n - X(X'X)^{-1}X')u}{\sigma^2} \sim \chi^2(\mathrm{tr}(I_n - X(X'X)^{-1}X'))$$

Note that $\mathrm{tr}(I_n - X(X'X)^{-1}X') = n - k$, because

$$\mathrm{tr}(I_n) = n$$

$$\mathrm{tr}(X(X'X)^{-1}X') = \mathrm{tr}((X'X)^{-1}X'X) = \mathrm{tr}(I_k) = k$$

[Review]

- $X'X \sim \chi^2(n)$ for $X \sim N(0, I_n)$.

- $\dfrac{X'X}{\sigma^2} \sim \chi^2(n)$ for $X \sim N(0, \sigma^2 I_n)$.

- $\dfrac{X'AX}{\sigma^2} \sim \chi^2(G)$, where $X \sim N(0, \sigma^2 I_n)$ and A is a symmetric idempotent $n \times n$ matrix of rank $G \leq n$.

Remember that $G = \text{Rank}(A) = \text{tr}(A)$ when A is symmetric and idempotent.

[End of Review]

Asymptotic Normality (without normality assumption on u): Using the central limit theorem, without normality assumption we can show that as $n \longrightarrow \infty$, under the condition of $\frac{1}{n}X'X \longrightarrow M$ we have the following result:

$$\frac{\hat{\beta}_j - \beta_j}{s\sqrt{a_{jj}}} \longrightarrow N(0,1),$$

where M denotes a $k \times k$ constant matrix.

Thus, we can construct the confidence interval and the testing procedure, using the t distribution under the normality assumption or the normal distribution without the normality assumption.

4 Properties of OLSE

1. Properties of $\hat{\beta}$: **BLUE (best linear unbiased estimator，最良線形不偏推定量)**, i.e., minimum variance within the class of linear unbiased estimators **(Gauss-Markov theorem，ガウス・マルコフの定理)**

   Proof:

   Consider another linear unbiased estimator, which is denoted by $\tilde{\beta} = Cy$.

   $$\tilde{\beta} = Cy = C(X\beta + u) = CX\beta + Cu,$$

   where C is a $k \times n$ matrix.

Taking the expectation of $\tilde{\beta}$, we obtain:

$$\mathrm{E}(\tilde{\beta}) = CX\beta + C\mathrm{E}(u) = CX\beta$$

Because we have assumed that $\tilde{\beta} = Cy$ is unbiased, $\mathrm{E}(\tilde{\beta}) = \beta$ holds.

That is, we need the condition: $CX = I_k$.

Next, we obtain the variance of $\tilde{\beta} = Cy$.

$$\tilde{\beta} = C(X\beta + u) = \beta + Cu.$$

Therefore, we have:

$$\mathrm{V}(\tilde{\beta}) = \mathrm{E}((\tilde{\beta} - \beta)(\tilde{\beta} - \beta)') = \mathrm{E}(Cuu'C') = \sigma^2 CC'$$

Defining $C = D + (X'X)^{-1}X'$, $\mathrm{V}(\tilde{\beta})$ is rewritten as:

$$\mathrm{V}(\tilde{\beta}) = \sigma^2 CC' = \sigma^2 (D + (X'X)^{-1}X')(D + (X'X)^{-1}X')'.$$

Moreover, because $\hat{\beta}$ is unbiased, we have the following:

$$CX = I_k = (D + (X'X)^{-1}X')X = DX + I_k.$$

Therefore, we have the following condition:

$$DX = 0.$$

Accordingly, $\mathrm{V}(\tilde{\beta})$ is rewritten as:

$$\begin{aligned}\mathrm{V}(\tilde{\beta}) &= \sigma^2 CC' = \sigma^2 (D + (X'X)^{-1}X')(D + (X'X)^{-1}X')' \\ &= \sigma^2 (X'X)^{-1} + \sigma^2 DD' = \mathrm{V}(\hat{\beta}) + \sigma^2 DD'\end{aligned}$$

Thus, $\mathrm{V}(\tilde{\beta}) - \mathrm{V}(\hat{\beta})$ is a positive definite matrix.

$\Longrightarrow \mathrm{V}(\tilde{\beta}_i) - \mathrm{V}(\hat{\beta}_i) > 0$

$\Longrightarrow \hat{\beta}$ is a minimum variance (i.e., best) linear unbiased estimator of β.

Note as follows:

$\Longrightarrow$ A is positive definite when $d'Ad > 0$ except $d = 0$.

$\Longrightarrow$ The ith diagonal element of A, i.e., a_{ii}, is positive (choose d such that the ith element of d is one and the other elements are zeros).

F Distribution (H_0 : $\beta = 0$):

1. If $u \sim N(0, \sigma^2 I_n)$, then $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$.

   Therefore, $\dfrac{(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)}{\sigma^2} \sim \chi^2(k)$.

2. **Proof:**

   Using $\hat{\beta}-\beta = (X'X)^{-1}X'u$, we obtain:

$$
\begin{aligned}
(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta) &= ((X'X)^{-1}X'u)'X'X(X'X)^{-1}X'u \\
&= u'X(X'X)^{-1}X'X(X'X)^{-1}X'u = u'X(X'X)^{-1}X'u
\end{aligned}
$$

Note that $X(X'X)^{-1}X'$ is symmetric and idempotent, i.e., $A'A = A$.

$$\frac{u'X(X'X)^{-1}X'u}{\sigma^2} \sim \chi^2\Big(\mathrm{tr}(X(X'X)^{-1}X')\Big)$$

The degree of freedom is given by:

$$\mathrm{tr}(X(X'X)^{-1}X') = \mathrm{tr}((X'X)^{-1}X'X) = \mathrm{tr}(I_k) = k$$

Therefore, we obtain:

$$\frac{u'X(X'X)^{-1}X'u}{\sigma^2} \sim \chi^2(k)$$

3. (*) Formula:

Suppose that $X \sim N(0, I_k)$.

If A is symmetric and idempotent, i.e., $A'A = A$, then $X'AX \sim \chi^2(\mathrm{tr}(A))$.

Here, $X = \dfrac{1}{\sigma}u \sim N(0, I_n)$ from $u \sim N(0, \sigma^2 I_n)$, and $A = X(X'X)^{-1}X'$.

4. **Sum of Residuals:** e is rewritten as:

$$e = (I_n - X(X'X)^{-1}X')u.$$

Therefore, the sum of residuals is given by:

$$e'e = u'(I_n - X(X'X)^{-1}X')u.$$

Note that $I_n - X(X'X)^{-1}X'$ is symmetric and idempotent.

We obtain the following result:

$$\frac{e'e}{\sigma^2} = \frac{u'(I_n - X(X'X)^{-1}X')u}{\sigma^2} \sim \chi^2\Big(\mathrm{tr}(I_n - X(X'X)^{-1}X')\Big),$$

where the trace is:

$$\mathrm{tr}(I_n - X(X'X)^{-1}X') = n - k.$$

Therefore, we have the following result:

$$\frac{e'e}{\sigma^2} = \frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k),$$

where

$$s^2 = \frac{1}{n-k}e'e.$$

5. We show that $\hat{\beta}$ is independent of e.

Proof:

Because $u \sim N(0, \sigma^2 I_n)$, we show that $\mathrm{Cov}(e, \hat{\beta}) = 0$.

$$\begin{aligned}
\mathrm{Cov}(e, \hat{\beta}) &= \mathrm{E}(e(\hat{\beta} - \beta)') = \mathrm{E}\Big((I_n - X(X'X)^{-1}X')u((X'X)^{-1}X'u)'\Big) \\
&= \mathrm{E}\Big((I_n - X(X'X)^{-1}X')uu'X(X'X)^{-1}\Big) = (I_n - X(X'X)^{-1}X')\mathrm{E}(uu')X(X'X)^{-1} \\
&= (I_n - X(X'X)^{-1}X')(\sigma^2 I_n)X(X'X)^{-1} = \sigma^2(I_n - X(X'X)^{-1}X')X(X'X)^{-1} \\
&= \sigma^2(X(X'X)^{-1} - X(X'X)^{-1}X'X(X'X)^{-1}) = \sigma^2(X(X'X)^{-1} - X(X'X)^{-1}) = 0.
\end{aligned}$$

Therefore, $\hat{\beta}$ is independent of e.

6. Therefore, we obtain the following distribution:

$$\frac{(\hat\beta-\beta)'X'X(\hat\beta-\beta)}{\sigma^2} = \frac{u'X(X'X)^{-1}X'u}{\sigma^2} \sim \chi^2(k),$$

$$\frac{e'e}{\sigma^2} = \frac{u'(I_n - X(X'X)^{-1}X')u}{\sigma^2} \sim \chi^2(n-k)$$

$\hat\beta$ is independent of e.

Accordingly, we can derive:

$$\frac{\dfrac{(\hat\beta-\beta)'X'X(\hat\beta-\beta)}{\sigma^2}\Big/k}{\dfrac{e'e}{\sigma^2}\Big/(n-k)} = \frac{(\hat\beta-\beta)'X'X(\hat\beta-\beta)/k}{s^2} \sim F(k, n-k)$$

Note as follows:

$$\frac{\dfrac{(\hat\beta-\beta)'X'X(\hat\beta-\beta)}{\sigma^2}\Big/k}{\dfrac{e'e}{\sigma^2}\Big/(n-k)} = \frac{u'X(X'X)^{-1}X'u/k}{u'(I_n - X(X'X)^{-1}X')u/(n-k)} \sim F(k, n-k),$$

because $X(X'X)^{-1}X'(I_n - X(X'X)^{-1}X') = 0$.

Under the null hypothesis $H_0:\ \beta = 0$, $\dfrac{\hat\beta'X'X\hat\beta/k}{s^2} \sim F(k, n-k)$.

Given data, $\dfrac{\hat\beta'X'X\hat\beta/k}{s^2}$ is compared with $F(k, n-k)$.

If $\dfrac{\hat\beta'X'X\hat\beta/k}{s^2}$ is in tha tail of the F distribution, the null hypothesis is rejected.

(*) Formulas — Review — :

- $\dfrac{U/n}{V/m} \sim F(n,m)$ when U $sim\chi^2(n)$, $V \sim \chi^2(m)$, and U is independent of V.

- When $X \sim N(0, I_n)$, A and B are $n \times n$ symmetric idempotent matrices, $\text{Rank}(A) = \text{tr}(A) = G$, $\text{Rank}(B) = \text{tr}(B) = K$ and $AB = 0$, then $\dfrac{X'AX/G}{X'BX/K} \sim F(G,K)$.

Note that the covariance of AX and BX is zero, which implies that AX is independent of BX under normality of X.

Coefficient of Determination (決定係数), R^2:

1. Definition of the Coefficient of Determination, R^2: $R^2 = 1 - \dfrac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \overline{y})^2}$

2. Numerator: $\displaystyle\sum_{i=1}^n e_i^2 = e'e$

3. Denominator: $\displaystyle\sum_{i=1}^n (y_i - \overline{y})^2 = y'(I_n - \frac{1}{n}ii')'(I_n - \frac{1}{n}ii')y = y'(I_n - \frac{1}{n}ii')y$

(*) Remark

$$\begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix} = y - \frac{1}{n} ii'y = (I_n - \frac{1}{n} ii')y,$$

where $i = (1, 1, \cdots, 1)'$.

4. In a matrix form, we can rewrite as: $R^2 = 1 - \dfrac{e'e}{y'(I_n - \frac{1}{n}ii')y}$

F Distribution and Coefficient of Determination:

$\Longrightarrow$ This will be discussed later.

Testing Linear Restrictions (F Distribution):

1. If $u \sim N(0, \sigma^2 I_n)$, then $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$.

   Consider testing the hypothesis $H_0 : R\beta = r$.

   $R : \; G \times k, \qquad \text{rank}(R) = G \leq k.$

   $R\hat{\beta} \sim N(R\beta, \sigma^2 R(X'X)^{-1}R')$.

Therefore, $\dfrac{(R\hat{\beta}-r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta}-r)}{\sigma^2} \sim \chi^2(G)$.

Note that $R\beta = r$.

(a) When $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$, the mean of $R\hat{\beta}$ is:

$$\mathrm{E}(R\hat{\beta}) = R\mathrm{E}(\hat{\beta}) = R\beta.$$

(b) When $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$, the variance of $R\hat{\beta}$ is:

$$\begin{aligned}\mathrm{V}(R\hat{\beta}) &= \mathrm{E}((R\hat{\beta}-R\beta)(R\hat{\beta}-R\beta)') = \mathrm{E}(R(\hat{\beta}-\beta)(\hat{\beta}-\beta)'R') \\ &= R\mathrm{E}((\hat{\beta}-\beta)(\hat{\beta}-\beta)')R' = R\mathrm{V}(\hat{\beta})R' = \sigma^2 R(X'X)^{-1}R'.\end{aligned}$$

2. We know that $\dfrac{(n-k)s^2}{\sigma^2} = \dfrac{e'e}{\sigma^2} = \dfrac{(y - X\hat{\beta})'(y - X\hat{\beta})}{\sigma^2} \sim \chi^2(n-k)$.

3. Under normality assumption on u, $\hat{\beta}$ is independent of e.

4. Therefore, we have the following distribution:

$$\frac{(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/G}{(y - X\hat{\beta})'(y - X\hat{\beta})/(n-k)} \sim F(G, n-k)$$

5. **Some Examples:**

(a) t Test:

The case of $G = 1$, $r = 0$ and $R = (0, \cdots, 1, \cdots, 0)$ (the ith element of R is one and the other elements are zero):

The test of $H_0:\ \beta_i = 0$ is given by:

$$\frac{(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/G}{s^2} = \frac{\hat{\beta}_i^2}{s^2 a_{ii}} \sim F(1, n-k),$$

where $s^2 = e'e/(n-k)$, $R\hat{\beta} = \hat{\beta}_i$ and

$$a_{ii} = R(X'X)^{-1}R' = \text{ the } i \text{ row and } i\text{th column of } (X'X)^{-1}.$$

*) Recall that $Y \sim F(1, m)$ when $X \sim t(m)$ and $Y = X^2$.

Therefore, the test of $H_0: \beta_i = 0$ is given by:

$$\frac{\hat{\beta}_i}{s\sqrt{a_{ii}}} \sim t(n-k).$$

(b) Test of structural change (Part 1):

$$y_i = \begin{cases} x_i\beta_1 + u_i, & i = 1, 2, \cdots, m \\ x_i\beta_2 + u_i, & i = m+1, m+2, \cdots, n \end{cases}$$

Assume that $u_i \sim N(0, \sigma^2)$.

In a matrix form,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \\ y_{m+1} \\ y_{m+2} \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & 0 \\ x_2 & 0 \\ \vdots & \vdots \\ x_m & 0 \\ 0 & x_{m+1} \\ 0 & x_{m+2} \\ \vdots & \vdots \\ 0 & x_n \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \\ u_{m+1} \\ u_{m+2} \\ \vdots \\ u_n \end{pmatrix}$$

Moreover, rewriting,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + u$$

Again, rewriting,

$$Y = X\beta + u$$

The null hypothesis is $H_0 : \beta_1 = \beta_2$.

Apply the F test, using $R = (I_k \quad -I_k)$ and $r = 0$.

In this case, $G = \text{rank}(R) = k$ and β is a $2k \times 1$ vector.

The distribution is $F(k, n - 2k)$.

(c) The hypothesis in which sum of the 1st and 2nd coefficients is equal to one:

$R = (1, 1, 0, \cdots, 0)$, $r = 1$

In this case, $G = \text{rank}(R) = 1$

The distribution of the test statistic is $F(1, n-k)$.

(d) Testing seasonality:

In the case of **quarterly data (四半期データ)**, the regression model is:

$$y = \alpha + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + X\beta_0 + u$$

$D_j = 1$ in the jth quarter and 0 otherwise, i.e., D_j, $j = 1, 2, 3$, are seasonal dummy variables.

Testing seasonality $\Longrightarrow H_0:\ \alpha_1 = \alpha_2 = \alpha_3 = 0$

$$\beta = \begin{pmatrix} \alpha \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_0 \end{pmatrix}, \qquad R = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}, \qquad r = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

In this case, $G = \text{rank}(R) = 3$, and β is a $k \times 1$ vector.

The distribution of the test statistic is $F(3, n-k)$.

(e) Cobb-Douglas Production Function:

Let Q_i, K_i and L_i be production, capital stock and labor.

We estimate the following production function:

$$\log(Q_i) = \beta_1 + \beta_2 \log(K_i) + \beta_3 \log(L_i) + u_i.$$

We test a linear homogeneous (一次同次) production function.

The null and alternative hypotheses are:

$$H_0:\ \beta_2 + \beta_3 = 1,$$

$$H_1:\ \beta_2 + \beta_3 \neq 1.$$

Then, set as follows:

$$R = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}, \qquad r = 1.$$

(f) Test of structural change (Part 2):

Test the structural change between time periods m and $m+1$.

In the case where both the constant term and the slope are changed, the regression model is as follows:

$$y_i = \alpha + \beta x_i + \gamma d_i + \delta d_i x_i + u_i,$$

where

$$d_i = \begin{cases} 0, & \text{for } i = 1, 2, \cdots, m, \\ 1, & \text{for } i = m+1, m+2, \cdots, n. \end{cases}$$

We consider testing the structural change at time $m + 1$.

The null and alternative hypotheses are as follows:

$$H_0: \ \gamma = \delta = 0,$$

$$H_1: \ \gamma \neq 0, \text{ or, } \delta \neq 0.$$

Then, set as follows:

$$R = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

(g) Multiple regression model:

Consider the case of two explanatory variables:

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i.$$

We want to test the hypothesis that neither x_i nor z_i depends on y_i.

In this case, the null and alternative hypotheses are as follows:

$$H_0: \ \beta = \gamma = 0,$$

$$H_1: \ \beta \neq 0, \ \text{or}, \ \gamma \neq 0.$$

Then, set as follows:

$$R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Coefficient of Determination R^2 and F distribution:

● The regression model:

$$y_i = x_i\beta + u_i = \beta_1 + x_{2i}\beta_2 + u_i$$

where

$$x_i = \begin{pmatrix} 1 & x_{2i} \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix},$$

$x_i :\ 1 \times k, \qquad x_{2i} :\ 1 \times (k-1), \qquad \beta :\ k \times 1, \qquad \beta_2 :\ (k-1) \times 1$

Define:

$$X_2 = \begin{pmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{pmatrix}$$

Then,

$$y = X\beta + u = (i \quad X_2)\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + u = i\beta_1 + X_2\beta_2 + u,$$

where the first column of X corresponds to a constant term, i.e.,

$$X = \begin{pmatrix} i & X_2 \end{pmatrix}, \qquad i = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

● Consider testing $H_0: \ \beta_2 = 0$.

The F distribution is set as follows:

$$R = \begin{pmatrix} 0 & I_{k-1} \end{pmatrix}, \qquad r = 0$$

where R is a $(k-1) \times k$ matrix and r is a $(k-1) \times 1$ vector.

$$\frac{(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/(k-1)}{e'e/(n-k)} \sim F(k-1, n-k)$$

We are going to show:

$$(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) = \hat{\beta}_2' X_2' M X_2 \hat{\beta}_2,$$

where $M = I_n - \frac{1}{n} ii'$.

Note that M is symmetric and idempotent, i.e., $M'M = M$.

$$\begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} = My$$

$R(X'X)^{-1}R'$ is given by:

$$\begin{aligned} R(X'X)^{-1}R' &= \begin{pmatrix} 0 & I_{k-1} \end{pmatrix} \left(\begin{pmatrix} i' \\ X_2' \end{pmatrix} \begin{pmatrix} i & X_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} 0 \\ I_{k-1} \end{pmatrix} \\ &= \begin{pmatrix} 0 & I_{k-1} \end{pmatrix} \begin{pmatrix} i'i & i'X_2 \\ X_2'i & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ I_{k-1} \end{pmatrix} \end{aligned}$$

(*) The inverse of a partitioned matrix:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_{11} and A_{22} are square nonsingular matrices.

$$A^{-1} = \begin{pmatrix} B_{11} & -B_{11}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}B_{11} & A_{22}^{-1} + A_{22}^{-1}A_{21}B_{11}A_{12}A_{22}^{-1} \end{pmatrix},$$

where $B_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$, or alternatively,

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}B_{22}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}B_{22} \\ -B_{22}A_{21}A_{11}^{-1} & B_{22} \end{pmatrix},$$

where $B_{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$

Go back to the F distribution.

$$\begin{pmatrix} i'i & i'X_2 \\ X_2'i & X_2'X_2 \end{pmatrix}^{-1} = \begin{pmatrix} \cdot & \cdots \\ \vdots & (X_2'X_2 - X_2'i(i'i)^{-1}i'X_2)^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} \cdot & \cdots \\ \vdots & (X_2'(I_n - \frac{1}{n}ii')X_2)^{-1} \end{pmatrix} = \begin{pmatrix} \cdot & \cdots \\ \vdots & (X_2'MX_2)^{-1} \end{pmatrix}$$

Therefore, we obtain:

$$\begin{pmatrix} 0 & I_{k-1} \end{pmatrix} \begin{pmatrix} i'i & i'X_2 \\ X_2'i & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ I_{k-1} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & I_{k-1} \end{pmatrix} \begin{pmatrix} \cdot & \cdots \\ \vdots & (X_2'MX_2)^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ I_{k-1} \end{pmatrix} = (X_2'MX_2)^{-1}.$$

Thus, under $H_0: \beta_2 = 0$, we obtain the following result:

$$\frac{(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/(k-1)}{e'e/(n-k)}$$

$$= \frac{\hat{\beta}_2' X_2' M X_2 \hat{\beta}_2/(k-1)}{e'e/(n-k)} \sim F(k-1, n-k)$$

● Coefficient of Determination R^2:

Define e as $e = y - X\hat{\beta}$. The coefficient of determinant, R^2, is

$$R^2 = 1 - \frac{e'e}{y'My},$$

where $M = I_n - \frac{1}{n}ii'$, I_n is a $n \times n$ identity matrix and i is a $n \times 1$ vector consisting of 1, i.e., $i = (1, 1, \cdots, 1)'$.

$$Me = My - MX\hat{\beta}.$$

When $X = \begin{pmatrix} i & X_2 \end{pmatrix}$ and $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$,

$$Me = e,$$

because $i'e = 0$, and

$$MX = M\begin{pmatrix} i & X_2 \end{pmatrix} = \begin{pmatrix} Mi & MX_2 \end{pmatrix} = \begin{pmatrix} 0 & MX_2 \end{pmatrix}$$

because $Mi = 0$.

$$MX\hat{\beta} = \begin{pmatrix} 0 & MX_2 \end{pmatrix}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = MX_2\hat{\beta}_2$$

Thus,

$$My = MX\hat{\beta} + Me \qquad \Longrightarrow \qquad My = MX_2\hat{\beta}_2 + e$$

Therefore, $y'My$ is given by: $\qquad y'My = \hat{\beta}_2'X_2'MX_2\hat{\beta}_2 + e'e,$

because $X_2'e = 0$ and $Me = e$.

The coefficient of determinant, R^2, is rewritten as:

$$R^2 = 1 - \frac{e'e}{y'My} \qquad \Longrightarrow \qquad e'e = (1 - R^2)y'My$$

$$R^2 = \frac{y'My - e'e}{y'My} = \frac{\hat{\beta}_2' X_2' M X_2 \hat{\beta}_2}{y'My} \qquad \Longrightarrow \qquad \hat{\beta}_2' X_2' M X_2 \hat{\beta}_2 = R^2 y'My$$

Therefore,

$$\begin{aligned} \frac{\hat{\beta}_2' X_2' M X_2 \hat{\beta}_2/(k-1)}{e'e/(n-k)} &= \frac{R^2 y'My/(k-1)}{(1-R^2)y'My/(n-k)} \\ &= \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k) \end{aligned}$$

Thus, using R^2, the null hypothesis $H_0 : \beta_2 = 0$ is easily tested.

5 Restricted OLS (制約付き最小二乗法)

1. Let $\tilde{\beta}$ be the restricted estimator.

   Consider the linear restriction: $R\beta = r$.

2. Minimize $(y - X\tilde{\beta})'(y - X\tilde{\beta})$ subject to $R\tilde{\beta} = r$.

   Let L be the Lagrangian for the minimization problem.

$$L = (y - X\tilde{\beta})'(y - X\tilde{\beta}) - 2\tilde{\lambda}'(R\tilde{\beta} - r)$$

Because $\tilde{\beta}$ and $\tilde{\lambda}$ minimize the Lagrangian L,

$$\frac{\partial L}{\partial \tilde{\beta}} = -2X'(y - X\tilde{\beta}) - 2R'\tilde{\lambda} = 0$$
$$\frac{\partial L}{\partial \tilde{\lambda}} = -2(R\tilde{\beta} - r) = 0.$$

(*) Remember that $\dfrac{\partial a'x}{\partial x} = a$ and $\dfrac{\partial x'Ax}{\partial x} = (A + A')x$.

From $\dfrac{\partial L}{\partial \tilde{\beta}} = 0$, we obtain:

$$\tilde{\beta} = (X'X)^{-1}X'y + (X'X)^{-1}R'\tilde{\lambda} = \hat{\beta} + (X'X)^{-1}R'\tilde{\lambda}.$$

Multiplying R from the left, we have:

$$R\tilde{\beta} = R\hat{\beta} + R(X'X)^{-1}R'\tilde{\lambda}.$$

Because $R\tilde{\beta} = r$ has to be satisfied, we have the following expression:

$$r = R\hat{\beta} + R(X'X)^{-1}R'\tilde{\lambda}.$$

Therefore, solving the above equation with respect to $\tilde{\lambda}$, we obtain:

$$\tilde{\lambda} = \Big(R(X'X)^{-1}R'\Big)^{-1}(r - R\hat{\beta})$$

Substituting $\tilde{\lambda}$ into $\tilde{\beta} = \hat{\beta} + (X'X)^{-1}R'\tilde{\lambda}$, the restricted OLSE is given by:

$$\tilde{\beta} = \hat{\beta} + (X'X)^{-1}R'\Big(R(X'X)^{-1}R'\Big)^{-1}(r - R\hat{\beta}).$$

(a) The expectation of $\tilde{\beta}$ is:

$$\begin{aligned} \mathrm{E}(\tilde{\beta}) &= \mathrm{E}(\hat{\beta}) + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\mathrm{E}(\hat{\beta})) \\ &= \beta + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) \\ &= \beta, \end{aligned}$$

because of $R\beta = r$.

Thus, it is shown that $\tilde{\beta}$ is unbiased.

(b) The variance of $\tilde{\beta}$ is as follows.

First, rewrite as follows:

$$
\begin{aligned}
(\tilde{\beta}-\beta) &= (\hat{\beta}-\beta) + (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(R\beta - R\hat{\beta}) \\
&= (\hat{\beta}-\beta) - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - R\beta) \\
&= (\hat{\beta}-\beta) - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(\hat{\beta} - \beta) \\
&= \left(I_k - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R\right)(\hat{\beta} - \beta) \\
&= W(\hat{\beta}-\beta),
\end{aligned}
$$

where $W \equiv I_k - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R$.

Then, we obtain the following variance:

$$
\begin{aligned}
\mathrm{V}(\tilde\beta) \equiv \mathrm{E}((\tilde\beta-\beta)(\tilde\beta-\beta)') &= \mathrm{E}(W(\hat\beta-\beta)(\hat\beta-\beta)'W') \\
&= W\mathrm{E}((\hat\beta-\beta)(\hat\beta-\beta)')W' = W\mathrm{V}(\hat\beta)W' = \sigma^2 W(X'X)^{-1}W' \\
&= \sigma^2\Big(I-(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)^{-1}R\Big)(X'X)^{-1} \\
&\qquad\qquad \times\Big(I-(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)^{-1}R\Big)' \\
&= \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)^{-1}R(X'X)^{-1} \\
&= \mathrm{V}(\hat\beta) - \sigma^2(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)^{-1}R(X'X)^{-1}
\end{aligned}
$$

Thus, $\mathrm{V}(\hat\beta)-\mathrm{V}(\tilde\beta)$ is positive definite.

3. Another solution:

Again, write the first-order condition for minimization:

$$\frac{\partial L}{\partial \tilde{\beta}} = -2X'(y - X\tilde{\beta}) - 2R'\tilde{\lambda} = 0,$$
$$\frac{\partial L}{\partial \tilde{\lambda}} = -2(R\tilde{\beta} - r) = 0,$$

which can be written as:

$$X'X\tilde{\beta} - R'\tilde{\lambda} = X'y,$$
$$R\tilde{\beta} = r.$$

Using the matrix form:

$$\begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix} \begin{pmatrix} \tilde{\beta} \\ -\tilde{\lambda} \end{pmatrix} = \begin{pmatrix} X'y \\ r \end{pmatrix}.$$

The solutions of $\tilde{\beta}$ and $-\tilde{\lambda}$ are given by:

$$\begin{pmatrix} \tilde{\beta} \\ -\tilde{\lambda} \end{pmatrix} = \begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ r \end{pmatrix}.$$

(*) Formula to the inverse matrix:

$$\begin{pmatrix} A & B \\ B' & D \end{pmatrix}^{-1} = \begin{pmatrix} E & F \\ F' & G \end{pmatrix},$$

where E, F and G are given by:

$$E = (A - BD^{-1}B')^{-1} = A^{-1} + A^{-1}B(D - B'A^{-1}B)^{-1}B'A^{-1}$$

$$F = -(A - BD^{-1}B')^{-1}BD^{-1} = -A^{-1}B(D - B'A^{-1}B)^{-1}$$

$$G = (D - B'A^{-1}B)^{-1} = D^{-1} + D^{-1}B'(A - BD^{-1}B')^{-1}BD^{-1}$$

In this case, E and F correspond to:

$$E = (X'X)^{-1} - (X'X)^{-1}R'\Big(R(X'X)^{-1}R'\Big)^{-1}R(X'X)^{-1}$$
$$F = (X'X)^{-1}R'\Big(R(X'X)^{-1}R'\Big)^{-1}.$$

Therefore, $\tilde{\beta}$ is derived as follows:

$$\begin{aligned}\tilde{\beta} &= EX'y + Fr \\ &= \hat{\beta} + (X'X)^{-1}R'\Big(R(X'X)^{-1}R'\Big)^{-1}(r - R\hat{\beta}).\end{aligned}$$

The variance is:

$$\mathrm{V}\begin{pmatrix} \tilde{\beta} \\ -\tilde{\lambda} \end{pmatrix} = \sigma^2 \begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix}^{-1}.$$

Therefore, $\mathrm{V}(\tilde{\beta})$ is:

$$\mathrm{V}(\tilde{\beta}) = \sigma^2 E = \sigma^2\Big((X'X)^{-1} - (X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)^{-1}R(X'X)^{-1}\Big)$$

Under the restriction: $R\beta = r$,

$$\mathrm{V}(\hat{\beta}) - \mathrm{V}(\tilde{\beta}) = \sigma^2(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)^{-1}R(X'X)^{-1}$$

is positive definite.

6 *F* Distribution (Restricted and Unrestricted OLSs)

1. As mentioned above, under the null hypothesis $H_0 : R\beta = r$,

$$\frac{(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/G}{(y - X\hat{\beta})'(y - X\hat{\beta})/(n-k)} \sim F(G, n-k),$$

   where $G = \mathrm{Rank}(R)$.

   Using $\tilde{\beta} = \hat{\beta} + (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(r - R\hat{\beta})$, the numerator is rewritten as follows:

$$(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) = (\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta}).$$

Moreover, the denominator is rewritten as follows:

$$
\begin{aligned}
(y - X\tilde{\beta})'(y - X\tilde{\beta}) =& (y - X\hat{\beta} - X(\tilde{\beta} - \hat{\beta}))'(y - X\hat{\beta} - X(\tilde{\beta} - \hat{\beta})) \\
=& (y - X\hat{\beta})'(y - X\hat{\beta}) + (\tilde{\beta} - \hat{\beta})'X'X(\tilde{\beta} - \hat{\beta}) \\
& - (y - X\hat{\beta})'X(\tilde{\beta} - \hat{\beta}) - (\tilde{\beta} - \hat{\beta})'X'(y - X\hat{\beta}) \\
=& (y - X\hat{\beta})'(y - X\hat{\beta}) + (\tilde{\beta} - \hat{\beta})'X'X(\tilde{\beta} - \hat{\beta}).
\end{aligned}
$$

$X'(y - X\hat{\beta}) = X'e = 0$ is utilized.

Summarizing, we have following representation:

$$
\begin{aligned}
(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) &= (\tilde{\beta} - \hat{\beta})'X'X(\tilde{\beta} - \hat{\beta}) \\
&= (y - X\tilde{\beta})'(y - X\tilde{\beta}) - (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= \tilde{u}'\tilde{u} - e'e,
\end{aligned}
$$

where e and $\tilde{u}$ are the restricted residual and the unrestricted residual, i.e., $e = y - X\hat{\beta}$ and $\tilde{u} = y - X\tilde{\beta}$.

Therefore, we obtain the following result:

$$
\frac{(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/G}{(y - X\hat{\beta})'(y - X\hat{\beta})/(n-k)} = \frac{(\tilde{u}'\tilde{u} - e'e)/G}{e'e/(n-k)} \sim F(G, n-k).
$$

7 Example: F Distribution (Restricted OLS and Unrestricted OLS)

Date file $\Longrightarrow$ cons99.txt (Next slide)

Each column denotes year, nominal household expenditures (家計消費，10 billion yen), household disposable income (家計可処分所得，10 billion yen) and household expenditure deflator (家計消費デフレータ，1990=100) from the left.

1955	5430.1	6135.0	18.1
1956	5974.2	6828.4	18.3
1957	6686.3	7619.5	19.0
1958	7169.7	8153.3	19.1
1959	8019.3	9274.3	19.7
1960	9234.9	10776.5	20.5
1961	10836.2	12869.4	21.8
1962	12430.8	14701.4	23.2
1963	14506.6	17042.7	24.9
1964	16674.9	19709.9	26.0
1965	18820.5	22337.4	27.8
1966	21680.6	25514.5	29.0
1967	24914.0	29012.6	30.1
1968	28452.7	34233.6	31.6
1969	32705.2	39486.3	32.9
1970	37784.1	45913.2	35.2
1971	42571.6	51944.3	37.5
1972	49124.1	60245.4	39.7
1973	59366.1	74924.8	44.1
1974	71782.1	93833.2	53.3
1975	83591.1	108712.8	59.4
1976	94443.7	123540.9	65.2
1977	105397.8	135318.4	70.1
1978	115960.3	147244.2	73.5
1979	127600.9	157071.1	76.0
1980	138585.0	169931.5	81.6
1981	147103.4	181349.2	85.4
1982	157994.0	190611.5	87.7
1983	166631.6	199587.8	89.5
1984	175383.4	209451.9	91.8
1985	185335.1	220655.6	93.9
1986	193069.6	229938.8	94.8
1987	202072.8	235924.0	95.3
1988	212939.9	247159.7	95.8
1989	227122.2	263940.5	97.7
1990	243035.7	280133.0	100.0
1991	255531.8	297512.9	102.5
1992	265701.6	309256.6	104.5
1993	272075.3	317021.6	105.9
1994	279538.7	325655.7	106.7
1995	283245.4	331967.5	106.2
1996	291458.5	340619.1	106.0
1997	298475.2	345522.7	107.3

Estimate using TSP 5.0.

```
LINE  ****************************************************
|       1  freq a;
|       2  smpl 1955 1997;
|       3  read(file='cons99.txt') year cons yd price;
|       4  rcons=cons/(price/100);
|       5  ryd=yd/(price/100);
|       6  d1=0.0;
|       7  smpl 1974 1997;
|       8  d1=1.0;
|       9  smpl 1956 1997;
|      10  d1ryd=d1*ryd;
|      11  olsq rcons c ryd;
|      12  olsq rcons c d1 ryd d1ryd;
|      13  end;
*****************************************************
```

```
                Equation   1
                ============

  Method of estimation = Ordinary Least Squares


Dependent variable: RCONS
Current sample:  1956 to 1997
Number of observations:  42

 Mean of dependent variable = 149038.
Std. dev. of dependent var. = 78147.9
   Sum of squared residuals = .127951E+10
      Variance of residuals = .319878E+08
   Std. error of regression = 5655.77
                  R-squared = .994890
```

```
        Adjusted R-squared = .994762
   Durbin-Watson statistic = .116873
  F-statistic (zero slopes) = 7787.70
 Schwarz Bayes. Info. Crit. = 17.4101
 Log of likelihood function = -421.469
```

Variable	Estimated Coefficient	Standard Error	t-statistic
C	-3317.80	1934.49	-1.71508
RYD	.854577	.968382E-02	88.2480

```
                Equation   2
                ============

  Method of estimation = Ordinary Least Squares


Dependent variable: RCONS
Current sample:  1956 to 1997
Number of observations:  42

 Mean of dependent variable = 149038.
Std. dev. of dependent var. = 78147.9
   Sum of squared residuals = .244501E+09
      Variance of residuals = .643423E+07
   Std. error of regression = 2536.58
                  R-squared = .999024
```

Adjusted R-squared = .998946
Durbin-Watson statistic = .420979
F-statistic (zero slopes) = 12959.1
Schwarz Bayes. Info. Crit. = 15.9330
Log of likelihood function = -386.714

Variable	Estimated Coefficient	Standard Error	t-statistic
C	4204.11	1440.45	2.91861
D1	-39915.3	3154.24	-12.6545
RYD	.786609	.015024	52.3561
D1RYD	.194495	.018731	10.3839

1. `Equation 1`

Significance test:

`Equation 1` is:

$$\mathtt{RCONS} = \beta_1 + \beta_2 \mathtt{RYD}$$

$H_0: \ \beta_2 = 0$

(No.1) t Test $\Longrightarrow$ Compare 88.2480 and $t(42-2)$.

(No.2) F Test $\Longrightarrow$ Compare $\dfrac{R^2/G}{(1-R^2)/(n-k)} = \dfrac{.994890/1}{(1-.994890)/(42-2)} =$ 7787.8 and $F(1,40)$. Note that $\sqrt{7787.8} = 88.2485$.

1% point of $F(1, 40) = 7.31$

$H_0: \beta_2 = 0$ is rejected.

2. `Equation 2:`

$$\texttt{RCONS} = \beta_1 + \beta_2\texttt{D1} + \beta_3\texttt{RYD} + \beta_4\texttt{RYD} \times \texttt{D1}$$

$H_0: \beta_2 = \beta_3 = \beta_4 = 0$

F Test $\Longrightarrow$ Compare $\dfrac{R^2/G}{(1-R^2)/(n-k)} = \dfrac{.999024/3}{(1-.999024)/(42-4)} = 12965.5$ and $F(3, 38)$.

1% point of $F(3, 38) = 4.34$

$H_0: \beta_2 = \beta_3 = \beta_4 = 0$ is rejected.

3. `Equation 1` vs. `Equation 2`

   Test the structural change between 1973 and 1974.

   `Equation 2` is:

$$\mathtt{RCONS} = \beta_1 + \beta_2 \mathtt{D1} + \beta_3 \mathtt{RYD} + \beta_4 \mathtt{RYD} \times \mathtt{D1}$$

   $H_0: \beta_2 = \beta_4 = 0$

Restricted OLS $\Longrightarrow$ `Equation 1`

Unrestricted OLS $\Longrightarrow$ `Equation 2`

$$\frac{(\tilde{u}'\tilde{u} - e'e)/G}{e'e/(n-k)} = \frac{(.127951\mathtt{E}+10 - .244501\mathtt{E}+09)/2}{.244501\mathtt{E}+09/(42-4)} = 80.43$$

which should be compared with $F(2, 38)$.

1% point of $F(2, 38) = 5.211 < 80.43$

$H_0:\ \beta_2 = \beta_4 = 0$ is rejected.

$\Longrightarrow$ The structure was changed in 1974.

8 Generalized Least Squares Method (GLS, 一般化最小自乗法)

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

2. **Heteroscedasticity (不等分散，不均一分散)**

$$\sigma^2\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{pmatrix}$$

First-Order Autocorrelation (一階の自己相関，系列相関)

In the case of time series data, the subscript is conventionally given by t, not i .

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

$$\sigma^2\Omega = \frac{\sigma_\epsilon^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

$$V(u_t) = \sigma^2 = \frac{\sigma_\epsilon^2}{1-\rho^2}$$

3. The Generalized Least Squares (GLS，一般化最小二乘法) estimator of β, denoted by b, solves the following minimization problem:

$$\min_{b} \ (y - Xb)'\Omega^{-1}(y - Xb)$$

The GLSE of β is:

$$b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

4. In general, when Ω is symmetric, Ω is decomposed as follows.

$$\Omega = A'\Lambda A$$

   Λ is a diagonal matrix, where the diagonal elements of Λ are given by the eigen values.

   A is a matrix consisting of eigen vectors.

   When Ω is a positive definite matrix, all the diagonal elements of Λ are positive.

5. There exists P such that $\Omega = PP'$ (i.e., take $P = A'\Lambda^{1/2}$). $\Longrightarrow$ $P^{-1}\Omega P'^{-1} = I_n$

Multiply P^{-1} on both sides of $y = X\beta + u$.

We have:

$$y^\star = X^\star\beta + u^\star,$$

where $\quad y^\star = P^{-1}y, \quad X^\star = P^{-1}X, \quad$ and $\quad u^\star = P^{-1}u$.

The variance of $u^\star$ is:

$$\mathrm{V}(u^\star) = \mathrm{V}(P^{-1}u) = P^{-1}\mathrm{V}(u)P'^{-1} = \sigma^2 P^{-1}\Omega P'^{-1} = \sigma^2 I_n.$$

because $\Omega = PP'$, i.e., $P^{-1}\Omega P'^{-1} = I_n$.

Accordingly, the regression model is rewritten as:

$$y^\star = X^\star\beta + u^\star, \qquad u^\star \sim (0, \sigma^2 I_n)$$

Apply OLS to the above model.

Let b be as estimator of β from the above model.

That is, the minimization problem is given by:

$$\min_b \; (y^\star - X^\star b)'(y^\star - X^\star b),$$

which is equivalent to:

$$\min_{b} \ (y - Xb)'\Omega^{-1}(y - Xb).$$

Solving the minimization problem above, we have the following estimator:

$$\begin{aligned} b &= (X^{\star\prime}X^{\star})^{-1}X^{\star\prime}y^{\star} \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \end{aligned}$$

which is called GLS (Generalized Least Squares) estimator.

b is rewritten as follows:

$$b = \beta + (X^{\star\prime} X^{\star})^{-1} X^{\star\prime} u^{\star} = \beta + (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} u$$

The mean and variance of b are given by:

$$\mathrm{E}(b) = \beta,$$

$$\mathrm{V}(b) = \sigma^2 (X^{\star\prime} X^{\star})^{-1} = \sigma^2 (X' \Omega^{-1} X)^{-1}.$$

6. Suppose that the regression model is given by:

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

In this case, when we use OLS, what happens?

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u$$

$$\mathrm{V}(\hat{\beta}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Compare GLS and OLS.

(a) Expectation:

$$\mathrm{E}(\hat{\beta}) = \beta, \quad \text{and} \quad \mathrm{E}(b) = \beta$$

Thus, both $\hat{\beta}$ and b are unbiased estimator.

(b) Variance:

$$\mathrm{V}(\hat{\beta}) = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

$$\mathrm{V}(b) = \sigma^2 (X'\Omega^{-1}X)^{-1}$$

Which is more efficient, OLS or GLS?.

$$
\begin{aligned}
\mathrm{V}(\hat{\beta}) - \mathrm{V}(b) &= \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1} - \sigma^2 (X'\Omega^{-1} X)^{-1} \\
&= \sigma^2 \Big((X'X)^{-1} X' - (X'\Omega^{-1} X)^{-1} X'\Omega^{-1} \Big) \Omega \\
&\qquad \times \Big((X'X)^{-1} X' - (X'\Omega^{-1} X)^{-1} X'\Omega^{-1} \Big)' \\
&= \sigma^2 A \Omega A'
\end{aligned}
$$

Ω is the variance-covariance matrix of u, which is a positive definite matrix.

Therefore, except for $\Omega = I_n$, $A\Omega A'$ is also a positive definite matrix.

This implies that $\mathrm{V}(\hat{\beta}_i) - \mathrm{V}(b_i) > 0$ for the ith element of β.

Accordingly, b is more efficient than $\hat{\beta}$.

7. If $u \sim N(0, \sigma^2\Omega)$, then $b \sim N(\beta, \sigma^2(X'\Omega^{-1}X)^{-1})$.

   Consider testing the hypothesis $H_0 : R\beta = r$.

   $R: \ G \times k, \quad \mathrm{rank}(R) = G \leq k.$

   $Rb \sim N(R\beta, \sigma^2 R(X'\Omega^{-1}X)^{-1}R').$

   Therefore, the following quadratic form is distributed as:

$$\frac{(Rb - r)'(R(X'\Omega^{-1}X)^{-1}R')^{-1}(Rb - r)}{\sigma^2} \sim \chi^2(G)$$

8. Because $(y^\star - X^\star b)'(y^\star - X^\star b)/\sigma^2 \sim \chi^2(n-k)$, we obtain:

$$\frac{(y - Xb)'\Omega^{-1}(y - Xb)}{\sigma^2} \sim \chi^2(n-k)$$

9. Furthermore, from the fact that b is independent of $y - Xb$, the following F distribution can be derived:

$$\frac{(Rb - r)'(R(X'\Omega^{-1}X)^{-1}R')^{-1}(Rb - r)/G}{(y - Xb)'\Omega^{-1}(y - Xb)/(n-k)} \sim F(G, n-k)$$

10. Let b be the unrestricted GLSE and $\tilde{b}$ be the restricted GLSE.

Their residuals are given by e and $\tilde{u}$, respectively.

$$e = y - Xb, \qquad \tilde{u} = y - X\tilde{b}$$

Then, the F test statistic is written as follows:

$$\frac{(\tilde{u}'\Omega^{-1}\tilde{u} - e'\Omega^{-1}e)/G}{e'\Omega^{-1}e/(n-k)} \sim F(G, n-k)$$

8.1 Example: Mixed Estimation (Theil and Goldberger Model)

A generalization of the restricted OLS $\Longrightarrow$ Stochastic linear restriction:

$$r = R\beta + v, \qquad \mathrm{E}(v) = 0 \text{ and } \mathrm{V}(v) = \sigma^2\Psi$$

$$y = X\beta + u, \qquad \mathrm{E}(u) = 0 \text{ and } \mathrm{V}(u) = \sigma^2 I_n$$

Using a matrix form,

$$\begin{pmatrix} y \\ r \end{pmatrix} = \begin{pmatrix} X \\ R \end{pmatrix}\beta + \begin{pmatrix} u \\ v \end{pmatrix}, \qquad \mathrm{E}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \mathrm{V}\begin{pmatrix} u \\ v \end{pmatrix} = \sigma^2 \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}$$

For estimation, we do not need normality assumption.

Applying GLS, we obtain:

$$b = \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} X \\ R \end{pmatrix} \right)^{-1} \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} y \\ r \end{pmatrix} \right)$$

$$= \left(X'X + R'\Psi^{-1}R \right)^{-1} \left(X'y + R'\Psi^{-1}r \right).$$

Mean and Variance of *b*: *b* is rewritten as follows:

$$b = \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} X \\ R \end{pmatrix} \right)^{-1} \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} y \\ r \end{pmatrix} \right)$$

$$= \beta + \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} X \\ R \end{pmatrix} \right)^{-1} \begin{pmatrix} u \\ v \end{pmatrix}$$

Therefore, the mean and variance are given by:

$$\mathrm{E}(b) = \beta \qquad \Longrightarrow \qquad b \text{ is unbiased.}$$

$$\begin{aligned} \mathrm{V}(b) &= \sigma^2 \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} X \\ R \end{pmatrix} \right)^{-1} \\ &= \sigma^2 \left(X'X + R'\Psi^{-1}R \right)^{-1} \end{aligned}$$

9 Maximum Likelihood Estimation (MLE, 最尤法)

$\Longrightarrow$ Review of Last Semester

1. The distribution function of $\{X_i\}_{i=1}^n$ is $f(x;\theta)$, where $x = (x_1, x_2, \cdots, x_n)$ and $\theta = (\mu, \Sigma)$.

   Note that X is a vector of random variables and x is a vector of their realizations (i.e., observed data).

   Likelihood function $L(\cdot)$ is defined as $L(\theta; x) = f(x;\theta)$.

Note that $f(x;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$ when $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed.

The maximum likelihood estimator (MLE) of θ is θ such that:

$$\max_{\theta} L(\theta; X). \qquad \Longleftrightarrow \qquad \max_{\theta} \log L(\theta; X).$$

MLE satisfies the following two conditions:

(a) $\dfrac{\partial \log L(\theta; X)}{\partial \theta} = 0.$

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

2. **Fisher's information matrix (フィッシャーの情報行列)** is defined as:

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

where we have the following equality:

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

Proof of the above equality:

$$\int L(\theta; x)\mathrm{d}x = 1$$

Take a derivative with respect to θ.

$$\int \frac{\partial L(\theta; x)}{\partial \theta} \mathrm{d}x = 0$$

(We assume that (i) the domain of x does not depend on θ and (ii) the derivative $\dfrac{\partial L(\theta; x)}{\partial \theta}$ exists.)

Rewriting the above equation, we obtain:

$$\int \frac{\partial \log L(\theta; x)}{\partial \theta} L(\theta; x) \mathrm{d}x = 0,$$

i.e.,

$$\mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right) = 0.$$

Again, differentiating the above with respect to θ, we obtain:

$$\int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x)\mathrm{d}x + \int \frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial L(\theta; x)}{\partial'\theta} \mathrm{d}x$$
$$= \int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x)\mathrm{d}x + \int \frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial \log L(\theta; x)}{\partial\theta'} L(\theta; x)\mathrm{d}x$$
$$= \mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\Big) + \mathrm{E}\Big(\frac{\partial \log L(\theta; X)}{\partial\theta} \frac{\partial \log L(\theta; X)}{\partial\theta'}\Big) = 0.$$

Therefore, we can derive the following equality:

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

where the second equality utilizes $\mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right) = 0$.

3. **Cramer-Rao Lower Bound (クラメール・ラオの下限):** $(I(\theta))^{-1}$

Suppose that an estimator of θ is given by $s(X)$.

The expectation of $s(X)$ is:

$$\mathrm{E}(s(X)) = \int s(x)L(\theta; x)\mathrm{d}x.$$

Differentiating the above with respect to θ,

$$\begin{aligned}\frac{\partial \mathrm{E}(s(X))}{\partial\theta} &= \int s(x)\frac{\partial L(\theta; x)}{\partial\theta}\mathrm{d}x = \int s(x)\frac{\partial \log L(\theta; x)}{\partial\theta}L(\theta; x)\mathrm{d}x \\ &= \mathrm{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial\theta}\right)\end{aligned}$$

For simplicity, let $s(X)$ and θ be scalars.

Then,

$$\left(\frac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2 = \left(\mathrm{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta}\right)\right)^2 = \rho^2 \mathrm{V}\left(s(X)\right) \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right)$$
$$\leq \mathrm{V}\left(s(X)\right) \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),$$

where ρ denotes the correlation coefficient between $s(X)$ and $\dfrac{\partial \log L(\theta; X)}{\partial \theta}$, i.e.,

$$\rho = \frac{\mathrm{Cov}\left(s(X), \dfrac{\partial \log L(\theta; X)}{\partial \theta}\right)}{\sqrt{\mathrm{V}\left(s(X)\right)}\sqrt{\mathrm{V}\left(\dfrac{\partial \log L(\theta; X)}{\partial \theta}\right)}}.$$

Note that $|\rho| \le 1$.

Therefore, we have the following inequality:

$$\left(\frac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2 \le \mathrm{V}(s(X))\,\mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),$$

i.e.,

$$\mathrm{V}(s(X)) \geq \frac{\left(\dfrac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2}{\mathrm{V}\left(\dfrac{\partial \log L(\theta; X)}{\partial \theta}\right)}$$

Especially, when $\mathrm{E}(s(X)) = \theta$,

$$\mathrm{V}(s(X)) \geq \frac{1}{-\mathrm{E}\left(\dfrac{\partial^2 \log L(\theta; X)}{\partial \theta^2}\right)} = (I(\theta))^{-1}.$$

Even in the case where $s(X)$ is a vector, the following inequality holds.

$$\mathrm{V}(s(X)) \geq (I(\theta))^{-1},$$

where $I(\theta)$ is defined as:

$$\begin{aligned} I(\theta) &= -\mathrm{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\right) \\ &= \mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\frac{\partial \log L(\theta; X)}{\partial\theta'}\right) = \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right). \end{aligned}$$

The variance of any unbiased estimator of θ is larger than or equal to $(I(\theta))^{-1}$.

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:

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

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:

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

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:

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

$L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)$ 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.

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

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:

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

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

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

(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.

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

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

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

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

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

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

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:

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

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:

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

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:

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

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:

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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.

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

$$\Longrightarrow \qquad \text{Apply OLS.}$$

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

$$X^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}x_1 \\ x_2 - \rho x_1 \\ \vdots \\ x_n - \rho x_{n-1} \end{pmatrix} = P^{-1}X \qquad \Longrightarrow$$

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:

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

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

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

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

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 $X_1, X_2, \cdots, X_n, \cdots$ have distribution functions $F_1, F_2, \cdots$, respectively.

   If

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

   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,

$$\lim_{i\to\infty} P(|Z_i - \theta| < \epsilon) = 1,$$

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 $\hat{\theta}_i$ is a consistent estimator of θ.

3. **Chebyshev's inequality:**

For $g(X) \geq 0$,

$$P(g(X) \geq k) \leq \frac{\mathrm{E}(g(X))}{k},$$

where k is a positive constant.

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

and Var(X) = Σ.

Then, we have the following inequality:

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

Note as follows:

$$\begin{aligned}\mathrm{E}((X-\mu)'(X-\mu)) &= \mathrm{E}\Big(\mathrm{tr}((X-\mu)'(X-\mu))\Big) = \mathrm{E}\Big(\mathrm{tr}((X-\mu)(X-\mu)')\Big) \\ &= \mathrm{tr}\Big(\mathrm{E}((X-\mu)(X-\mu)')\Big) = \mathrm{tr}(\Sigma).\end{aligned}$$

5. **Example 1 (Univariate Case):**

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

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

   Use Chebyshev's inequality.

If $n \longrightarrow \infty$,

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

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)**

Univariate Case: $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed as $X_i \sim (\mu,\ \sigma^2)$.

Then,

$$\frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1),$$

which implies

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0,\sigma^2).$$

Multivariate Case: $X_1, X_2, \cdots, X_n$ are mutually independently and identically 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 and identically 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

$$\Sigma = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\Sigma_i\right).$$

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 $\theta_1, \theta_2, \cdots, \theta_k$.

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 $\alpha_1, \alpha_2, \cdots, \alpha_k$ are given by:

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

11 Consistency and Asymptotic Normality of OLSE

Regression model:

$$y = X\beta + u, \qquad u \sim (0, \sigma^2 I_n)$$

Consistency:

1. Let $\hat{\beta}_n = (X'X)^{-1}X'y$ be the OLS with sample size n.

   Consistency: As n is large, $\hat{\beta}_n$ converges to β.

2. Assume the stationarity assumption for X, i.e.,

$$\frac{1}{n}X'X \longrightarrow M_{xx}.$$

Then, we have the following result:

$$\frac{1}{n}X'u \longrightarrow 0.$$

Proof:

According to Chebyshev's inequality, for $g(Z) \geq 0$,

$$P(g(Z) \geq k) \leq \frac{\mathrm{E}(g(Z))}{k},$$

where k is a positive constant.

Set $g(Z) = Z'Z$, and $Z = \frac{1}{n}X'u$.

Apply Chebyshev's inequality.

$$\begin{aligned} \mathrm{E}\Big((\frac{1}{n}X'u)'\frac{1}{n}X'u\Big) &= \frac{1}{n^2}\mathrm{E}\Big(u'XX'u\Big) = \frac{1}{n^2}\mathrm{E}\Big(\mathrm{tr}(u'XX'u)\Big) = \frac{1}{n^2}\mathrm{E}\Big(\mathrm{tr}(XX'uu')\Big) \\ &= \frac{1}{n^2}\mathrm{tr}\Big(XX'\mathrm{E}(uu')\Big) = \frac{\sigma^2}{n^2}\mathrm{tr}(XX') = \frac{\sigma^2}{n^2}\mathrm{tr}(X'X) = \frac{\sigma^2}{n}\mathrm{tr}(\frac{1}{n}X'X). \end{aligned}$$

Therefore,

$$P\Big((\frac{1}{n}X'u)'\frac{1}{n}X'u \geq k\Big) \leq \frac{\sigma^2}{nk}\mathrm{tr}(\frac{1}{n}X'X) \longrightarrow 0 \times \mathrm{tr}(M_{xx}) = 0$$

Note that from the assumption,

$$\frac{1}{n}X'X \longrightarrow M_{xx}.$$

Therefore, we have:

$$(\frac{1}{n}X'u)'\frac{1}{n}X'u \longrightarrow 0,$$

which implies:

$$\frac{1}{n}X'u \longrightarrow 0,$$

because $(\frac{1}{n}X'u)'\frac{1}{n}X'u$ indicates a quadratic form.

3. Note that

$$\frac{1}{n}X'X \longrightarrow M_{xx}$$

results in

$$(\frac{1}{n}X'X)^{-1} \longrightarrow M_{xx}^{-1}$$

$\Longrightarrow$ Slutsky's Theorem

(*) **Slutsky's Theorem** $\quad g(\hat{\theta}) \longrightarrow g(\theta)$, when $\hat{\theta} \longrightarrow \theta$.

4. OLS is given by:

$$\hat{\beta}_n = \beta + (X'X)^{-1}X'u$$

$$= \beta + (\frac{1}{n}X'X)^{-1}(\frac{1}{n}X'u).$$

Therefore,

$$\hat{\beta}_n \longrightarrow \beta + M_{xx}^{-1} \times 0 = \beta$$

Thus, OLSE is a consitent estimator.

Asymptotic Normality:

1. Asymptotic Normality of OLSE

$$\sqrt{n}(\hat{\beta}_n - \beta) \longrightarrow N(0.\sigma^2 M_{xx}^{-1}), \quad \text{when } n \longrightarrow \infty.$$

2. **Central Limit Theorem:** Greenberg and Webster (1983)

   $Z_1, Z_2, \cdots, Z_n$ are mutually indelendently distributed with mean μ and variance Σ_i.

Then, we have the following result:

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

where

$$\Sigma = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\Sigma_i\right).$$

The distribution of Z_i is not assumed.

3. Define $Z_i = x_i' u_i$. Then, $\Sigma_i = \mathrm{Var}(Z_i) = \sigma^2 x_i' x_i$.

4. Σ is defined as:

$$\Sigma = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma^2 x_i' x_i\right) = \sigma^2 \lim_{n\to\infty}\left(\frac{1}{n}X'X\right) = \sigma^2 M_{xx},$$

where

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

5. Applying Central Limit Theorem (Greenberg and Webster (1983), we obtain the following:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i' u_i = \frac{1}{\sqrt{n}} X'u \longrightarrow N(0, \sigma^2 M_{xx}).$$

On the other hand, from $\hat{\beta}_n = \beta + (X'X)^{-1}X'u$, we can rewrite as:

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}X'X\right)^{-1} \frac{1}{\sqrt{n}} X'u.$$

$$\mathrm{Var}\left(\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{\sqrt{n}}X'u\right) = \mathrm{E}\left(\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{\sqrt{n}}X'u\left(\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{\sqrt{n}}X'u\right)'\right)$$

$$= \Big(\frac{1}{n}X'X\Big)^{-1}\Big(\frac{1}{n}X'\mathrm{E}(uu')X\Big)\Big(\frac{1}{n}X'X\Big)^{-1}$$
$$= \sigma^2\Big(\frac{1}{n}X'X\Big)^{-1} \longrightarrow \sigma^2 M_{xx}^{-1}.$$

Therefore,

$$\sqrt{n}(\hat{\beta} - \beta) \longrightarrow N(0, \sigma^2 M_{xx}^{-1})$$

$\Longrightarrow$ Asymptotic normality (漸近的正規性) of OLSE

The distribution of u_i is not assumed.

12 Instrumental Variable (操作変数法)

12.1 Measurement Error (測定誤差)

Errors in Variables

1. True regression model:

$$y = \tilde{X}\beta + u$$

2. Observed variable:

$$X = \tilde{X} + V$$

V: is called the **measurement error (測定誤差 or 観測誤差)**.

3. For the elements which do not include measurement errors in X, the corresponding elements in V are zeros.

4. Regression using observed variable:

$$y = X\beta + (u - V\beta)$$

OLS of β is:

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'(u - V\beta)$$

5. Assumptions:

(a) The measurement error in X is uncorrelated with $\tilde{X}$ in the limit. i.e.,

$$\text{plim}\Big(\frac{1}{n}\tilde{X}'V\Big) = 0.$$

Therefore, we obtain the following:

$$\text{plim}\Big(\frac{1}{n}X'X\Big) = \text{plim}\Big(\frac{1}{n}\tilde{X}'\tilde{X}\Big) + \text{plim}\Big(\frac{1}{n}V'V\Big) = \Sigma + \Omega$$

(b) u is not correlated with V.

u is not correlated with $\tilde{X}$.

That is,

$$\operatorname{plim}\left(\frac{1}{n}V'u\right) = 0, \qquad \operatorname{plim}\left(\frac{1}{n}\tilde{X}'u\right) = 0.$$

6. OLSE of β is:

$$\hat{\beta} = \beta + (X'X)^{-1}X'(u - V\beta) = \beta + (X'X)^{-1}(\tilde{X} + V)'(u - V\beta).$$

Therefore, we obtain the following:

$$\operatorname{plim}\hat{\beta} = \beta - (\Sigma + \Omega)^{-1}\Omega\beta$$

7. **Example: The Case of Two Variables:**

The regression model is given by:

$$y_t = \alpha + \beta \tilde{x}_t + u_t, \qquad x_t = \tilde{x}_t + v_t.$$

Under the above model,

$$\Sigma = \text{plim}\Big(\frac{1}{n}\tilde{X}'\tilde{X}\Big) = \text{plim}\begin{pmatrix} 1 & \frac{1}{n}\sum \tilde{x}_i \\ \frac{1}{n}\sum \tilde{x}_i & \frac{1}{n}\sum \tilde{x}_i^2 \end{pmatrix} = \begin{pmatrix} 1 & \mu \\ \mu & \mu^2 + \sigma^2 \end{pmatrix},$$

where μ and σ^2 represent the mean and variance of $\tilde{x}_i$.

$$\Omega = \text{plim}\Big(\frac{1}{n}V'V\Big) = \text{plim}\begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{n}\sum v_i^2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}.$$

Therefore,

$$\operatorname{plim}\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} = \begin{pmatrix}\alpha\\ \beta\end{pmatrix} - \left(\begin{pmatrix}1 & \mu\\ \mu & \mu^2+\sigma^2\end{pmatrix} + \begin{pmatrix}0 & 0\\ 0 & \sigma_v^2\end{pmatrix}\right)^{-1}\begin{pmatrix}0 & 0\\ 0 & \sigma_v^2\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$$

$$= \begin{pmatrix}\alpha\\ \beta\end{pmatrix} - \frac{1}{\sigma^2+\sigma_v^2}\begin{pmatrix}-\mu\sigma_v^2\beta\\ \sigma_v^2\beta\end{pmatrix}$$

Now we focus on β.

$\hat{\beta}$ is not consistent. because of:

$$\operatorname{plim}(\hat{\beta}) = \beta - \frac{\sigma_v^2\beta}{\sigma^2+\sigma_v^2} = \frac{\beta}{1+\sigma_v^2/\sigma^2} < \beta$$

12.2 Instrumental Variable (IV) Method (操作変数法 or IV 法)

Instrumental Variable (IV)

1. Consider the regression model: $y = X\beta + u$ and $u \sim N(0, \sigma^2 I_n)$.

   In the case of $\mathrm{E}(X'u) \neq 0$, OLSE of β is inconsistent.

2. **Proof:**

$$\hat{\beta} = \beta + (\frac{1}{n}X'X)^{-1}\frac{1}{n}X'u \longrightarrow \beta + M_{xx}^{-1}M_{xu},$$

where

$$\frac{1}{n}X'X \longrightarrow M_{xx}, \qquad \frac{1}{n}X'u \longrightarrow M_{xu} \neq 0$$

3. Find the Z which satisfies $\frac{1}{n}Z'u \longrightarrow M_{zu} = 0$.

   Multiplying Z' on both sides of the regression model: $y = X\beta + u$,

$$Z'y = Z'X\beta + Z'u$$

   Dividing n on both sides of the above equation, we take plim on both sides.

Then, we obtain the following:

$$\operatorname{plim}\left(\frac{1}{n}Z'y\right) = \operatorname{plim}\left(\frac{1}{n}Z'X\right)\beta + \operatorname{plim}\left(\frac{1}{n}Z'u\right) = \operatorname{plim}\left(\frac{1}{n}Z'X\right)\beta.$$

Accordingly, we obtain:

$$\beta = \left(\operatorname{plim}\left(\frac{1}{n}Z'X\right)\right)^{-1} \operatorname{plim}\left(\frac{1}{n}Z'y\right).$$

Therefore, we consider the following estimator:

$$\beta_{IV} = (Z'X)^{-1}Z'y,$$

which is taken as an estimator of β.

$\Longrightarrow$ **Instrumental Variable Method (操作変数法 or IV 法)**

4. Assume the followings:

$$\frac{1}{n}Z'X \longrightarrow M_{zx}, \qquad \frac{1}{n}Z'Z \longrightarrow M_{zz}, \qquad \frac{1}{n}Z'u \longrightarrow 0$$

5. **Distribution of $\boldsymbol{\beta_{IV}}$:**

$$\beta_{IV} = (Z'X)^{-1}Z'y = (Z'X)^{-1}Z'(X\beta + u) = \beta + (Z'X)^{-1}Z'u,$$

which is rewritten as:

$$\sqrt{n}(\beta_{IV} - \beta) = \Big(\frac{1}{n}Z'X\Big)^{-1}\Big(\frac{1}{\sqrt{n}}Z'u\Big)$$

Applying the Central Limit Theorem to $\left(\frac{1}{\sqrt{n}}Z'u\right)$, we have the following result:

$$\frac{1}{\sqrt{n}}Z'u \longrightarrow N(0, \sigma^2 M_{zz}).$$

Therefore,

$$\sqrt{n}(\beta_{IV} - \beta) = \left(\frac{1}{n}Z'X\right)^{-1}\left(\frac{1}{\sqrt{n}}Z'u\right) \longrightarrow N(0, \sigma^2 M_{zx}^{-1} M_{zz} M_{zx}'^{-1})$$

$\Longrightarrow$ Consistency and Asymptotic Normality

6. The variance of β_{IV} is given by:

$$V(\beta_{IV}) = s^2 (Z'X)^{-1} Z'Z (X'Z)^{-1},$$

where

$$s^2 = \frac{(y - X\beta_{IV})'(y - X\beta_{IV})}{n - k}.$$

12.3 Two-Stage Least Squares Method (2 段階最小二乗法, 2SLS or TSLS)

1. Regression Model:

$$y = X\beta + u, \qquad u \sim N(0, \sigma^2 I),$$

   In the case of $\mathrm{E}(X'u) \neq 0$, OLSE is not consistent.

2. Find the variable Z which satisfies $\dfrac{1}{n}Z'u \longrightarrow M_{zu} = 0$.

3. Use $Z = \hat{X}$ for the instrumental variable.

$\hat{X}$ is the predicted value which regresses X on the other exogenous variables, say W.

That is, consider the following regression model:

$$X = WB + V.$$

Estimate B by OLS.

Then, we obtain the prediction:

$$\hat{X} = W\hat{B},$$

where $\hat{B} = (W'W)^{-1}W'X$.

Or, equivalently,

$$\hat{X} = W(W'W)^{-1}W'X.$$

$\hat{X}$ is used for the instrumental variable of X.

4. The IV method is rewritten as:

$$\beta_{IV} = (\hat{X}'X)^{-1}\hat{X}'y = (X'W(W'W)^{-1}W'X)^{-1}X'W(W'W)^{-1}W'y.$$

Furthermore, β_{IV} is written as follows:

$$\beta_{IV} = \beta + (X'W(W'W)^{-1}W'X)^{-1}X'W(W'W)^{-1}W'u.$$

Therefore, we obtain the following expression:

$$\sqrt{n}(\beta_{IV} - \beta) = \left(\Big(\frac{1}{n}X'W\Big)\Big(\frac{1}{n}W'W\Big)^{-1}\Big(\frac{1}{n}XW'\Big)'\right)^{-1}\Big(\frac{1}{n}X'W\Big)\Big(\frac{1}{n}W'W\Big)^{-1}\Big(\frac{1}{\sqrt{n}}W'u\Big)$$
$$\longrightarrow N\Big(0, \sigma^2(M_{xw}M_{ww}^{-1}M'_{xw})^{-1}\Big).$$

5. Clearly, there is no correlation between W and u at least in the limit, i.e.,

$$\text{plim}\Big(\frac{1}{n}W'u\Big) = 0.$$

6. **Remark:**

$$\hat{X}'X = X'W(W'W)^{-1}W'X = X'W(W'W)^{-1}W'W(W'W)^{-1}W'X = \hat{X}'\hat{X}.$$

Therefore,

$$\beta_{IV} = (\hat{X}'X)^{-1}\hat{X}'y = (\hat{X}'\hat{X})^{-1}\hat{X}'y,$$

which implies the OLS estimator of β in the regression model: $y = \hat{X}\beta + u$ and $u \sim N(0, \sigma^2 I_n)$.

Example:

$$y_t = \alpha x_t + \beta z_t + u_t, \qquad u_t \sim (0, \sigma^2).$$

Suppose that x_t is correlated with u_t but z_t is not correlated with u_t.

- 1st Step:

Estimate the following regression model:

$$x_t = \gamma w_t + \delta z_t + \cdots + v_t,$$

by OLS. $\Longrightarrow$ Obtain $\hat{x}_t$ through OLS.

- 2nd Step:

Estimate the following regression model:

$$y_t = \alpha \hat{x}_t + \beta z_t + u_t,$$

by OLS. $\Longrightarrow$ α_{iv} and β_{iv}

Note as follows. Estimate the following regression model:

$$z_t = \gamma_2 w_t + \delta_2 z_t + \cdots + v_{2t},$$

by OLS.

$\Longrightarrow$ $\hat{\gamma}_2 = 0$, $\hat{\delta}_2 = 1$, and the other coefficient estimates are zeros. i.e., $\hat{z}_t = z_t$.

Eviews Command:

```
tsls y x z @ w z ...
```

13 Large Sample Tests

13.1 Wald, LM and LR Tests

$\theta : K \times 1$

$h(\theta) : G \times 1$ vector function, $G \leq K$

$\theta : K \times 1$

The null hypothesis $H_0 : h(\theta) = 0 \implies G$ restrictions

$\tilde{\theta} : k \times 1$, restricted maximum likelihood estimate

$\hat{\theta} : k \times 1$, unrestricted maximum likelihood estimate

$I(\theta) : k \times k$, information matrix, i.e.,

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

$\log L(\theta)$: log-likelihood function

$R_\theta = \dfrac{\partial h(\theta)}{\partial\theta'} : G \times k$

$F_\theta = \dfrac{\partial \log L(\theta)}{\partial\theta} : k \times 1$

1. **Wald Test (ワルド検定):** $W = h(\hat{\theta})'\Big(R_{\hat{\theta}}(I(\hat{\theta}))^{-1}R'_{\hat{\theta}}\Big)^{-1}h(\hat{\theta})$

   (a) $h(\theta) \approx h(\hat{\theta}) + \dfrac{\partial h(\hat{\theta})}{\partial\theta'}(\theta - \hat{\theta}) \quad \Longleftarrow \quad h(\theta)$ is linearized around $\theta = \hat{\theta}$.

Under the null hypothesis $h(\theta) = 0$,

$$h(\hat{\theta}) \approx \frac{\partial h(\hat{\theta})}{\partial \theta'}(\hat{\theta} - \theta) = R_{\hat{\theta}}(\hat{\theta} - \theta)$$

(b) $\hat{\theta}$ is MLE.

From the properties of MLE,

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

That is, approximately, we have the following result:

$$(\hat{\theta} - \theta) \sim N\Big(0, (I(\theta))^{-1}\Big).$$

(c) The distribution of $h(\hat{\theta})$ is approximately given by:

$$h(\hat{\theta}) \sim N\Big(0, R_{\hat{\theta}}(I(\theta))^{-1}R'_{\hat{\theta}}\Big)$$

(d) Therefore, the $\chi^2(G)$ distribution is derived as follows:

$$h(\hat{\theta})\Big(R_{\hat{\theta}}(I(\theta))^{-1}R'_{\hat{\theta}}\Big)^{-1}h(\hat{\theta})' \longrightarrow \chi^2(G).$$

Furthermore, from the fact that $I(\hat{\theta}) \longrightarrow I(\theta)$ as $n \longrightarrow \infty$ (i.e., convergence in probability, 確率収束), we can replace θ by $\hat{\theta}$ as follows:

$$h(\hat{\theta})\Big(R_{\hat{\theta}}(I(\hat{\theta}))^{-1}R'_{\hat{\theta}}\Big)^{-1}h(\hat{\theta})' \longrightarrow \chi^2(G).$$

2. **Lagrange Multiplier Test (ラグランジェ乗数検定):** $LM = F'_{\tilde{\theta}}(I(\tilde{\theta}))^{-1}F_{\tilde{\theta}}$

(a) MLE with the constraint $h(\theta) = 0$:

$$\max_{\theta} \log L(\theta), \quad \text{subject to} \quad h(\theta) = 0$$

The Lagrangian function:

$$L = \log L(\theta) + \lambda h(\theta)$$

(b) For maximization, we have the following two equations:

$$\frac{\partial L}{\partial \theta} = \frac{\partial \log L(\theta)}{\partial \theta} + \lambda \frac{\partial h(\theta)}{\partial \theta} = 0$$
$$\frac{\partial L}{\partial \lambda} = h(\theta) = 0$$

(c) Mean and variance of $\dfrac{\partial \log L(\theta)}{\partial \theta}$ are given by:

$$\mathrm{E}\Big(\frac{\partial \log L(\theta)}{\partial \theta}\Big) = 0, \qquad \mathrm{V}\Big(\frac{\partial \log L(\theta)}{\partial \theta}\Big) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}\Big) = I(\theta).$$

(d) Therefore, using the central limit theorem,

$$\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta)}{\partial \theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial \theta} \longrightarrow N\Big(0, \lim_{n\to\infty}\Big(\frac{1}{n}I(\theta)\Big)\Big)$$

(e) Therefore,

$$\frac{\partial \log L(\theta)}{\partial \theta}(I(\theta))^{-1}\frac{\partial \log L(\theta)}{\partial \theta'} \longrightarrow \chi^2(G)$$

Because MLE is consistent, i.e., $\tilde{\theta} \longrightarrow \theta$, we have the result:

$$F'_{\tilde{\theta}}(I(\tilde{\theta}))^{-1}F_{\tilde{\theta}} \longrightarrow \chi^2(G).$$

3. **Likelihood Ratio Test (尤度比検定):** $LR = -2\log\lambda \longrightarrow \chi^2(G)$

$$\lambda = \frac{L(\tilde{\theta})}{L(\hat{\theta})}$$

(a) By Taylor series expansion evaluated at $\theta = \hat{\theta}$, $\log L(\theta)$ is given by:

$$\begin{aligned}
\log L(\theta) &= \log L(\hat{\theta}) + \frac{\partial \log L(\hat{\theta})}{\partial \theta}(\theta - \hat{\theta}) \\
&\qquad + \frac{1}{2}(\theta - \hat{\theta})' \frac{\partial^2 \log L(\hat{\theta})}{\partial \theta \partial \theta'}(\theta - \hat{\theta}) + \cdots \\
&= \log L(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})' \frac{\partial^2 \log L(\hat{\theta})}{\partial \theta \partial \theta'}(\theta - \hat{\theta}) + \cdots
\end{aligned}$$

Note that $\dfrac{\partial \log L(\hat{\theta})}{\partial \theta} = 0$ because $\hat{\theta}$ is MLE.

$$\begin{aligned} -2(\log L(\theta) - \log L(\hat{\theta})) &\approx -(\theta - \hat{\theta})' \Big(\frac{\partial^2 \log L(\hat{\theta})}{\partial \theta \partial \theta'} \Big) (\theta - \hat{\theta}) \\ &= \sqrt{n}(\hat{\theta} - \theta)' \Big(-\frac{1}{n} \frac{\partial^2 \log L(\hat{\theta})}{\partial \theta \partial \theta'} \Big) \sqrt{n}(\hat{\theta} - \theta) \\ &\longrightarrow \chi^2(G) \end{aligned}$$

Note:

(1) $\hat{\theta} \longrightarrow \theta,$

(2) $-\dfrac{1}{n}\dfrac{\partial^2 \log L(\hat{\theta})}{\partial\theta\partial\theta'} \longrightarrow -\lim_{n\to\infty}\Big(\dfrac{1}{n}\mathrm{E}\Big(\dfrac{\partial^2 \log L(\hat{\theta})}{\partial\theta\partial\theta'}\Big)\Big) = \lim_{n\to\infty}\Big(\dfrac{1}{n}I(\theta)\Big),$

(3) $\sqrt{n}(\hat{\theta}-\theta) \longrightarrow N\Big(0, \lim_{n\to\infty}\Big(\dfrac{1}{n}I(\theta)\Big)\Big).$

(b) Under $H_0: h(\theta) = 0$,

$$-2(\log L(\tilde{\theta}) - \log L(\hat{\theta})) \longrightarrow \chi^2(G).$$

Remember that $h(\tilde{\theta}) = 0$ is always satisfied.

For proof, see Theil (1971, p.396).

4. All of W, LM and LR are asymptotically distributed as $\chi^2(G)$ random variables under the null hypothesis $H_0 : h(\theta) = 0$.

5. Under some comditions, we have $W \geq LR \geq LM$. See Engle (1981) “Wald, Likelihood and Lagrange Multiplier Tests in Econometrics,” Chap. 13 in *Handbook of Econometrics*, Vol.2, Grilliches and Intriligator eds, North-Holland.

13.2 Example: W, LM and LR Tests

Date file $\Longrightarrow$ cons99.txt (same data as before)

Each column denotes year, nominal household expenditures (家計消費，10 billion yen), household disposable income (家計可処分所得，10 billion yen) and household expenditure deflator (家計消費デフレータ，1990=100) from the left.

1955	5430.1	6135.0	18.1
1956	5974.2	6828.4	18.3
1957	6686.3	7619.5	19.0
1958	7169.7	8153.3	19.1
1959	8019.3	9274.3	19.7
1960	9234.9	10776.5	20.5
1961	10836.2	12869.4	21.8
1962	12430.8	14701.4	23.2
1963	14506.6	17042.7	24.9
1964	16674.9	19709.9	26.0
1965	18820.5	22337.4	27.8
1966	21680.6	25514.5	29.0
1967	24914.0	29012.6	30.1
1968	28452.7	34233.6	31.6
1969	32705.2	39486.3	32.9
1970	37784.1	45913.2	35.2
1971	42571.6	51944.3	37.5
1972	49124.1	60245.4	39.7
1973	59366.1	74924.8	44.1
1974	71782.1	93833.2	53.3
1975	83591.1	108712.8	59.4
1976	94443.7	123540.9	65.2
1977	105397.8	135318.4	70.1
1978	115960.3	147244.2	73.5
1979	127600.9	157071.1	76.0
1980	138585.0	169931.5	81.6
1981	147103.4	181349.2	85.4
1982	157994.0	190611.5	87.7
1983	166631.6	199587.8	89.5
1984	175383.4	209451.9	91.8
1985	185335.1	220655.6	93.9
1986	193069.6	229938.8	94.8
1987	202072.8	235924.0	95.3
1988	212939.9	247159.7	95.8
1989	227122.2	263940.5	97.7
1990	243035.7	280133.0	100.0
1991	255531.8	297512.9	102.5
1992	265701.6	309256.6	104.5
1993	272075.3	317021.6	105.9
1994	279538.7	325655.7	106.7
1995	283245.4	331967.5	106.2
1996	291458.5	340619.1	106.0
1997	298475.2	345522.7	107.3

```
          PROGRAM
LINE  ******************************************************
|           freq a;
|           smpl 1955 1997;
|           read(file='cons99.txt') year cons yd price;
|           rcons=cons/(price/100);
|           ryd=yd/(price/100);
|           lyd=log(ryd);
|           olsq rcons c ryd;
|           olsq @res @res(-1);
|           ar1  rcons c ryd;
|           olsq rcons c lyd;
|           param a1 0 a2 0 a3 1;
|           frml eq rcons=a1+a2*((ryd**a3)-1.)/a3;
|           lsq(tol=0.00001,maxit=100) eq;
|           a3=1.15;
|           rryd=((ryd**a3)-1.)/a3;
|           ar1  rcons c rryd;
|           end;
******************************************************
```

Equation 1
============

Method of estimation = Ordinary Least Squares

Dependent variable: RCONS
Current sample: 1955 to 1997
Number of observations: 43

```
       Mean of dep. var. = 146270.          LM het. test = .207443 [.649]
  Std. dev. of dep. var. = 79317.2         Durbin-Watson = .115101 [.000,.000]
Sum of squared residuals = .129697E+10  Jarque-Bera test = 9.47539 [.009]
   Variance of residuals = .316335E+08   Ramsey's RESET2 = 53.6424 [.000]
Std. error of regression = 5624.36       F (zero slopes) = 8311.90 [.000]
               R-squared = .995092       Schwarz B.I.C. = 435.051
      Adjusted R-squared = .994972       Log likelihood = -431.289
```

Variable	Estimated Coefficient	Standard Error	t-statistic	P-value
C	-2919.54	1847.55	-1.58022	[.122]
RYD	.852879	.935486E-02	91.1696	[.000]

```
                              Equation   2
                              ============
                 Method of estimation = Ordinary Least Squares

Dependent variable: @RES
Current sample:  1956 to 1997
Number of observations:  42

       Mean of dep. var. = -95.5174
  Std. dev. of dep. var. = 5588.52
Sum of squared residuals = .146231E+09
   Variance of residuals = .356662E+07
Std. error of regression = 1888.55
               R-squared = .885884
      Adjusted R-squared = .885884
            LM het. test = .760256 [.383]
           Durbin-Watson = 1.40409 [.023,.023]
             Durbin's h = 1.97732 [.048]
        Durbin's h alt. = 1.91077 [.056]
        Jarque-Bera test = 6.49360 [.039]
```

Ramsey's RESET2 = .186107 [.668]
Schwarz B.I.C. = 377.788
Log likelihood = -375.919

Variable	Estimated Coefficient	Standard Error	t-statistic	P-value
@RES(-1)	.950693	.053301	17.8362	[.000]

Equation 3
============

FIRST-ORDER SERIAL CORRELATION OF THE ERROR
Objective function: Exact ML (keep first obs.)

Dependent variable: RCONS
Current sample: 1955 to 1997
Number of observations: 43

Mean of dep. var.	= 146270.	R-squared	= .999480
Std. dev. of dep. var.	= 79317.2	Adjusted R-squared	= .999454
Sum of squared residuals	= .145826E+09	Durbin-Watson	= 1.38714
Variance of residuals	= .364564E+07	Schwarz B.I.C.	= 391.061
Std. error of regression	= 1909.36	Log likelihood	= -385.419

Parameter	Estimate	Standard Error	t-statistic	P-value
C	1672.42	6587.40	.253881	[.800]
RYD	.840011	.027182	30.9032	[.000]
RHO	.945025	.045843	20.6143	[.000]

```
                                Equation   4
                                ============
                      Method of estimation = Ordinary Least Squares

Dependent variable: RCONS
Current sample:  1955 to 1997
Number of observations:  43

       Mean of dep. var. = 146270.          LM het. test = 2.21031 [.137]
  Std. dev. of dep. var. = 79317.2          Durbin-Watson = .029725 [.000,.000]
Sum of squared residuals = .256040E+11  Jarque-Bera test = 3.72023 [.156]
   Variance of residuals = .624487E+09   Ramsey's RESET2 = 344.855 [.000]
Std. error of regression = 24989.7       F (zero slopes) = 382.117 [.000]
               R-squared = .903100        Schwarz B.I.C. = 499.179
      Adjusted R-squared = .900737        Log likelihood = -495.418

           Estimated    Standard
Variable  Coefficient     Error        t-statistic   P-value
C         -.115228E+07  66538.5        -17.3175      [.000]
LYD       109305.       5591.69        19.5478       [.000]
```

NONLINEAR LEAST SQUARES
========================

CONVERGENCE ACHIEVED AFTER 84 ITERATIONS

Number of observations = 43 Log likelihood = -414.362
Schwarz B.I.C. = 420.004

Parameter	Estimate	Standard Error	t-statistic	P-value
A1	16544.5	2615.60	6.32530	[.000]
A2	.063304	.024133	2.62307	[.009]
A3	1.21694	.031705	38.3839	[.000]

Standard Errors computed from quadratic form of analytic first derivatives (Gauss)

Equation: EQ
Dependent variable: RCONS

Mean of dep. var. = 146270.

```
  Std. dev. of dep. var. = 79317.2
Sum of squared residuals = .590213E+09
   Variance of residuals = .147553E+08
Std. error of regression = 3841.27
               R-squared = .997766
      Adjusted R-squared = .997655
            LM het. test = .174943 [.676]
           Durbin-Watson = .253234 [.000,.000]
```

Equation 5
============

FIRST-ORDER SERIAL CORRELATION OF THE ERROR
Objective function: Exact ML (keep first obs.)

Dependent variable: RCONS
Current sample: 1955 to 1997
Number of observations: 43

Mean of dep. var.	= 146270.	R-squared	= .999470
Std. dev. of dep. var.	= 79317.2	Adjusted R-squared	= .999443
Sum of squared residuals	= .140391E+09	Durbin-Watson	= 1.43657
Variance of residuals	= .350977E+07	Schwarz B.I.C.	= 389.449
Std. error of regression	= 1873.44	Log likelihood	= -383.807

Parameter	Estimate	Standard Error	t-statistic	P-value
C	12034.8	3346.47	3.59628	[.000]
RRYD	.140723	.282614E-02	49.7933	[.000]
RHO	.876924	.068199	12.8583	[.000]

1. `Equation 1` vs. `Equation 3` (Test of Serial Correlation)

`Equation 1` is:

$$\texttt{RCONS}_t = \beta_1 + \beta_2 \texttt{RYD}_t + u_t, \qquad \epsilon_t \sim \text{ iid } N(0, \sigma_\epsilon^2)$$

`Equation 3` is:

$$\texttt{RCONS}_t = \beta_1 + \beta_2 \texttt{RYD}_t + u_t, \qquad u_t = \rho u_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{ iid } N(0, \sigma_\epsilon^2)$$

The null hypothesis is $H_0 : \rho = 0$

Restricted MLE $\Longrightarrow$ `Equation 1`

Unrestricted MLE $\Longrightarrow$ `Equation 3`

The log-likelihood function of `Equation 3` is:

$$\log L(\beta, \sigma_\epsilon^2, \rho) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_\epsilon^2) + \frac{1}{2}\log(1-\rho^2) - \frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{n}(\mathtt{RCONS}_t^* - \beta_1 \mathtt{CONST}_t^* - \beta_2 \mathtt{RYD}_t^*)^2,$$

where

$$\mathtt{RCONS}_t^* = \begin{cases} \sqrt{1-\rho^2}\mathtt{RCONS}_t, & \text{for } t = 1, \\ \mathtt{RCONS}_t - \rho \mathtt{RCONS}_{t-1}, & \text{for } t = 2, 3, \cdots, n, \end{cases}$$

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

$$\texttt{RYD}_t^* = \begin{cases} \sqrt{1-\rho^2}\texttt{RYD}_t, & \text{for } t = 1, \\ \texttt{RYD}_t - \rho\texttt{RYD}_{t-1}, & \text{for } t = 2, 3, \cdots, n. \end{cases}$$

- MLE with the restriction $\rho = 0$ (`Equation 1`) solves:

$$\max_{\beta,\sigma_\epsilon^2} \log L(\beta, \sigma_\epsilon^2, 0)$$

Restricted MLE $\Longrightarrow \tilde{\beta}, \tilde{\sigma}_\epsilon^2$

`Log of likelihood function = -431.289`

- MLE without the restriction $\rho = 0$ (`Equation 3`) solves:

$$\max_{\beta, \sigma_\epsilon^2, \rho} \log L(\beta, \sigma_\epsilon^2, \rho)$$

  Unrestricted MLE $\Longrightarrow \hat{\beta}, \hat{\sigma}_\epsilon^2, \hat{\rho}$

  `Log of likelihood function = -385.419`

The likelihood ratio test statistic is:

$$\begin{aligned}-2\log(\lambda) &= -2\log\Big(\frac{L(\tilde{\beta}, \tilde{\sigma}_\epsilon^2, 0)}{L(\hat{\beta}, \hat{\sigma}_\epsilon^2, \hat{\rho})}\Big) = -2\Big(\log L(\tilde{\beta}, \tilde{\sigma}_\epsilon^2, 0) - \log L(\hat{\beta}, \hat{\sigma}_\epsilon^2, \hat{\rho})\Big) \\ &= -2\Big(-431.289 - (-385.419)\Big) = 91.74.\end{aligned}$$

The asymptotic distribution is given by:

$$-2\log(\lambda) \sim \chi^2(G),$$

where G is the number of the restrictions, i.e., $G = 1$ in this case.

The 1% upper probability point of $\chi^2(1)$ is 6.635.

$$91.74 > 6.635$$

Therefore, $H_0:\ \rho = 0$ is rejected.

There is serial correlation in the error term.

2. `Equation 1` (Test of Serial Correlation $\longrightarrow$ Lagrange Multiplier Test)

`Equation 2` is:

$$@\mathtt{RES}_t = \rho @\mathtt{RES}_{t-1} + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2),$$

where $@\mathtt{RES}_t = \mathtt{RCONS}_t - \hat{\beta}_1 - \hat{\beta}_2 \mathtt{RYD}_t$, and $\hat{\beta}_1$ and $\hat{\beta}_2$ are OLSEs.

The null hypothesis is $H_0 : \rho = 0$

```
 @RES(-1)   .950693        .053301        17.8362        [.000]
```

Therefore, the Lagrange multiplier test statistic is $17.8362^2 = 318.13 > 6.635$.

$H_0: \ \rho = 0$ is rejected.

3. `Equation 3` (Test of Serial Correlation $\longrightarrow$ Wald Test)

   `Equation 3` is:

$$\texttt{RCONS}_t = \beta_1 + \beta_2 \texttt{RYD}_t + u_t, \qquad u_t = \rho u_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{ iid } N(0, \sigma_\epsilon^2)$$

   The null hypothesis is $H_0: \ \rho = 0$

```
 RHO          .945025        .045843        20.6143        [.000]
```

   The Wald test statistics is $20.6143^2 = 424.95$, which is compared with $\chi^2(1)$.

4. `Equation 1` vs. `NONLINEAR LEAST SQUARES` (Choice of Functional Form – linear):

`NONLINEAR LEAST SQUARES` estimates:

$$\mathtt{RCONS}_t = a1 + a2\frac{\mathtt{RYD}_t^{a3} - 1}{a3} + u_t.$$

When $a3 = 1$, we have:

$$\mathtt{RCONS}_t = (a1 - a2) + a2\mathtt{RYD}_t + u_t,$$

which is equivalent to `Equation 1`.

The null hypothesis is $H_0:\ a3 = 1$, where $G = 1$.

- MLE with $a3 = 1$ MLE (`Equation 1`)

  `Log of likelihood function = -431.289`

- MLE without $a3 = 1$ (`NONLINEAR LEAST SQUARES`)

  `Log of likelihood function = -414.362`

The likelihood ratio test statistic is given by:

$$-2\log(\lambda) = -2\Big(-431.289 - (-414.362)\Big) = 33.854.$$

The 1% upper probability point of $\chi^2(1)$ is 6.635.

$$33.854 > 6.635$$

$H_0: a3 = 1$ is rejected.

Therefore, the functional form of the regression model is not linear.

5. `Equation 4` vs. `NONLINEAR LEAST SQUARES` (Choice of Functional Form – log-linear):

   In `NONLINEAR LEAST SQUARES`, i.e.,

$$\mathtt{RCONS}_t = a1 + a2\frac{\mathtt{RYD}_t^{a3} - 1}{a3} + u_t,$$

   if $a3 = 0$, we have:

$$\mathtt{RCONS}_t = a1 + a2\log(\mathtt{RYD}_t) + u_t,$$

which is equivalent to `Equation 3`.

The null hypothesis is $H_0: \; a3 = 0$, where $G = 1$.

- MLE with $a3 = 0$ (`Equation 3`)

  `Log of likelihood function = -495.418`

- MLE without $a3 = 0$ (`NONLINEAR LEAST SQUARES`)

  `Log of likelihood function = -414.362`

The likelihood ratio test statistic is:

$$-2\log(\lambda) = -2\Big(-495.418 - (-414.362)\Big) = 162.112 > 6.635.$$

Therefore, $H_0: \ a3 = 0$ is rejected.

As a result, the functional form of the regression model is not log-linear, either.

6. `Equation 1` vs. `Equation 5` (Simultaneous Test of Serial Correlation and Linear Function):

   `Equation 5` is:

$$\mathtt{RCONS}_t = a1 + a2\frac{\mathtt{RYD}_t^{a3} - 1}{a3} + u_t, \qquad u_t = \rho u_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{ iid } N(0, \sigma_\epsilon^2)$$

   The null hypothesis is $H_0: \ a3 = 1, \ \rho = 0$

Restricted MLE $\Longrightarrow$ `Equation 1`

Unrestricted MLE $\Longrightarrow$ `Equation 4`

Remark: In Lines 14–16 of `PROGRAM`, we have estimated `Equation 4`, given $a3 = 0.00, 0.01, 0.02, \cdots$.

As a result, $a3 = 1.15$ gives us the maximum log-likelihood.

The likelihood ratio test statistic is:

$$-2\log(\lambda) = -2\Big(-431.289 - (-383.807)\Big) = 94.964.$$

$-2\log(\lambda) \sim \chi^2(2)$ in this case.

The 1% upper probability point of $\chi^2(2)$ is 9.210.

$$94.964 > 9.210$$

$H_0: \; a3 = 1, \rho = 0$ is rejected.

`Equation 3` vs. `Equation 5` vs. (Taking into account serially correlated errors, Choice of Functional Form – linear):

The null hypothesis is $H_0 : \ a3 = 1, \ \rho = 0$

From `Equation 3`,

```
 Log likelihood = -385.419
```

From `Equation 5`,

```
 Log likelihood = -383.807
```

$$2(-383.807 - (-385.419)) = 3.224 < 6.635.$$

$H_0: \ a3 = 1$ is not rejected, given $\rho \neq 0$.

Thus, if serial correlation is taken into account, the regression model is linear.

14 Time Series Analysis (時系列分析)

14.1 Introduction

Textbooks

- J.D. Hamilton (1994) *Econometric Analysis*
  沖本・井上訳 (2006)『時系列解析 (上・下)』
- A.C. Harvey (1981) *Time Series Models*
  国友・山本訳 (1985)『時系列モデル入門』
- 沖本竜義 (2010)『経済・ファイナンスデータの計量時系列分析』

1. **Stationarity (定常性) :**

Let $y_1, y_2, \cdots, y_T$ be time series data.

(a) **Weak Stationarity (弱定常性) :**

$$\mathrm{E}(y_t) = \mu,$$

$$\mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)) = \gamma(\tau), \qquad \tau = 0, 1, 2, \cdots$$

The first and second moments depend on time difference, not time itself.

(b) **Strong Stationarity (強定常性) :**

Let $f(y_{t_1}, y_{t_2}, \cdots, y_{t_r})$ be the joint distribution of $y_{t_1}, y_{t_2}, \cdots, y_{t_r}$.

$$f(y_{t_1}, y_{t_2}, \cdots, y_{t_r}) = f(y_{t_1+\tau}, y_{t_2+\tau}, \cdots, y_{t_r+\tau})$$

All the moments are same for all τ.

2. **Auto-covariance Function (自己共分散関数) :**

$$\mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)) = \gamma(\tau), \qquad \tau = 0, 1, 2, \cdots$$

$$\gamma(\tau) = \gamma(-\tau)$$

3. **Auto-correlation Function (自己相関関数) :**

$$\rho(\tau) = \frac{\mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu))}{\sqrt{\mathrm{Var}(y_t)}\sqrt{\mathrm{Var}(y_{t-\tau})}} = \frac{\gamma(\tau)}{\gamma(0)}$$

Note that $\mathrm{Var}(y_t) = \mathrm{Var}(y_{t-\tau}) = \gamma(0)$.

4. **Sample Mean (標本平均) :**

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

5. **Sample Auto-covariance (標本自己共分散) :**

$$\hat{\gamma}(\tau) = \frac{1}{T}\sum_{t=\tau+1}^{T} (y_t - \hat{\mu})(y_{t-\tau} - \hat{\mu})$$

6. **Correlogram (コレログラム, or 標本自己相関関数) :**

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}$$

7. **Lag Operator (ラグ作要素) :**

$$L^{\tau} y_t = y_{t-\tau}, \qquad \tau = 1, 2, \cdots$$

8. **Likelihood Function (尤度関数)** — Innovation Form :

The joint distribution of $y_1, y_2, \cdots, y_T$ is written as:

$$
\begin{aligned}
f(y_1, , y_2, \cdots, y_T) &= f(y_T|y_{T-1}, \cdots, y_1) f(y_{T-1}, \cdots, y_1) \\
&= f(y_T|y_{T-1}, \cdots, y_1) f(y_{T-1}|y_{T-2}, \cdots, y_1) f(y_{T-2}, \cdots, y_1) \\
&\vdots \\
&= f(y_T|y_{T-1}, \cdots, y_1) f(y_{T-1}|y_{T-2}, \cdots, y_1) \;\cdots\; f(y_2|y_1) f(y_1) \\
&= f(y_1) \prod_{t=2}^{T} f(y_t|y_{t-1}, \cdots, y_1).
\end{aligned}
$$

Therefore, the log-likelihood function is given by:

$$\log f(y_1, y_2, \cdots, y_T) = \log f(y_1) + \sum_{t=2}^{T} \log f(y_t | y_{t-1}, \cdots, y_1).$$

Under the normality assumption, $f(y_t|y_{t-1}, \cdots, y_1)$ is given by the normal distribution with conditional mean $\mathrm{E}(y_t|y_{t-1}, \cdots, y_1)$ and conditional variance $\mathrm{Var}(y_t|y_{t-1}, \cdots, y_1)$.

14.2 Autoregressive Model (自己回帰モデル or AR モデル)

1. **AR(p) Model :**

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

which is rewritten as:

$$\phi(L) y_t = \epsilon_t,$$

where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p.$$

2. **Stationarity (定常性) :**

   Suppose that all the p solutions of x from $\phi(x) = 0$ are real numbers

   When the p solutions are greater than one, y_t is stationary.

   Suppose that the p solutions include imaginary numbers.

   When the p solutions are outside unit circle, y_t is stationary.

Example: AR(1) Model: $y_t = \phi_1 y_{t-1} + \epsilon_t$

1. The stationarity condition is: the solution of $\phi(x) = 1 - \phi_1 x = 0$, i.e., $x = 1/\phi_1$, is greater than one in absolute value, or equivalently, $-1 < \phi_1 < 1$.

2. Rewriting the AR(1) model,

$$\begin{aligned}
y_t &= \phi_1 y_{t-1} + \epsilon_t \\
&= \phi_1^2 y_{t-2} + \epsilon_t + \phi_1 \epsilon_{t-1} \\
&= \phi_1^3 y_{t-3} + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} \\
&\ \vdots \\
&= \phi_1^s y_{t-s} + \epsilon_t + \phi_1 \epsilon_{t-1} + \cdots + \phi_1^{s-1} \epsilon_{t-s+1}.
\end{aligned}$$

As s is large, ϕ_1^s approaches zero. $\Longrightarrow$ Stationarity condition

3. For stationarity, $y_t = \phi_1 y_{t-1} + \epsilon_t$ is rewritten as:

$$y_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots$$

4. Mean and Variance of AR(1) process, μ

$$\begin{aligned}
\mu = \mathrm{E}(y_t) &= \mathrm{E}(\epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots) \\
&= \mathrm{E}(\epsilon_t) + \phi_1 \mathrm{E}(\epsilon_{t-1}) + \phi_1^2 \mathrm{E}(\epsilon_{t-2}) + \cdots = 0 \\
\gamma(0) = \mathrm{V}(y_t) &= \mathrm{V}(\epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots) \\
&= \mathrm{V}(\epsilon_t) + \mathrm{V}(\phi_1 \epsilon_{t-1}) + \mathrm{V}(\phi_1^2 \epsilon_{t-2}) + \cdots \\
&= \sigma_\epsilon^2 + \phi_1^2 \sigma_\epsilon^2 + \phi_1^4 \sigma_\epsilon^2 + \cdots = \frac{\sigma_\epsilon^2}{1 - \phi_1^2}
\end{aligned}$$

5. Autocovariance and autocorrelation functions of the AR(1) process:

Rewriting the AR(1) process, we have:

$$y_t = \phi_1^\tau y_{t-\tau} + \epsilon_t + \phi_1 \epsilon_{t-1} + \cdots + \phi_1^{\tau-1} \epsilon_{t-\tau+1}.$$

Therefore, the autocovariance function of AR(1) process is:

$$\begin{aligned}
\gamma(\tau) &= \mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)) = \mathrm{E}(y_t y_{t-\tau}) \\
&= \mathrm{E}\Big((\phi_1^\tau y_{t-\tau} + \epsilon_t + \phi_1 \epsilon_{t-1} + \cdots + \phi_1^{\tau-1} \epsilon_{t-\tau+1}) y_{t-\tau}\Big) \\
&= \phi_1^\tau \mathrm{E}(y_{t-\tau} y_{t-\tau}) + \mathrm{E}(\epsilon_t y_{t-\tau}) + \phi_1 \mathrm{E}(\epsilon_{t-1} y_{t-\tau}) + \cdots + \phi_1^{\tau-1} \mathrm{E}(\epsilon_{t-\tau+1} y_{t-\tau}) \\
&= \phi_1^\tau \gamma(0).
\end{aligned}$$

The autocorrelation function of AR(1) process is:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \phi_1^\tau.$$

6. Another solution:

   Multiply $y_{t-\tau}$ on both sides of the AR(1) process and take the expectation:

$$\mathrm{E}(y_t y_{t-\tau}) = \phi_1 \mathrm{E}(y_{t-1} y_{t-\tau}) + \mathrm{E}(\epsilon_t y_{t-\tau})$$

$$\gamma(\tau) = \begin{cases} \phi_1 \gamma(\tau - 1), & \text{for } \tau \neq 0, \\ \phi_1 \gamma(\tau - 1) + \sigma_\epsilon^2, & \text{for } \tau = 0. \end{cases}$$

Using $\gamma(\tau) = \gamma(-\tau)$, $\gamma(\tau)$ for $\tau = 0$ is given by:

$$\gamma(0) = \phi_1\gamma(1) + \sigma_\epsilon^2 = \phi_1^2\gamma(0) + \sigma_\epsilon^2.$$

Note that $\gamma(1) = \phi_1\gamma(0)$.

Therefore, $\gamma(0)$ is given by:

$$\gamma(0) = \frac{\sigma_\epsilon^2}{1 - \phi_1^2}$$

7. Estimation of AR(1) model:

(a) Likelihood function

$$\begin{aligned}
\log f(y_T, \cdots, y_1) &= \log f(y_1) + \sum_{t=2}^{T} \log f(y_t | y_{t-1}, \cdots, y_1) \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma_\epsilon^2}{1-\phi_1^2}\right) - \frac{1}{\sigma_\epsilon^2/(1-\phi_1^2)} y_1^2 \\
&\quad -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma_\epsilon^2) - \frac{1}{\sigma_\epsilon^2}\sum_{t=2}^{T}(y_t - \phi_1 y_{t-1})^2
\end{aligned}$$

$$= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma_\epsilon^2) - \frac{1}{2}\log\left(\frac{1}{1-\phi_1^2}\right)$$

$$-\frac{1}{2\sigma_\epsilon^2/(1-\phi_1^2)}y_1^2 - \frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}(y_t - \phi_1 y_{t-1})^2$$

Note as follows:

$$f(y_1) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2/(1-\phi_1^2)}}\exp\left(-\frac{1}{2\sigma_\epsilon^2/(1-\phi_1^2)}y_1^2\right)$$

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

$$\frac{\partial \log f(y_T, \cdots, y_1)}{\partial \sigma_\epsilon^2} = -\frac{T}{2}\frac{1}{\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4/(1-\phi_1^2)}y_1^2 + \frac{1}{2\sigma_\epsilon^4}\sum_{t=2}^{T}(y_t - \phi_1 y_{t-1})^2 = 0$$

$$\frac{\partial \log f(y_T, \cdots, y_1)}{\partial \phi_1} = -\frac{\phi_1}{1-\phi_1^2} + \frac{\phi_1}{\sigma_\epsilon^2}y_1^2 + \frac{1}{\sigma_\epsilon^2}\sum_{t=2}^{T}(y_t - \phi_1 y_{t-1})y_{t-1} = 0$$

The MLE of ϕ_1 and σ_ϵ^2 satisfies the above two equation.

$$\tilde{\sigma}_\epsilon^2 = \frac{1}{T}\left((1-\tilde{\phi}_1^2)y_1^2 + \sum_{t=2}^{T}(y_t - \tilde{\phi}_1 y_{t-1})^2\right)$$

$$\tilde{\phi}_1 = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} + \left(\tilde{\phi}_1 y_1^2 - \frac{\tilde{\sigma}_\epsilon^2 \tilde{\phi}_1}{1-\tilde{\phi}_1^2}\right)\Big/\sum_{t=2}^{T} y_{t-1}^2$$

(b) Ordinary Least Squares (OLS) Method

$$S(\phi_1) = \sum_{t=2}^{T} (y_t - \phi_1 y_{t-1})^2$$

is minimized with respect to ϕ_1.

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{T} y_{t-1} y_t}{\sum_{t=2}^{T} y_{t-1}^2} = \phi_1 + \frac{\sum_{t=2}^{T} y_{t-1}\epsilon_t}{\sum_{t=2}^{T} y_{t-1}^2} = \phi_1 + \frac{(1/T)\sum_{t=2}^{T} y_{t-1}\epsilon_t}{(1/T)\sum_{t=2}^{T} y_{t-1}^2}$$
$$\longrightarrow \phi_1 + \frac{\mathrm{E}(y_{t-1}\epsilon_t)}{\mathrm{E}(y_{t-1}^2)} = \phi_1$$

OLSE of ϕ_1 is a consistent estimator.

The following equations are utilized.

$$\mathrm{E}(y_{t-1}\epsilon_t) = 0$$

$$\mathrm{E}(y_{t-1}^2) = \mathrm{Var}(y_{t-1}) = \gamma(0)$$

8. Asymptotic distribution of OLSE $\hat{\phi}_1$:

$$\sqrt{T}(\hat{\phi}_1 - \phi_1) \longrightarrow N(0, 1 - \phi_1^2)$$

Proof:

$y_{t-1}\epsilon_t$, $t = 1, 2, \cdots, T$, are distributed with mean zero and variance $\dfrac{\sigma_\epsilon^4}{1 - \phi_1^2}$.

From the central limit theorem,

$$\frac{(1/T)\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\sqrt{\sigma_\epsilon^4/(1-\phi_1^2)}/\sqrt{T}} \longrightarrow N(0,1)$$

Rewriting,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow N(0, \frac{\sigma_\epsilon^4}{1-\phi_1^2}).$$

Next,

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \mathrm{E}(y_{t-1}^2) = \gamma(0) = \frac{\sigma_\epsilon^2}{1-\phi_1^2}$$

yields:

$$\sqrt{T}(\hat{\phi}_1 - \phi_1) = \frac{(1/\sqrt{T})\sum_{t=1}^{T} y_{t-1}\epsilon_t}{(1/T)\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow N(0, 1-\phi_1^2)$$

9. Some formulas:

   (a) Central Limit Theorem

   Random variables $x_1, x_2, \cdots, x_T$ are mutually independently distributed with mean μ and variance σ^2.

   Define $\overline{x} = (1/T)\sum_{t=1}^{T} x_t$.

Then,

$$\frac{\overline{x} - \mathrm{E}(\overline{x})}{\sqrt{\mathrm{V}(\overline{x})}} = \frac{\overline{x} - \mu}{\sigma/\sqrt{T}} \longrightarrow N(0, 1)$$

(b) Central Limit Theorem II

Random variables x_1, x_2, $\cdots$, x_T are distributed with mean μ and variance σ^2.

Define $\overline{x} = (1/T)\sum_{t=1}^{T} x_t$.

Then,

$$\frac{\overline{x} - \mathrm{E}(\overline{x})}{\sqrt{\mathrm{V}(\overline{x})}} \longrightarrow N(0, 1)$$

(c) Let x and y be random variables.

y converges in distribution to a distribution, and x converges in probability to a fixed value.

Then, xy converges in distribution.

For example, consider:

$$y \longrightarrow N(\mu, \sigma^2), \qquad x \longrightarrow c.$$

Then, we obtain:

$$xy \longrightarrow N(c\mu, c^2\sigma^2)$$

10. **AR(1) +drift:** $y_t = \mu + \phi_1 y_{t-1} + \epsilon_t$

Mean:

Using the lag operator,

$$\phi(L) y_t = \mu + \epsilon_t$$

where $\phi(L) = 1 - \phi_1 L$.

Multiply $\phi(L)^{-1}$ on both sides. Then, when $|\phi_1| < 1$, we have:

$$y_t = \phi(L)^{-1}\mu + \phi(L)^{-1}\epsilon_t.$$

Taking the expectation on both sides,

$$
\begin{aligned}
\mathrm{E}(y_t) &= \phi(L)^{-1}\mu + \phi(L)^{-1}\mathrm{E}(\epsilon_t) \\
&= \phi(1)^{-1}\mu = \frac{\mu}{1-\phi_1}
\end{aligned}
$$

15 Unit Root (単位根) and Cointegration (共和分)

15.1 Unit Root (単位根) Test (Dickey-Fuller (DF) Test)

1. Why is a unit root problem important?

   (a) Economic variables increase over time in general.

   One of the assumptions of OLS is stationarity on y_t and x_t.

   This assumption implies that $\dfrac{1}{T}X'X$ converges to a fixed matrix as T is large.

That is, asymptotic normality of OLS estimator does not hold.

(b) In nonstationary time series, the unit root is the most important.

In the case of unit root, OLSE of the first-order autoregressive coefficient is consistent.

OLSE is $\sqrt{T}$-consistent in the case of stationary AR(1) process, but OLSE is T-consistent in the case of nonstationay AR(1) process.

(c) A lot of economic variables increase over time.

It is important to check an economic variable is trend stationary (i.e., $y_t = a_0 + a_1 t + \epsilon_t$) or difference stationary (i.e., $y_t = b_0 + y_{t-1} + \epsilon_t$).

Consider k-step ahead prediction for both cases.

$$\text{(Trend Stationarity)} \qquad y_{t+k|t} = a_0 + a_1(t+k)$$

$$\text{(Difference Stationarity)} \qquad y_{t+k|t} = b_0 k + y_t$$

2. **The Case of $|\phi_1| < 1$:**

$$y_t = \phi_1 y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{i.i.d. } N(0, \sigma_\epsilon^2), \qquad y_0 = 0, \qquad t = 1, \cdots, T$$

Then, OLSE of ϕ_1 is:

$$\hat{\phi}_1 = \frac{\displaystyle\sum_{t=1}^{T} y_{t-1} y_t}{\displaystyle\sum_{t=1}^{T} y_{t-1}^2}.$$

In the case of $|\phi_1| < 1$,

$$\hat{\phi}_1 = \phi_1 + \frac{\displaystyle\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\displaystyle\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow \phi_1 + \frac{\mathrm{E}(y_{t-1}\epsilon_t)}{\mathrm{E}(y_{t-1}^2)} = \phi_1.$$

Note as follows:

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow \mathrm{E}(y_{t-1}\epsilon_t) = 0.$$

By the central limit theorem,

$$\frac{\overline{y\epsilon} - \mathrm{E}(\overline{y\epsilon})}{\sqrt{\mathrm{V}(\overline{y\epsilon})}} \longrightarrow N(0,1)$$

where

$$\overline{y\epsilon} = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t.$$

$$
\begin{aligned}
\mathrm{E}(\overline{y\epsilon}) &= 0,\\
\mathrm{V}(\overline{y\epsilon}) &= \mathrm{V}\Big(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t\Big) = \mathrm{E}\Big(\big(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t\big)^2\Big)\\
&= \frac{1}{T^2}\mathrm{E}\Big(\sum_{t=1}^{T}\sum_{s=1}^{T} y_{t-1}y_{s-1}\epsilon_t\epsilon_s\Big) = \frac{1}{T^2}\mathrm{E}\Big(\sum_{t=1}^{T} y_{t-1}^2\epsilon_t^2\Big) = \frac{1}{T}\sigma_\epsilon^2\gamma(0).
\end{aligned}
$$

Therefore,

$$
\frac{\overline{y\epsilon}}{\sqrt{\sigma_\epsilon^2\gamma(0)/T}} = \frac{1}{\sigma_\epsilon\sqrt{\gamma(0)}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow N(0,1),
$$

which is rewritten as:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow N(0, \sigma_\epsilon^2\gamma(0)).$$

Using $\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \mathrm{E}(y_{t-1}^2) = \gamma(0)$, we have the following asymptotic distribution:

$$\sqrt{T}(\hat{\phi}_1 - \phi_1) = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow N\left(0, \frac{\sigma_\epsilon^2}{\gamma(0)}\right) = N\left(0, 1-\phi_1^2\right).$$

Note that $\gamma(0) = \dfrac{\sigma_\epsilon^2}{1 - \phi_1^2}$.

3. In the case of $\phi_1 = 1$, as expected, we have:

$$\sqrt{T}(\hat{\phi}_1 - 1) \longrightarrow 0.$$

That is, $\hat{\phi}_1$ has the distribution which converges in probability to $\phi_1 = 1$ (i.e., degenerated distribution).

Is this true?

4. **The Case of $\boldsymbol{\phi_1 = 1}$:** $\Longrightarrow$ Random Walk Process

$y_t = y_{t-1} + \epsilon_t$ with $y_0 = 0$ is written as:

$$y_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \cdots + \epsilon_1.$$

Therefore, we can obtain:

$$y_t \sim N(0, \sigma_\epsilon^2 t).$$

The variance of y_t depends on time t. $\implies$ y_t is nonstationary.

5. Remember that $\hat{\phi}_1 = \phi_1 + \dfrac{\sum y_{t-1}\epsilon_t}{\sum y_{t-1}^2}$.

   (a) First, consider the numerator $\sum y_{t-1}\epsilon_t$.

We have $y_t^2 = (y_{t-1} + \epsilon_t)^2 = y_{t-1}^2 + 2y_{t-1}\epsilon_t + \epsilon_t^2$.

Therefore, we obtain:

$$y_{t-1}\epsilon_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - \epsilon_t^2).$$

Taking into account $y_0 = 0$, we have:

$$\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}y_T^2 - \frac{1}{2}\sum_{t=1}^{T}\epsilon_t^2.$$

Divided by $\sigma_\epsilon^2 T$ on both sides, we have the following:

$$\frac{1}{\sigma_\epsilon^2 T}\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}\left(\frac{y_T}{\sigma_\epsilon\sqrt{T}}\right)^2 - \frac{1}{2\sigma_\epsilon^2}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2.$$

From $y_t \sim N(0, \sigma_\epsilon^2 t)$, we obtain the following result:

$$\left(\frac{y_T}{\sigma_\epsilon \sqrt{T}}\right)^2 \sim \chi^2(1).$$

Moreover, the second term is derived from:

$$\frac{1}{T}\sum_{t=1}^{T} \epsilon_t^2 \longrightarrow \mathrm{E}(\epsilon_t^2) = \sigma_\epsilon^2.$$

Therefore,

$$\frac{1}{\sigma_\epsilon^2 T}\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}\left(\frac{y_T}{\sigma_\epsilon \sqrt{T}}\right)^2 - \frac{1}{2\sigma_\epsilon^2}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \longrightarrow \frac{1}{2}(\chi^2(1) - 1).$$

(b) Next, consider $\sum y_{t-1}^2$.

$$\mathrm{E}\left(\sum_{t=1}^{T} y_{t-1}^2\right) = \sum_{t=1}^{T} \mathrm{E}(y_{t-1}^2) = \sum_{t=1}^{T} \sigma_\epsilon^2 (t-1) = \sigma_\epsilon^2 \frac{T(T-1)}{2}.$$

Thus, we obtain the following result:

$$\frac{1}{T^2}\mathrm{E}\left(\sum_{t=1}^{T} y_{t-1}^2\right) \longrightarrow \text{ a fixed value.}$$

Therefore,

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \text{ a distribution.}$$

6. Summarizing the results up to now, $T(\hat{\phi}_1 - \phi_1)$, not $\sqrt{T}(\hat{\phi}_1 - \phi_1)$, has limiting distribution in the case of $\phi_1 = 1$.

$$T(\hat{\phi}_1 - \phi_1) = \frac{(1/T)\sum y_{t-1}\epsilon_t}{(1/T^2)\sum y_{t-1}^2} \longrightarrow \text{a distribution.}$$

7. Basic Concepts of Random Walk Process:

   (a) Model: $y_t = y_{t-1} + \epsilon_t, \qquad y_0 = 0, \qquad \epsilon_t \sim N(0,1).$

   Then,

$$y_t = \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1.$$

Therefore,

$$y_t \sim N(0, t).$$

$\Longrightarrow$ Nonstationary Process (i.e., variance depends on time t.)

Difference between y_s and y_t ($s > t$) is:

$$y_s - y_t = \epsilon_s + \epsilon_{s-1} + \cdots + \epsilon_{t+2} + \epsilon_{t+1}.$$

The distribution of $y_s - y_t$ is:

$$y_s - y_t \sim N(0, s - t).$$

(b) Rewrite as follows:

$$\begin{aligned} y_t &= y_{t-1} + \epsilon_t \\ &= y_{t-1} + e_{1,t} + e_{2,t} + \cdots + e_{N,t}, \end{aligned}$$

where $\epsilon_t = e_{1,t} + e_{2,t} + \cdots + e_{N,t}$.

$e_{1,t}, e_{2,t}, \cdots, e_{N,t}$ are iid with $e_{i,t} \sim N(0, 1/N)$.

That is, suppose that there are N subperiods between time t and time $t+1$.

The limit when $N \to \infty$ is a **continuous time (連続時間)** process known as **standard Brownian motion** or **Wiener process**.

The value of this process at time r is denoted by $W(r)$ for $0 \le r \le 1$.

Definition:

Standard Brownian motion $W(r)$ denotes a continuous-time variable at time r and a stochastic function.

$W(r)$ for $r \in [0, 1]$ satisfies the following:

i. $W(0) = 0$

ii. For any time periods $0 \leq r_1 < r_2 < \cdots < r_k \leq 1$, $W(r_2) - W(r_1)$, $W(r_3) - W(r_2)$, $\cdots$, $W(r_k) - W(r_{k-1})$ are independently multivariate normal with $W(s) - W(t) \sim N(0, s - t)$ for $s > t$.

iii. $W(r)$ is continuous in r with probability 1.

An example:

$$\sigma W(r) \sim N(0, \sigma^2 r),$$

which denotes the Brownian motion with variance σ^2.

Another example;

$$W(r)^2 \sim r \times \chi^2(1).$$

(c) Assume $\epsilon_t \sim$ iid $(0, \sigma_\epsilon^2)$. Define $X_T(r)$ for $r \in [0,1]$ as follows:

$$X_T(r) = \begin{cases} 0, & 0 \le r < \dfrac{1}{T} \\ \dfrac{\epsilon_1}{T}, & \dfrac{1}{T} \le r < \dfrac{2}{T} \\ \dfrac{\epsilon_1 + \epsilon_2}{T}, & \dfrac{2}{T} \le r < \dfrac{3}{T} \\ \vdots & \vdots \\ \dfrac{\epsilon_1 + \epsilon_2 + \cdots + \epsilon_T}{T}, & r = 1 \end{cases}$$

Let $[Tr]$ be the largest integer which is less than or equal to $T \times r$.

$$X_T(r) \equiv \frac{1}{T} \sum_{t=1}^{[Tr]} \epsilon_t, \qquad \sqrt{T} X_T(r) \longrightarrow N(0, r\sigma_\epsilon^2).$$

Note that

$$\frac{1}{T} \sum_{t=1}^{[Tr]} \epsilon_t = \frac{[Tr]}{T} \frac{1}{[Tr]} \sum_{t=1}^{[Tr]} \epsilon_t,$$

$$\frac{[Tr]}{T} \longrightarrow r, \qquad \frac{1}{\sqrt{[Tr]}} \sum_{t=1}^{[Tr]} \epsilon_t \longrightarrow N(0, \sigma_\epsilon^2),$$

$$\sqrt{T} X_T(r) = \frac{[Tr]}{T} \sqrt{\frac{T}{[Tr]}} \frac{1}{\sqrt{[Tr]}} \sum_{t=1}^{[Tr]} \epsilon_t, \qquad \sqrt{\frac{T}{[Tr]}} \longrightarrow \frac{1}{\sqrt{r}}.$$

Therefore, we obtain:

$$\sqrt{T} X_T(r) \longrightarrow N(0, r\sigma_\epsilon^2).$$

Moreover, we have the following results:

$$\frac{\sqrt{T} X_T(r)}{\sigma_\epsilon} \longrightarrow N(0, r) = W(r),$$

$$\frac{\sqrt{T}\,(X_T(r_2) - X_T(r_1))}{\sigma_\epsilon} \longrightarrow W(r_2) - W(r_1) = N(0, r_2 - r_1).$$

For example, consider:

$$X_T(1) = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t.$$

Then,

$$\frac{\sqrt{T}X_T(1)}{\sigma_\epsilon} = \frac{1}{\sigma_\epsilon\sqrt{T}}\sum_{t=1}^{T}\epsilon_t \longrightarrow W(1) = N(0,1).$$

(d) Consider $y_t = y_{t-1} + \epsilon_t,\ y_0 = 0$ and $\epsilon_t \sim N(0, \sigma_\epsilon^2)$.

$X_T(r)$ is defined as follows:

$$X_T(r) = \begin{cases} 0, & 0 \leq r < \frac{1}{T}, \\ \frac{y_1}{T}, & \frac{1}{T} \leq r < \frac{2}{T}, \\ \frac{y_2}{T}, & \frac{2}{T} \leq r < \frac{3}{T}, \\ \quad\vdots & \quad\vdots \\ \frac{y_{T-1}}{T}, & \frac{T-1}{T} \leq r < 1, \\ \frac{y_T}{T}, & r = 1. \end{cases}$$

Define $S_T(r)$ as follows:

$$S_T(r) = \begin{cases} 0, & 0 \le r < \frac{1}{T}, \\ \frac{y_1^2}{T}, & \frac{1}{T} \le r < \frac{2}{T}, \\ \frac{y_2^2}{T}, & \frac{2}{T} \le r < \frac{3}{T}, \\ \vdots & \vdots \\ \frac{y_{T-1}^2}{T}, & \frac{T-1}{T} \le r < 1, \\ \frac{y_T^2}{T}, & r = 1. \end{cases}$$

To obtain $\int_0^1 X_T(r)\mathrm{d}r$ and $\int_0^1 S_T(r)\mathrm{d}r$, we compute a sum of rectangulars as follows:

$$\int_0^1 X_T(r)\mathrm{d}r \approx \frac{y_1}{T}\left(\frac{2}{T}-\frac{1}{T}\right)+\frac{y_2}{T}\left(\frac{3}{T}-\frac{2}{T}\right)+\ \cdots\ +\frac{y_{T-1}}{T}\left(1-\frac{T-1}{T}\right)$$

$$=\frac{y_1}{T^2}+\frac{y_2}{T^2}+\ \cdots\ +\frac{y_{T-1}}{T^2}=\frac{1}{T^2}\sum_{t=1}^{T} y_t,$$

$$\int_0^1 S_T(r)\mathrm{d}r \approx \frac{y_1^2}{T}\left(\frac{2}{T}-\frac{1}{T}\right)+\frac{y_2^2}{T}\left(\frac{3}{T}-\frac{2}{T}\right)+\ \cdots\ +\frac{y_{T-1}^2}{T}\left(1-\frac{T-1}{T}\right)$$

$$=\frac{y_1^2}{T^2}+\frac{y_2^2}{T^2}+\ \cdots\ +\frac{y_{T-1}^2}{T^2}=\frac{1}{T^2}\sum_{t=1}^{T} y_t^2.$$

We have already known that $\sqrt{T}X_T(r) \longrightarrow \sigma_\epsilon W(r)$.

Therefore,

$$\int_0^1 \sqrt{T}X_T(r)\mathrm{d}r \longrightarrow \sigma_\epsilon \int_0^1 W(r)\mathrm{d}r.$$

That is,

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} y_t \longrightarrow \sigma_\epsilon \int_0^1 W(r)\mathrm{d}r.$$

From $S_T(r) \equiv \left(\sqrt{T} X_T(r)\right)^2$,

$$S_T(r) \equiv \left(\sqrt{T} X_T(r)\right)^2 \longrightarrow \sigma_\epsilon^2 \left(W(r)\right)^2,$$

which is called the continuous mapping theorem.

(*) **Continuous Mapping Theorem (連続写像定理):**

if $x_T \longrightarrow x$ (convergence in distribution) and $g(\cdot)$ is a continuous function, then $g(x_T) \longrightarrow g(x)$ (convergence in distribution).

Threfore, we have the follwoing result:

$$\int_0^1 S_T(r)\mathrm{d}r \longrightarrow \sigma_\epsilon^2 \int_0^1 (W(r))^2 \, \mathrm{d}r.$$

That is,

$$\frac{1}{T^2} \sum_{t=1}^{T} y_t^2 \longrightarrow \sigma_\epsilon^2 \int_0^1 (W(r))^2 \, \mathrm{d}r.$$

8. Asymptotic Distribution of AR(1) Model:

(a) $\boldsymbol{H_0 : y_t = y_{t-1} + \epsilon_t}$ **and** $\boldsymbol{H_1 : y_t = \phi_1 y_{t-1} + \epsilon_t}$ **for** $\boldsymbol{|\phi_1| < 1}$

OLSE of ϕ_1, denoted by $\hat{\phi}_1$, is given by:

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2} = \phi_1 + \frac{\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

Using $\phi_1 = 1$ and some formulas shown above, we obtain:

$$T(\hat{\phi}_1 - 1) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\epsilon_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow \frac{\frac{1}{2}\Big((W(1))^2 - 1\Big)}{\int_0^1 (W(r))^2 \,\mathrm{d}r}$$

Remember that

$$T^{-1} \sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow \frac{1}{2}\sigma_\epsilon^2 \left((W(1))^2 - 1\right)$$

and

$$T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \sigma_\epsilon^2 \int_0^1 (W(r))^2 \, \mathrm{d}r,$$

where $(W(1))^2 = \chi^2(1)$.

We say that $\hat{\phi}_1$ is **super-consistent (超一致性)** or ***T*-consistent**.

Remember that when $|\phi_1| < 1$ we have $\sqrt{T}(\hat{\phi}_1 - \phi_1) \longrightarrow N(0, 1 - \phi_1^2)$, and in this case we say that $\hat{\phi}_1$ is $\boldsymbol{\sqrt{T}}$**-consistent**.

Conventional t test statistic is given by:

$$t = \frac{\hat{\phi}_1 - 1}{s_\phi},$$

where

$$s_\phi = \left(s^2 \Big/ \sum_{t=1}^{T} y_{t-1}^2 \right)^{1/2} \quad \text{and} \quad s^2 = \frac{1}{T-1} \sum_{t=1}^{T} (y_t - \hat{\phi}_1 y_{t-1})^2.$$

Next, consider t statistic.

The t test statistic, denoted by t, is represented as follows:

$$t = \frac{\hat{\phi}_1 - 1}{s_\phi} = \frac{T(\hat{\phi}_1 - 1)}{T s_\phi}$$

The denominator is:

$$\begin{aligned} T s_\phi &= \left(s^2 \Big/ \frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^2 \right)^{1/2} \\ &\longrightarrow \left(\sigma_\epsilon^2 \Big/ \left(\sigma_\epsilon^2 \int_0^1 (W(r))^2 \, \mathrm{d}r \right) \right)^{1/2} = \left(\int_0^1 (W(r))^2 \, \mathrm{d}r \right)^{-1/2}, \end{aligned}$$

where $s^2 \longrightarrow \sigma_\epsilon^2$ is utilized.

Therefore, we have the following asymptotic distribution:

$$t = \frac{\hat{\phi}_1 - 1}{s_\phi} \longrightarrow \frac{\frac{1}{2}\left((W(1))^2 - 1\right)}{\int_0^1 (W(r))^2 \, \mathrm{d}r} \Bigg/ \left(\int_0^1 (W(r))^2 \, \mathrm{d}r\right)^{-1/2}$$

$$= \frac{\frac{1}{2}\left((W(1))^2 - 1\right)}{\left(\int_0^1 (W(r))^2 \, \mathrm{d}r\right)^{1/2}}.$$

Therefore, the distribution of the t statistic shown above is different from the t distribution.

(b) $\boldsymbol{H_0 : y_t = y_{t-1} + \epsilon_t}$ **and** $\boldsymbol{H_1 : y_t = \alpha_0 + \phi_1 y_{t-1} + \epsilon_t}$ **for** $\boldsymbol{|\phi_1| < 1}$

$$\begin{pmatrix} \hat{\alpha}_0 \\ \hat{\phi}_1 \end{pmatrix} = \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_t \\ \sum y_{t-1} y_t \end{pmatrix}$$

$$= \begin{pmatrix} \alpha_0 \\ \phi_1 \end{pmatrix} + \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum \epsilon_t \\ \sum y_{t-1} \epsilon_t \end{pmatrix}$$

In the true model, $\alpha_0 = 0$ and $\phi_1 = 1$.

$$\begin{pmatrix} \hat{\alpha}_0 \\ \hat{\phi}_1 - 1 \end{pmatrix} = \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum \epsilon_t \\ \sum y_{t-1} \epsilon_t \end{pmatrix}$$

$$= \begin{pmatrix} O_p(T) & O_p(T^{3/2}) \\ O_p(T^{3/2}) & O_p(T^2) \end{pmatrix}^{-1} \begin{pmatrix} O_p(T^{1/2}) \\ O_p(T) \end{pmatrix}$$

(*) For random variable x and constant k, $x = O_p(k)$ implies that x/k converges in distribution.

To change each element of the matrices to $O_p(1)$, we use the following matrix:

$$\Gamma = \begin{pmatrix} T^{1/2} & 0 \\ 0 & T \end{pmatrix}.$$

Multiplying the above matrix from the left, we obtain the following:

$$\Gamma \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\phi}_1 - 1 \end{pmatrix} = \begin{pmatrix} T^{1/2}\hat{\alpha}_0 \\ T(\hat{\phi}_1 - 1) \end{pmatrix} = \Gamma \begin{pmatrix} O_p(T) & O_p(T^{3/2}) \\ O_p(T^{3/2}) & O_p(T^2) \end{pmatrix}^{-1} \Gamma\Gamma^{-1} \begin{pmatrix} O_p(T^{1/2}) \\ O_p(T) \end{pmatrix}$$

$$= \left(\Gamma^{-1}\begin{pmatrix} O_p(T) & O_p(T^{3/2}) \\ O_p(T^{3/2}) & O_p(T^2) \end{pmatrix}\Gamma^{-1}\right)^{-1}\Gamma^{-1}\begin{pmatrix} O_p(T^{1/2}) \\ O_p(T) \end{pmatrix}$$

$$= \left(\Gamma^{-1}\begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}\Gamma^{-1}\right)^{-1}\Gamma^{-1}\begin{pmatrix} \sum \epsilon_t \\ \sum y_{t-1}\epsilon_t \end{pmatrix}$$

$$= \begin{pmatrix} 1 & T^{-3/2}\sum y_{t-1} \\ T^{-3/2}\sum y_{t-1} & T^{-2}\sum y_{t-1}^2 \end{pmatrix}^{-1}\begin{pmatrix} T^{-1/2}\sum \epsilon_t \\ T^{-1}\sum y_{t-1}\epsilon_t \end{pmatrix}.$$

Each matrix converges in distribution as follows:

$$\begin{pmatrix} 1 & T^{-3/2}\sum y_{t-1} \\ T^{-3/2}\sum y_{t-1} & T^{-2}\sum y_{t-1}^2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \sigma_\epsilon \int_0^1 W(r)\mathrm{d}r \\ \sigma_\epsilon \int_0^1 W(r)\mathrm{d}r & \sigma_\epsilon^2 \int_0^1 (W(r))^2\,\mathrm{d}r \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\epsilon \end{pmatrix} \begin{pmatrix} 1 & \int_0^1 W(r)\mathrm{d}r \\ \int_0^1 W(r)\mathrm{d}r & \int_0^1 (W(r))^2\,\mathrm{d}r \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\epsilon \end{pmatrix},$$

$$\begin{pmatrix} T^{-1/2}\sum \epsilon_t \\ T^{-1}\sum y_{t-1}\epsilon_t \end{pmatrix} \longrightarrow \begin{pmatrix} \sigma_\epsilon W(1) \\ \frac{1}{2}\sigma_\epsilon^2\left((W(1))^2 - 1\right) \end{pmatrix} = \sigma_\epsilon \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\epsilon \end{pmatrix} \begin{pmatrix} W(1) \\ \frac{1}{2}\left((W(1))^2 - 1\right) \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} T^{1/2}\hat{\alpha}_0 \\ T(\hat{\phi}_1 - 1) \end{pmatrix} \longrightarrow \left[\begin{pmatrix} 1 & 0 \\ 0 & \sigma_\epsilon \end{pmatrix} \begin{pmatrix} 1 & \int_0^1 W(r)\mathrm{d}r \\ \int_0^1 W(r)\mathrm{d}r & \int_0^1 (W(r))^2\,\mathrm{d}r \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\epsilon \end{pmatrix} \right]^{-1}$$

$$\times \sigma_\epsilon \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\epsilon \end{pmatrix} \begin{pmatrix} W(1) \\ \frac{1}{2}\left((W(1))^2 - 1\right) \end{pmatrix}.$$

Finally, $T(\hat{\phi}_1 - 1)$ converges to the following distribution:

$$T(\hat{\phi}_1 - 1) \longrightarrow \frac{\frac{1}{2}\Big((W(1))^2 - 1\Big) - W(1) \displaystyle\int_0^1 W(r)\mathrm{d}r}{\displaystyle\int_0^1 (W(r))^2\,\mathrm{d}r - \left(\int_0^1 W(r)\mathrm{d}r\right)^2}.$$

The t test statistic is:

$$t = \frac{\hat{\phi}_1 - 1}{\left(s_\phi^2\right)^{1/2}} = \frac{T(\hat{\phi}_1 - 1)}{\left(T^2 s_\phi^2\right)^{1/2}},$$

where

$$s_\phi^2 = s^2 \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$s^2 = \frac{1}{T-2} \sum_{t=1}^{T} (y_t - \hat{\alpha}_0 - \hat{\phi}_1 y_{t-1})^2.$$

The denominator $T^2 s_\phi^2$ converges in distribution as follows:

$$T^2 s_\phi^2 \longrightarrow \sigma_\epsilon^2 \begin{pmatrix} 0 & 1 \end{pmatrix} \left(\begin{pmatrix} 1 & 0 \\ 0 & \sigma_\epsilon \end{pmatrix} \begin{pmatrix} 1 & \int_0^1 W(r)\mathrm{d}r \\ \int_0^1 W(r)\mathrm{d}r & \int_0^1 (W(r))^2\,\mathrm{d}r \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\epsilon \end{pmatrix} \right)^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$= \frac{1}{\int_0^1 (W(r))^2\,\mathrm{d}r - \left(\int_0^1 W(r)\mathrm{d}r \right)^2}$$

Thus, the t test statistic converges to the following distribution:

$$t \longrightarrow \frac{\frac{1}{2}\left((W(1))^2 - 1\right) - W(1)\int_0^1 W(r)\mathrm{d}r}{\left(\int_0^1 (W(r))^2\,\mathrm{d}r - \left(\int_0^1 W(r)\mathrm{d}r\right)^2\right)^{1/2}}.$$

(c) $\boldsymbol{H_0 : y_t = \alpha_0 + y_{t-1} + \epsilon_t}$ **and** $\boldsymbol{H_1 : y_t = \alpha_0 + \phi_1 y_{t-1} + \epsilon_t}$ **for** $|\boldsymbol{\phi_1}| < 1$

$$\begin{pmatrix} T^{1/2}(\hat{\alpha}_0 - \alpha_0) \\ T^{3/2}(\hat{\phi}_1 - 1) \end{pmatrix} \longrightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_\epsilon^2 \begin{pmatrix} 1 & \frac{\alpha_0}{2} \\ \frac{\alpha_0}{2} & \frac{\alpha_0^2}{3} \end{pmatrix}\right).$$

(abbr.)

(d) $\boldsymbol{H_0 : y_t = \alpha_0 + y_{t-1} + \epsilon_t}$ **and**

$\boldsymbol{H_1 : y_t = \alpha_0 + \alpha_1 t + \phi_1 y_{t-1} + \epsilon_t}$ **for** $|\boldsymbol{\phi_1}| < \mathbf{1}$

(abbr.)

9. The distributions of the t statistic: $\dfrac{\hat{\phi}_1 - 1}{s_\phi}$

t Distribution

T	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
25	–2.49	–2.06	–1.71	–1.32	1.32	1.71	2.06	2.49
50	–2.40	–2.01	–1.68	–1.30	1.30	1.68	2.01	2.40
100	–2.36	–1.98	–1.66	–1.29	1.29	1.66	1.98	2.36
250	–2.34	–1.97	–1.65	–1.28	1.28	1.65	1.97	2.34
500	–2.33	–1.96	–1.65	–1.28	1.28	1.65	1.96	2.33
∞	–2.33	–1.96	–1.64	–1.28	1.28	1.64	1.96	2.33

(a) $H_0: \; y_t = y_{t-1} + \epsilon_t$

$H_1: \; y_t = \phi_1 y_{t-1} + \epsilon_t$ **for** $\phi_1 < 1$ **or** $-1 < \phi_1$

T	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
25	−2.66	−2.26	−1.95	−1.60	0.92	1.33	1.70	2.16
50	−2.62	−2.25	−1.95	−1.61	0.91	1.31	1.66	2.08
100	−2.60	−2.24	−1.95	−1.61	0.90	1.29	1.64	2.03
250	−2.58	−2.23	−1.95	−1.62	0.89	1.29	1.63	2.01
500	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	2.00
∞	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	2.00

(b) $\boldsymbol{H_0 : \; y_t = y_{t-1} + \epsilon_t}$

$\boldsymbol{H_1 : \; y_t = \alpha_0 + \phi_1 y_{t-1} + \epsilon_t}$ **for** $\boldsymbol{\phi_1 < 1}$ **or** $\boldsymbol{-1 < \phi_1}$

T	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
25	−3.75	−3.33	−3.00	−2.63	−0.37	0.00	0.34	0.72
50	−3.58	−3.22	−2.93	−2.60	−0.40	−0.03	0.29	0.66
100	−3.51	−3.17	−2.89	−2.58	−0.42	−0.05	0.26	0.63
250	−3.46	−3.14	−2.88	−2.57	−0.42	−0.06	0.24	0.62
500	−3.44	−3.13	−2.87	−2.57	−0.43	−0.07	0.24	0.61
∞	−3.43	−3.12	−2.86	−2.57	−0.44	−0.07	0.23	0.60

(d) $\boldsymbol{H_0: \; y_t = \alpha_0 + y_{t-1} + \epsilon_t}$

$\boldsymbol{H_1: \; y_t = \alpha_0 + \alpha_1 t + \phi_1 y_{t-1} + \epsilon_t}$ **for** $\boldsymbol{\phi_1 < 1}$ **or** $\boldsymbol{-1 < \phi_1}$

T	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
25	–4.38	–3.95	–3.60	–3.24	–1.14	–0.80	–0.50	–0.15
50	–4.15	–3.80	–3.50	–3.18	–1.19	–0.87	–0.58	–0.24
100	–4.04	–3.73	–3.45	–3.15	–1.22	–0.90	–0.62	–0.28
250	–3.99	–3.69	–3.43	–3.13	–1.23	–0.92	–0.64	–0.31
500	–3.98	–3.68	–3.42	–3.13	–1.24	–0.93	–0.65	–0.32
∞	–3.96	–3.66	–3.41	–3.12	–1.25	–0.94	–0.66	–0.33

15.2 Serially Correlated Errors

Consider the case where the error term is serially correlated.

15.2.1 Augmented Dickey-Fuller (ADF) Test

Consider the following AR(p) model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t, \qquad \epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2),$$

which is rewritten as:

$$\phi(L) y_t = \epsilon_t.$$

When the above model has a unit root, we have $\phi(1) = 0$, i.e., $\phi_1 + \phi_2 + \cdots + \phi_p = 1$. The above AR($p$) model is written as:

$$y_t = \rho y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + +\delta_{p-1} \Delta y_{t-p+1} + \epsilon_t,$$

where $\rho = \phi_1 + \phi_2 + \cdots + \phi_p$ and $\delta_j = -(\phi_{j+1} + \phi_{j+2} + \cdots + \phi_p)$.

The null and alternative hypotheses are:

$$H_0:\ \rho = 1 \text{ (Unit root)},$$

$$H_1:\ \rho < 1 \text{ (Stationary)}.$$

Use the t test, where we have the same asymptotic distributions.

We can utilize the same tables as before.

Choose p by AIC or SBIC.

Use $N(0,1)$ to test $H_0: \ \delta_j = 0$ against $H_1: \ \delta_j \neq 0$ for $j = 1, 2, \cdots, p-1$.

Reference

Kurozumi (2008) “Economic Time Series Analysis and Unit Root Tests: Development and Perspective,” *Japan Statistical Society*, Vol.38, Series J, No.1, pp.39 – 57.

Download the above paper from:

`http://ci.nii.ac.jp/vol_issue/nels/AA11989749/ISS0000426576_ja.html`

15.3 Cointegration (共和分)

1. For a scalar y_t, when $(1-L)^d y_t$ is stationary, we write $y_t \sim I(d)$.

   When $\Delta y_t = y_t - y_{t-1}$ is stationary, we write $\Delta y_t \sim I(0)$ or $y_t \sim I(1)$.

2. **Definition of Cointegration:**

   Suppose that each series in a $g \times 1$ vector y_t is $I(1)$, i.e., each series has unit root, and that a linear combination of each series (i.e, $a'y_t$ for a nonzero vector a) is $I(0)$, i.e., stationary.

   Then, we say that y_t has a cointegration.

3. **Example:**

Suppose that $y_t = (y_{1,t},\ y_{2,t})'$ is the following vector autoregressive process:

$$y_{1,t} = \gamma y_{2,t} + \epsilon_{1,t},$$

$$y_{2,t} = y_{2,t-1} + \epsilon_{2,t}.$$

Then,

$$\Delta y_{1,t} = \gamma \epsilon_{2,t} + \epsilon_{1,t} - \epsilon_{1,t-1}, \quad \text{(MA(1) process)},$$

$$\Delta y_{2,t} = \epsilon_{2,t},$$

where both $y_{1,t}$ and $y_{2,t}$ are $I(1)$ processes.

The linear combination $y_{1,t} - \gamma y_{2,t}$ is $I(0)$.

In this case, we say that $y_t = (y_{1,t},\ y_{2,t})'$ is cointegrated with $a = (1,\ -\gamma)$.

$a = (1,\ -\gamma)$ is called the **cointegrating vector (共和分ベクトル)**, which is not unique. Therefore, the first element of a is set to be one.

4. Suppose that $y_t \sim I(1)$ and $x_t \sim I(1)$.

   For the regression model $y_t = x_t\beta + u_t$, OLS does not work well if we do not have the β which satisfies $u_t \sim I(0)$.

   $\Longrightarrow$ **Spurious regression (見せかけの回帰)**

5. Suppose that $y_t \sim I(1)$, y_t is a $g \times 1$ vector and $y_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix}$.

$y_{2,t}$ is a $k \times 1$ vector, where $k = g - 1$.

Consider the following regression model:

$$y_{1,t} = \alpha + \gamma' y_{2,t} + u_t, \qquad t = 1, 2, \cdots, T.$$

OLSE is given by:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \sum y'_{2,t} \\ \sum y_{2,t} & \sum y_{2,t} y'_{2,t} \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{1,t} \\ \sum y_{1,t} y_{2,t} \end{pmatrix}.$$

Next, consider testing the null hypothesis $H_0: \; R\gamma = d$, where R is a $G \times k$

matrix ($G \leq k$) and r is a $G \times 1$ vector. G denotes the number of the linear restrictions.

The F statistic, denoted by F, is given by:

$$F = \frac{1}{G}(R\hat{\gamma} - d)' \left(s^2 \begin{pmatrix} 0 & R \end{pmatrix} \begin{pmatrix} T & \sum y'_{2,t} \\ \sum y_{2,t} & \sum y_{2,t} y'_{2,t} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ R' \end{pmatrix} \right)^{-1} (R\hat{\gamma} - d),$$

where

$$s^2 = \frac{1}{T - g} \sum_{t=1}^{T} (y_{1,t} - \hat{\alpha} - \hat{\gamma}' y_{2,t})^2.$$

When we have the γ such that $y_{1,t} - \gamma y_{2,t}$ is stationary, OLSE of γ, i.e., $\hat{\gamma}$, is not statistically equal to zero.

When the sample size T is large enough, H_0 is rejected by the F test.

6. Phillips, P.C.B. (1986) "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, Vol.33, pp.95 – 131.

Consider a $g \times 1$ vector y_t whose first difference is described by:

$$\Delta y_t = \Psi(L)\epsilon_t = \sum_{s=0}^{\infty} \Psi_s \epsilon_{t-s},$$

for ϵ_t an i.i.d. $g \times 1$ vector with mean zero , variance $\mathrm{E}(\epsilon_t \epsilon_t') = PP'$, and finite fourth moments and where $\{s\Psi_s\}_{s=0}^{\infty}$ is absolutely summable.

Let $k = g - 1$ and $\Lambda = \Psi(1)P$.

Partition y_t as $y_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix}$ and $\Lambda\Lambda'$ as $\Lambda\Lambda' = \begin{pmatrix} \Sigma_{11} & \Sigma'_{21} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, where $y_{1,t}$ and Σ_{11} are scalars, $y_{2,t}$ and Σ_{21} are $k \times 1$ vectors, and Σ_{22} is a $k \times k$ matrix.

Suppose that $\Lambda\Lambda'$ is nonsingular,and define $\sigma_1^2 = \Sigma_{11} - \Sigma'_{21}\Sigma_{22}^{-1}\Sigma_{21}$.

Let L_{22} denote the Cholesky factor of Σ_{22}^{-1}, i.e., L_{22} is the lower triangular matrix satisfying $\Sigma_{22}^{-1} = L_{22}L'_{22}$.

Then, (a) – (c) hold.

(a) OLSEs of α and γ in the regression model $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$, denoted by $\hat{\alpha}_T$ and $\hat{\gamma}_T$, are characterized by:

$$\begin{pmatrix} T^{-1/2}\hat{\alpha}_T \\ \hat{\gamma}_T - \Sigma_{22}^{-1}\Sigma_{21} \end{pmatrix} \longrightarrow \begin{pmatrix} \sigma_1 h_1 \\ \sigma_1 L_{22} h_2 \end{pmatrix},$$

where

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} 1 & \int_0^1 W_2(r)' \mathrm{d}r \\ \int_0^1 W_2(r)\mathrm{d}r & \int_0^1 W_2(r)W_2(r)'\mathrm{d}r \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 W_1(r)\mathrm{d}r \\ \int_0^1 W_2(r)W_1(r)\mathrm{d}r \end{pmatrix},$$

where $W_1(r)$ and $W_2(r)$ denote scalar and g-dimensional standard Brownian motions, and $W_1(r)$ is independent of $W_2(r)$.

(b) The sum of squared residuals, denoted by $\mathrm{RSS}_T = \sum_{t=1}^{T} \hat{u}_t^2$, satisfies

$$T^{-2}\mathrm{RSS}_T \longrightarrow \sigma_1^2 H,$$

where

$$H = \int_0^1 (W_1(r))^2 \mathrm{d}r - \left(\begin{pmatrix} \int_0^1 W_1(r)\mathrm{d}r \\ \int_0^1 W_2(r)W_1(r)\mathrm{d}r \end{pmatrix}' \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \right)^{-1}.$$

(c) The F test satisfies:

$$T^{-1}F \longrightarrow \frac{1}{G}(\sigma_1 R^* h_2 - d^*)' \times\left(\sigma_1^2 H \begin{pmatrix} 0 & R^* \end{pmatrix} \begin{pmatrix} 1 & \int_0^1 W_2(r)'\mathrm{d}r \\ \int_0^1 W_2(r)\mathrm{d}r & \int_0^1 W_2(r) W_2^*(r)'\mathrm{d}r \end{pmatrix}^{-1} \begin{pmatrix} 0 & R^* \end{pmatrix}'\right)^{-1} \times(\sigma_1 R^* h_2 - d^*),$$

where $R^* = RL_{22}$ and $d^* = d - R\Sigma_{22}^{-1}\Sigma_{21}$.

(a) indicates that OLSE $\hat{\gamma}_T$ is not consistent.

(b) indicates that $s^2 = \dfrac{1}{T-g}\sum_{t=1}^{T}\hat{u}_t^2$ diverges.

(c) indicates that F diverges.

$\Longrightarrow$ **Spurious regression (見せかけの回帰)**

7. **Resolution for Spurious Regression:**

Suppose that $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$ is a spurious regression.

(1) Estimate $y_{1,t} = \alpha + \gamma' y_{2,t} + \phi y_{1,t-1} + \delta y_{2,t-1} + u_t$.

Then, $\hat{\gamma}_T$ is $\sqrt{T}$-consistent, and the t test statistic goes to the standard normal distribution under $H_0: \gamma = 0$.

(2) Estimate $\Delta y_{1,t} = \alpha + \gamma' \Delta y_{2,t} + u_t$. Then, $\hat{\alpha}_T$ and $\hat{\beta}_T$ are $\sqrt{T}$-consistent, and the t test and F test make sense.

(3) Estimate $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$ by the Cochrane-Orcutt method, assuming that u_t is the first-order serially correlated error.

Usually, choose (2).

However, there are two exceptions.

(i) The true value of ϕ in (1) above is not one, i.e., less than one.

(ii) $y_{1,t}$ and $y_{2,t}$ are the cointegrated processes.

In these two cases, taking the first difference leads to the misspecified regression.

8. **Cointegrating Vector:**

Suppose that each element of y_t is $I(1)$ and that $a'y_t$ is $I(0)$.

a is called a **cointegrating vector (共和分ベクトル)**, which is not unique.

Set $z_t = a'y_t$, where z_t is scalar, and a and y_t are $g \times 1$ vectors.

For $z_t \sim I(0)$ (i.e., stationary),

$$T^{-1}\sum_{t=1}^{T} z_t^2 = T^{-1}\sum_{t=1}^{T}(a'y_t)^2 \longrightarrow \mathrm{E}(z_t^2).$$

For $z_t \sim I(1)$ (i.e., nonstationary, i.e., a is not a cointegrating vector),

$$T^{-2}\sum_{t=1}^{T}(a'y_t)^2 \longrightarrow \lambda^2 \int_0^1 (W(r))^2\,\mathrm{d}r,$$

where $W(r)$ denotes a standard Brownian motion and λ^2 indicates variance of $(1 - L)z_t$.

If a is not a cointegrating vector, $T^{-1}\sum_{t=1}^{T} z_t^2$ diverges.

$\Longrightarrow$ We can obtain a consistent estimate of a cointegrating vector by minimizing $\sum_{t=1}^{T} z_t^2$ with respect to a, where a normalization condition on a has to be imposed.

The estimator of the a including the normalization condition is super-consistent (T-consistent).

Stock, J.H. (1987) "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, Vol.55, pp.1035 – 1056.

Proposition:

Let $y_{1,t}$ be a scalar, $y_{2,t}$ be a $k \times 1$ vector, and $(y_{1,t}, y'_{2,t})'$ be a $g \times 1$ vector, where $g = k + 1$.

Consider the following model:

$$y_{1,t} = \alpha + \gamma' y_{2,t} + u_{1,t}$$

$$\Delta y_{2,t} = u_{2,t}$$

$$\begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = \Psi(L)\epsilon_t$$

ϵ_t is a $g \times 1$ i.i.d. vector with $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{E}(\epsilon_t\epsilon_t') = PP'$.

OLSE is given by:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \sum y_{2,t}' \\ \sum y_{2,t} & \sum y_{2,t}y_{2,t}' \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{1,t} \\ \sum y_{1,t}y_{2,t} \end{pmatrix}.$$

Define λ_1, which is a $g \times 1$ vector, and Λ_2, which is a $k \times g$ matrix, as follows:

$$\Psi(1)\,P = \begin{pmatrix} \lambda_1{}' \\ \Lambda_2 \end{pmatrix}.$$

Then, we have the following results:

$$\begin{pmatrix} T^{1/2}(\hat{\alpha}-\alpha) \\ T(\hat{\gamma}-\gamma) \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \left(\Lambda_2 \int W(r)\mathrm{d}r\right)' \\ \Lambda_2 \int W(r)\mathrm{d}r & \Lambda_2 \left(\int (W(r))\,(W(r))'\,\mathrm{d}r\right)\Lambda_2{}' \end{pmatrix}^{-1} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix},$$

where

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} \lambda_1{}'W(1) \\ \Lambda_2 \left(\int W(r)\,(\mathrm{d}W(r))'\right)\lambda_1 + \sum_{\tau=0}^{\infty} \mathrm{E}(u_{2,t}u_{1,t+\tau}) \end{pmatrix}.$$

$W(r)$ denotes a g-dimensional standard Brownian motion.

1) OLSE of the cointegrating vector is consistent even though u_t is serially correlated.

2) The consistency of OLSE implies that $T^{-1}\sum \hat{u}_t^2 \longrightarrow \sigma^2$.

3) Because $T^{-1}\sum(y_{1,t} - \bar{y}_1)^2$ goes to infinity, a coefficient of determination, R^2, goes to one.

15.4 Testing Cointegration

15.4.1 Engle-Granger Test

$y_t \sim I(1)$

$y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$

- $u_t \sim I(0) \implies$ Cointegration
- $u_t \sim I(1) \implies$ Spurious Regression

Estimate $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$ by OLS, and obtain $\hat{u}_t$.

Estimate $\hat{u}_t = \rho \hat{u}_{t-1} + \delta_1 \Delta \hat{u}_{t-1} + \delta_2 \Delta \hat{u}_{t-2} + \cdots + \delta_{p-1} \Delta \hat{u}_{t-p+1} + e_t$ by OLS.

ADF Test:

- $H_0: \rho = 1$ (Sprious Regression)
- $H_1: \rho < 1$ (Cointegration)

$\Longrightarrow$ **Engle-Granger Test**

For example, see Engle and Granger (1987), Phillips and Ouliaris (1990) and Hansen (1992).

Asymmptotic Distribution of Residual-Based ADF Test for Cointegration

# of Refressors,	(a) Regressors have no drift				(b) Some regressors have drift			
excluding constant	1%	2.5%	5%	10%	1%	2.5%	5%	10%
1	−3.96	−3.64	−3.37	−3.07	−3.96	−3.67	−3.41	−3.13
2	−4.31	−4.02	−3.77	−3.45	−4.36	−4.07	−3.80	−3.52
3	−4.73	−4.37	−4.11	−3.83	−4.65	−4.39	−4.16	−3.84
4	−5.07	−4.71	−4.45	−4.16	−5.04	−4.77	−4.49	−4.20
5	−5.28	−4.98	−4.71	−4.43	−5.36	−5.02	−4.74	−4.46

J.D. Hamilton (1994), *Time Series Analysis*, p.766.

The Other Topics

- Generalized Method of Moments (一般化積率法，GMM)
- System of Equations (Seemingly Unrelated Regression (SUR), Simultaneous Equation (連立方程式), and etc.)
- Panel Data (パネル・データ)
- Discrete Dependent Variable, and Limited Dependent Variable
- Bayesian Estimation (ベイズ推定)
- Semiparametric and Nonparametric Regressions and Tests (セミパラメトリック，ノンパラメトリック推定・検定)
- $\cdots$

Exam — Jan. 29, 2014 (AM8:50-10:20), and # 509

- 60 - 70% from two homeworks (2 つの宿題から 60 - 70%)
- 30 - 40% of new questions (30 - 40% の新しい問題)
- Questions are written in English, and answers should be in English or Japanese.

(出題は英語，解答は英語または日本語)

- With no carrying in (持ち込みなし)