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:

$$P(|X-\mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}, \qquad \text{where } \mu = \mathrm{E}(X),\ \sigma^2 = \mathrm{V}(X) \text{ and any } \epsilon > 0.$$

[End of Review]

Replace X, $\mathrm{E}(X)$ and $\mathrm{V}(X)$ 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:

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

Replacing $(1/n)\sum_{i=1}^{n}(x_i - \overline{x})^2$ by its converged value m, we have:

$$\frac{\sqrt{n}(\hat{\beta}_2 - \beta_2)}{\sigma / \sqrt{m}} \longrightarrow N(0, 1),$$

which implies

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

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.

[Review] Confidence Interval (信頼区間，区間推定)):

Suppose $X_1, X_2, \cdots, X_n$ are iid with mean μ and variance σ^2. $\longrightarrow$ No N assumption

From CLT, $\dfrac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0, 1)$.

Replacing σ^2 by $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, we have: $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \longrightarrow N(0, 1)$.

That is, for large n,

$$P\Big(-1.96 < \frac{\overline{X} - \mu}{S/\sqrt{n}} < 1.96\Big) = 0.95, \text{ i.e., } P\Big(\overline{X} - 1.96\frac{S}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{S}{\sqrt{n}}\Big) = 0.95.$$

Note that 1.96 is obtained from the normal distribution table.

Then, replacing the estimators $\overline{X}$ and S^2 by the estimates $\overline{x}$ and s^2, we obtain the 95% confidence interval of μ as follows:

$$(\overline{x} - 1.96\frac{s}{\sqrt{n}},\ \overline{x} + 1.96\frac{s}{\sqrt{n}}).$$

[End of Review]

Going back to OLS, we have:

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

Therefore,

$$P\Big(-2.576 < \frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} < 2.576\Big) = 0.99,$$

i.e.,

$$P\Big(\hat{\beta}_2 - 2.576\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} < \beta_2 < \hat{\beta}_2 + 2.576\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\Big) = 0.99.$$

Note that 2.576 is 0.005 value of $N(0, 1)$, which comes from the statistical table.

Thus, the 99% confidence interval of β_2 is:

$$\Big(\hat{\beta}_2 - 2.576\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}},\ \hat{\beta}_2 + 2.576\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\Big),$$

where $\hat{\beta}_2$ and s^2 should be replaced by the observed data.

[Review] Testing the Hypothesis (仮説検定):

Suppose that $X_1, X_2, \cdots, X_n$ are iid with mean μ and variance σ^2.

From CLT, $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \longrightarrow N(0,1)$, where $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, which is known as the unbiased estimator of σ^2.

- The null hypothesis $H_0: \mu = \mu_0$, where μ_0 is a fixed number.
- The alternative hypothesis $H_1: \mu \neq \mu_0$

Under the null hypothesis, in large sample we have the following disribution:

$$\frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim N(0,1).$$

Replacing $\overline{X}$ and S^2 by $\overline{x}$ and s^2, compare $\dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$ and $N(0,1)$.

H_0 is rejected at significance level 0.05 when $\left|\dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}\right| > 1.96$.

[End of Review]

In the case of OLS, the hypotheses are as follows:

- The null hypothesis $H_0 : \beta_2 = \beta_2^*$
- The alternative hypothesis $H_1 : \beta_2 \neq \beta_2^*$

Under H_0, in large sample,

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

Replacing $\hat{\beta}_2$ and s^2 by the observed data, compare $\dfrac{\hat{\beta}_2 - \beta_2^*}{s / \sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$ and $N(0, 1)$.

H_0 is rejected at significance level 0.05 when $\left|\dfrac{\hat{\beta}_2 - \beta_2^*}{s / \sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\right| > 1.96$.

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]   Content of Special Lectures in Economics (Statistical Analysis)

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]