Econometrics II TA Session #1*

Makoto SHIMOSHIMIZU†

Room 4, October 8, 2019

Contents

---

*All comments welcome!

†E-mail: vgm067sm@student.econ.osaka-u.ac.jp

1 Review of OLS Estimator

In this section, we riview some contents related to the **Ordinary Least Squares (OLS) estimator**. Consider a **multiple regression model**, where a linear relationship between a **dependent** or **explained variable** and multiple **explanatory** or **independent variables** is considered from an n sample. The linear regression model becomes

$$y_i = \beta_1 x_{i,1} + \cdots + \beta_{i,k} x_{i,k} + u_i = X_i \beta + u_i, \tag{1.1}$$

where $X_i = (x_{i,1}, \ldots, x_{i,k})$ is a $1 \times k$ vector for $i \in \{1, \ldots, n\}$ and $\beta = (\beta_1, \ldots, \beta_k)'$ is a $k \times 1$ vector. Denoting by

$$\begin{aligned} y &:= (y_1, \ldots, y_n)' \in \mathbb{R}^n; \\ u &:= (u_1, \ldots, u_n)' \in \mathbb{R}^n; \\ X &:= \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,k} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,k} \end{pmatrix} \in \mathbb{R}^{n \times k}, \end{aligned}$$

we can write the <u>stacked regression system</u> as follows:

$$y = X\beta + u \left(\iff \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\in \mathbb{R}^n} = \underbrace{\begin{pmatrix} x_{1,1} & \cdots & x_{1,k} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,k} \end{pmatrix}}_{\in \mathcal{M}_{n \times k}(\mathbb{R})} \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}}_{\in \mathbb{R}^k} + \underbrace{\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}}_{\in \mathbb{R}^n} \right).$$

1.1 OLS Estimator: Derivation

To derive the OLS estimator, we first consider the following minimization problem.

Definition 1.1 (OLS Estimator for a Multivariate Regression Model)**.** The OLS estimator (for a multivariate regression model) is a vector $\beta_{OLS} \in \mathbb{R}^k$ which satisfies the minimum distance between y and the vectorial space of $\mathbb{R}^n$ generated by X for the Euclidian norm:

$$\beta_{OLS} = \arg\min_{\beta} \|y - X\beta\|_2^2 = \arg\min_{\beta} (y - X\beta)' (y - X\beta).$$

This definition implies that the OLS estimator is an estimator which minimizes the sum of the residual sum of squares. We obtain the OLS estimator as follows.

Theorem 1.1 (OLS Estimator for a Multivariate Regression Model)**.** Suppose

H1: $X_1, \ldots, X_k$ are independent,

then the OLS estimator β_{OLS} exists uniquely and satisfies

$$\beta_{OLS} = (X'X)^{-1} (X'y). \tag{1.2}$$

Proof. To obtain the OLS estimator, we have to confirm the first and second order condition for the minimization problem of the following loss function $S(\beta)$:

$$\arg\min_{\beta} \; \|y - X\beta\|_2^2 =: \arg\min_{\beta} \; S(\beta).$$

The first order condition becomes

$$\begin{aligned}\frac{\partial}{\partial \beta}\|y - X\beta\|_2^2 &= \frac{\partial}{\partial \beta}(y - X\beta)'(y - X\beta) \\ &= -2X'y + 2X'X\beta = \mathbf{0}.\end{aligned}$$

Recall that $y'X\beta \in \mathbb{R}$, and thereby $y'X\beta = (y'X\beta)' = \beta'X'y(\in \mathbb{R})$. Thus, the OLS estimator, denoted as β_{OLS}, satisfies this equation, and hence

$$(X'X)\,\beta_{OLS} = X'y.$$

From the assumption **H1**, the inverse matrix $(X'X)^{-1}$ exists, with $X = (X_1', \ldots, X_k')' \in \mathcal{M}_{n\times k}(\mathbb{R})$, whose columns are independent so that $X'X$ is a full rank matrix, and therefore we can obtain the OLS estimator in the form of Eq. (1.2). The second order condition becomes

$$\frac{\partial}{\partial \beta \partial \beta'}\|y - X\beta\|_2^2 = 2X'X.$$

By assumption **H1**, $X'X$ is a positive definite matrix. This shows that the loss function $S(\beta)$ has a minimum at the OLS estimator β_{OLS}. □

From this theorem, we can confirm that the OLS estimator expressed as Eq. (1.2) is a random variable since we can rewrite it as follows:

$$\beta_{OLS} = \beta + (X'X)^{-1}X'u. \tag{1.3}$$

Therefore, we can consider the mean and variance of the OLS estimator. First, we see the mean of the OLS estimator, which will be used to prove that the OLS estimator is an unbiased estimator.

Proposition 1.1 (Mean of the OLS Estimator)**.** Suppose

H2: $\mathbb{E}[u_i | X] = 0$ for all $i \in \{1, \ldots, n\}$,

then the conditional expectation of the OLS estimator β_{OLS} becomes

$$\mathbb{E}[\beta_{OLS}|X] = \beta. \tag{1.4}$$

Proof. Calculating the expectation of β yields

$$\begin{aligned}\mathbb{E}[\beta_{OLS}|X] &= \mathbb{E}\left[(X'X)^{-1}X'(X\beta + u)\,\middle|\,X\right] \\ &= \beta + (X'X)^{-1}X' \underbrace{\mathbb{E}\left[u\middle|X\right]}_{=0(\text{from } \mathbf{H2})} = \beta,\end{aligned}$$

which proves Eq. (1.4). □

Corollary 1.1 (Unconditional Expectation of the OLS estimator)**.** The conditional expectation of the OLS estimator is same as the unconditional one:

$$\mathbb{E}[\beta_{OLS}] = \beta.$$

from the **law of iterated expectation** mentioned below.

Lemma 1.1 (Law of Iterated Expectation)**.** For any two random variables x and y,

$$\mathbb{E}[y] = \mathbb{E}_x \left[\mathbb{E}[y|x]\right], \tag{1.5}$$

where $\mathbb{E}_x$ is the expectation over the values of x.

The proof is omitted (left as an exercise for students). From this, we have

$$\mathbb{E}[\beta_{OLS}] = \mathbb{E}\big[\underbrace{\mathbb{E}[\beta_{OLS}|X]}_{=\beta}\big] = \mathbb{E}[\beta] = \beta.$$

The variance of the OLS estimator, which is the minimum variance in the class of linear OLS estimator, becomes as follows.

Proposition 1.2 (Variance of the OLS Estimator)**.** Suppose [**H1**–**H2**] holds and assume

H3 $\mathbb{V}[u_i|X] = \sigma^2$ for all $i \in \{1, \dots, n\}$;

H4 $\mathbb{E}[u_i u_j|X] = 0$ for all $i \neq j$ and $i, j \in \{1, \dots, n\}$,

the conditional variance of the OLS estimator β_{OLS} becomes

$$\mathbb{V}[\beta_{OLS}|X] = \sigma^2 \left(X'X\right)^{-1}, \tag{1.6}$$

and the unconditional variance becomes

$$\mathbb{V}[\beta_{OLS}] = \sigma^2 \mathbb{E}\left[\left(X'X\right)^{-1}\right]. \tag{1.7}$$

Proof. From the Eq, (1.3) and Eq, (1.4),

$$\beta_{OLS} - \mathbb{E}[\beta_{OLS}\big|X] = \beta_{OLS} - \beta = \left(X'X\right)^{-1} X'u.$$

Therefore,

$$\begin{aligned}
\mathbb{V}[\beta_{OLS}\big|X] &= \mathbb{E}\left[\left(\beta_{OLS} - \mathbb{E}[\beta_{OLS}\big|X]\right)\left(\beta_{OLS} - \mathbb{E}[\beta_{OLS}\big|X]\right)'\Big|X\right] \\
&= \mathbb{E}\left[\left(X'X\right)^{-1} X'uu'X \left(X'X\right)^{-1}\big|X\right] \\
&= \left(X'X\right)^{-1} X'\mathbb{E}\left[uu'\big|X\right] X \left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1} X'\sigma^2 I_n X \left(X'X\right)^{-1} \\
&= \sigma^2 \left(X'X\right)^{-1}.
\end{aligned}$$

This implies Eq. (1.6) holds. Thus,

$$\mathbb{V}[\beta_{OLS}] = \mathbb{E}\left[\mathbb{V}[\beta_{OLS}\big|X]\right] + \mathbb{V}\left[\mathbb{E}[\beta_{OLS}\big|X]\right] = \mathbb{E}\left[\sigma^2\left(X'X\right)^{-1}\right] + \mathbb{V}[\beta] = \sigma^2\mathbb{E}\left[\left(X'X\right)^{-1}\right],$$

which proves Eq. (1.7). See [2] for the proof of the first equality. □

1.2 OLS Estimator: Properties

Here we exhibit some properties of the OLS estimator.

Theorem 1.2 (Properties of the OLS Estimator)**.** The OLS estimator obtained above has the following properties.

(i) **unbiasedness** Under the assumption **H2**, the OLS estimator β_{OLS} becomes an unbiased estimator:

$$\mathbb{E}[\beta_{OLS}] = \beta. \tag{1.8}$$

(ii) **consistency** Under the following assumption:

H5 $X'X$ is positive definite;

H6 For all i, for all k, l, the moments of $\mathbb{E}\left[|x_{ik}x_{il}|\right]$ exist and $\mathbb{E}\left[X'X\right]$ is p.d.,

as well as [**H1**–**H4**], the OLS estimator $\beta_{OLS} = \left(X'X\right)^{-1}\left(X'y\right)$ satisfies

$$\beta_{OLS} \xrightarrow{p} \beta \quad \text{or} \quad \operatorname*{plim}_{n\to\infty} \beta_{OLS} = \beta. \tag{1.9}$$

(iii) **efficiency** Under the assumption [**H1**–**H4**], the variance of the OLS estimator is the minimum one in the class of linear unbiased estimator.

Proof. We can prove the above theorem directly as follows.

(i) **unbiasedness** This property is shown above (in Corollary 1.1).

(ii) **consistency** From Eq. (1.3), we have

$$\beta_{OLS} = \beta + \left(X'X\right)^{-1} X'u = \beta + \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} X_i'u_i\right).$$

By taking the probability limit on both sides, we have

$$\begin{aligned}\operatorname*{plim}_{n\to\infty} \beta_{OLS} &= \operatorname*{plim}_{n\to\infty} \left[\beta + \left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'u\right)\right] \\ &= \beta + \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1} \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'u_i\right). \end{aligned} \tag{1.10}$$

Here we apply the **convergence of the product of random variables in probability**, which we will discuss in the following. From the **weak law of large numbers** (WLLN),

$$\frac{1}{n}\sum_{i=1}^{n} X_i'X_i \xrightarrow{p} \mathbb{E}\left[X_i'X_i\right] < \infty; \tag{1.11}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i'u_i \xrightarrow{p} \mathbb{E}\left[X_i'u_i\right] = \mathbf{0}(\in \mathbb{R}^k). \tag{1.12}$$

$\mathbb{E}[X_i'u_i] = 0$ holds from the **orthogonal condition** with respect to X and u. In addition,

$$\operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1} = \mathbb{E}[X_i'X_i]^{-1} \tag{1.13}$$

holds from the **continuous mapping theorem**. Thus, substituting Eq. (1.11) and Eq. (1.12) into Eq. (1.10) results in

$$\operatorname*{plim}_{n\to\infty}\beta_{OLS} = \beta + \mathbb{E}[X_i'X_i]^{-1}\mathbf{0} = \beta,$$

which indicates that $\beta_{OLS} \xrightarrow{p} \beta$.

(iii) **efficiency** As for the efficiency of the OLS estimator, the **Gauss–Markov theorem** for a multiple regression model , explained in the appendix A, support the efficiency.

□

1.3 OLS Estimator: Asymptotic Normality

In this section, we derive the asymptotic distribution of an OLS estimator to observe how the distribution changes as $n \to \infty$.

Theorem 1.3 (Asymptotic Normality of an OLS Estimator)**.** Let β_{OLS} be the OLS estimator obtained under the assumption [**H1–H6**]. Then, the OLS estimator asymptotically follows a normal distribution as follows:

$$\sqrt{n}(\beta_{OLS} - \beta) \xrightarrow{d} N_{\mathbb{R}^k}\left(\mathbf{0}, \sigma^2\left(\mathbb{E}[X_i'X_i]\right)^{-1}\right).$$

Proof. From Eq. (1.3), we have

$$\begin{aligned}\beta_{OLS} &= \beta + (X'X)^{-1}X'u \\ &= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'u_i\right).\end{aligned}$$

Therefore,

$$\sqrt{n}(\beta_{OLS} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i'u_i\right). \tag{1.14}$$

From the **Lindeberg–Feller central limit theorem** (Lindeberg–Feller CLT) as well as the **weak law of large numbers** (WLLN) and **continuous mapping theorem**, we have

$$\begin{aligned}&\left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1} \xrightarrow{\mathbb{P}} \mathbb{E}[X_i'X_i]^{-1}; \\ &\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i'u_i\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'u_i - \mathbf{0}\right) \xrightarrow{d} N_{\mathbb{R}^k}\left(\mathbf{0}, \mathbb{V}[X_i'u_i]\right),\end{aligned}$$

since from the orthogonal condition,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i' u_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i' u_i] = \mathbf{0}.$$

Then,

$$\begin{aligned}
\mathbb{V}[X_i' u_i] &= \mathbb{E}\left[\mathbb{V}[X_i' u_i \big| X_i]\right] + \mathbb{V}\big[\underbrace{\mathbb{E}[X_i' u_i \big| X_i]}_{=0}\big] \\
&= \mathbb{E}\left[X_i' \mathbb{V}[u_i \big| X_i] X_i\right] \\
&= \mathbb{E}\left[X_i' \sigma^2 X_i\right] \\
&= \sigma^2 \mathbb{E}\left[X_i' X_i\right] < \infty,
\end{aligned}$$

Therefore, from Eq. (1.14) and the **Slutsky's theorem**,

$$\sqrt{n}(\beta_{OLS} - \beta) \xrightarrow{d} \mathbb{E}\left[X_i' X_i\right]^{-1} \mathbf{B},$$

where

$$\mathbf{B} \sim N_{\mathbb{R}^k}\left(\mathbf{0}, \sigma^2 \mathbb{E}\left[X_i' X_i\right]\right).$$

From the following relation:

$$\mathbf{B} \sim N_{\mathbb{R}^k}\left(\mathbf{0}, \sigma^2 \mathbb{E}\left[X_i' X_i\right]\right) \Longrightarrow \mathbb{E}\left[X_i' X_i\right]^{-1} \mathbf{B} \sim N_{\mathbb{R}^k}\left(\mathbf{0}, \sigma^2 \mathbb{E}\left[X_i' X_i\right]^{-1}\right),$$

we obtain

$$\sqrt{n}(\beta_{OLS} - \beta) \xrightarrow{d} N_{\mathbb{R}^k}\left(\mathbf{0}, \sigma^2 \mathbb{E}\left[X_i' X_i\right]^{-1}\right).$$

□

2 Review of GLS Estimator

In this section, we riview some contents related to the **Generalized Least Squares (GLS) estimator**. Consider a **linear heteroscedastic model**, where one considers a linear relationship between a **dependent** or **explained variable** and multiple **explanatory** or **independent variables** from an n sample under the assumption of heteroscedastic error terms:

$$y = X\beta + u. \tag{2.1}$$

The definition is given as follows.

Definition 2.1 (Linear Regression with Heteroscedastic Errors)**.** We call **a linear heteroscedastic model** a model where the random vector y linearly depends on k explanatory variables X as Eq. (2.1) with the assumptions:

GH1: $\mathbb{E}[u \big| X] = \mathbf{0}$;

GH2: $\mathbb{V}[u \big| X] = \mathbb{E}[u'u \big| X] := \Omega = \Sigma(X, \theta)$ is **positive definite**;

GH3: $X'\Omega^{-1}X$ is positive definite.

2.1 GLS Estimator: Derivation

To derive the GLS estimator, we first consider the following minimization problem.

Definition 2.2 (Generalized Least Squares (GLS) Estimator)**.** The GLS estimator is a vector $\beta_{GLS} \in \mathbb{R}^k$ which satisfies the following minimization problem:

$$\beta_{GLS} = \arg\min_{\beta} \ \|y - X\beta\|^2_{\Omega^{-1}} := \arg\min_{\beta} \ (y - X\beta)' \, \Omega^{-1} \, (y - X\beta) \, .$$

The GLS estimator obtained from the above definition becomes as follows.

Theorem 2.1 (Generalized Least Squares (GLS) Estimator)**.** Suppose [**H1**–**H3**] holds. Then the GLS estimator β exists, is unique and satisfies

$$\beta_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y. \tag{2.2}$$

Proof. Define $\Omega^{-1/2}$ such that $\Omega^{-1/2}\Omega\Omega^{-1/2} = I_n$. Then, multiplying both sides of Eq. (2.1) by $\Omega^{-1/2}$ from the left results in

$$\Omega^{-1/2}y = \Omega^{-1/2}X\beta + \Omega^{-1/2}u.$$

By denoting $y^* := \Omega^{-1/2}y$, $X^* := \Omega^{-1/2}X$ and $u^* := \Omega^{-1/2}u$, we have

$$y^* = X^*\beta + u^*, \tag{2.3}$$

where $u^* \sim N_{\mathbb{R}^{n\times n}}\left(\mathbf{0}, I_n\right)$. Note that $u \sim N_{\mathbb{R}^{n\times n}}\left(\mathbf{0}, \Omega\right) \implies \Omega^{-1/2}u \sim N_{\mathbb{R}^{n\times n}}\left(\mathbf{0}, I_n\right)$. The model assumptions can be reported under this transformation:

GH1': $\mathbb{E}[u^* \big| X^*] = \mathbf{0}$;

GH2': $\mathbb{V}[u^* \big| X^*] := \Omega^{-1/2}\mathbb{V}[u \big| X]\Omega^{-1/2'} = \Omega^{-1/2}\Omega\Omega^{-1/2'} = I_n$;

GH3': $X^{*'}X^*$ is positive definite.

Thus, the GLS estimator is the OLS estimator of the regression coefficients of y^* on X^*:

$$\left(X^{*'}X^*\right)^{-1}\left(X^{*'}y^*\right) = \left(X\Omega^{-1/2'}\Omega^{-1/2}X'\right)^{-1}\left(X\Omega^{-1/2'}\Omega^{-1/2}y\right) = \left(X\Omega^{-1}X'\right)^{-1} X\Omega^{-1}y = \beta_{GLS}, \tag{2.4}$$

which proves Eq. (2.2). □

To obtain the GLS estimator, we have another method, as in the derivation of the OLS estimator, by verifying the first and second order conditions of the following minimization problem:

$$\arg\min_{\beta} \ \|y - X\beta\|^2_{\Omega^{-1}} =: \arg\min_{\beta} \ S^*(\beta).$$

The first order condition becomes

$$\begin{aligned} \frac{\partial}{\partial\beta}\|y - X\beta\|^2_{\Omega^{-1}} &= \frac{\partial}{\partial\beta}\left(y - X\beta\right)' \Omega^{-1}\left(y - X\beta\right) \\ &= -2X'\Omega^{-1}y + 2X'\Omega^{-1}X\beta = \mathbf{0}. \end{aligned}$$

Recall that $y'\Omega^{-1}X\beta \in \mathbb{R}$ and thereby $y'\Omega^{-1}X\beta = \beta'X'\Omega^{-1}y (\in \mathbb{R})$. The GLS estimator for this model, denoted by β_{GLS}, satisfies this equation, and hence

$$\left(X'\Omega^{-1}X\right)\beta_{GLS} = X'\Omega^{-1}y.$$

From assumption **GH3**, the inverse matrix $\left(X'\Omega^{-1}X\right)^{-1}$ exists, with $X = (X_1, \ldots, X_k) \in \mathcal{M}_{n\times k}(\mathbb{R})$, and therefore we can obtain the GLS estimator in the form of Eq. (2.2). The second order condition becomes

$$\frac{\partial}{\partial\beta\partial\beta'}\|y - X\beta\|^2_{\Omega^{-1}} = 2X'\Omega^{-1}X.$$

By assumption **GH3**, the loss function $S^*(\beta)$ has a minimum, which is the GLS estimator β_{GLS}.

From the Theorem 2.1, we can confirm that the GLS estimator expressed as Eq. (2.2) is a random variable since we can rewrite it as follows:

$$\beta_{GLS} = \beta + \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}u. \tag{2.5}$$

Therefore, we can consider the mean and variance of the GLS estimator. First, we see the mean of the GLS estimator, which will be used to prove that the GLS estimator is an unbiased estimator.

Proposition 2.1 (Mean of the GLS Estimator)**.** Under the assumption [**GH1–GH3**], the conditional expectation of the GLS estimator β_{GLS} becomes

$$\mathbb{E}[\beta_{GLS}|X] = \beta. \tag{2.6}$$

Proof. Calculating the expectation of β_{GLS} yields

$$\begin{aligned}\mathbb{E}[\beta_{GLS}|X] &= \mathbb{E}\left[\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}(X\beta + u)\middle|X\right] \\ &= \beta + \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}\mathbb{E}\left[u\middle|X\right] \\ &= \beta,\end{aligned}$$

which proves Eq. (2.6). □

Corollary 2.1 (Unconditional Expectation of the GLS estimator)**.** The conditional expectation of the GLS estimator is same as the unconditional one:

$$\mathbb{E}[\beta_{GLS}] = \beta.$$

from the **law of iterated expectation**:

$$\mathbb{E}[\beta_{GLS}] = \mathbb{E}\big[\mathbb{E}[\beta_{GLS}|X]\big] = \mathbb{E}[\beta] = \beta.$$

The variance of the GLS estimator, which is the minimum variance in the class of linear GLS estimator, becomes as follows.

Proposition 2.2 (Variance of the GLS Estimator)**.** Under the assumption [**GH1–GH3**], the conditional variance of the GLS estimator β_{GLS} becomes

$$\mathbb{V}[\beta_{GLS}|X] = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1}, \tag{2.7}$$

and the unconditional variance becomes

$$\mathbb{V}[\beta_{GLS}] = \mathbb{E}\left[\left(X'\Omega^{-1}X\right)^{-1}\right]. \tag{2.8}$$

Proof. From the Eq, (2.5) and Eq, (2.6),

$$\beta_{GLS} - \mathbb{E}[\beta_{GLS}\big|X] = \beta_{GLS} - \beta = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}u.$$

Therefore,

$$\mathbb{V}[\beta_{GLS}\big|X] = \mathbb{E}\left[\left(\beta_{OLS} - \mathbb{E}[\beta_{GLS}\big|X]\right)\left(\beta_{OLS} - \mathbb{E}[\beta_{GLS}\big|X]\right)'\Big|X\right] = \left(X'\Omega^{-1}X\right)^{-1}.$$

This implies Eq. (2.7) holds. Thus,

$$\mathbb{V}[\beta_{GLS}] = \mathbb{E}\left[\mathbb{V}[\beta_{GLS}\big|X]\right] + \mathbb{V}\left[\mathbb{E}[\beta_{GLS}\big|X]\right] = \mathbb{E}\left[\left(X'\Omega^{-1}X\right)^{-1}\right] + \underbrace{\mathbb{V}[\beta]}_{=0} = \mathbb{E}\left[\left(X'\Omega^{-1}X\right)^{-1}\right],$$

which proves Eq. (2.8). □

2.2 GLS Estimator: Properties

Here we exhibit some properties of the GLS estimator.

Theorem 2.2 (Properties of the GLS Estimator)**.** Under the assumption [**GH1–GH3**], the GLS estimator obtained above has the following properties:

(i) **unbiasedness** The GLS estimator β_{GLS} becomes an unbiased estimator:

$$\mathbb{E}[\beta_{GLS}] = \beta. \tag{2.9}$$

(ii) **consistency** Under the additional assumption:

GH4: A $= \mathbb{E}[X_i'\Omega^{-1}X_i]$ is non singular;

as well as [**GH1–GH3**], the GLS estimator β_{GLS} satisfies

$$\beta_{GLS} \xrightarrow{p} \beta \quad \text{or} \quad \operatorname*{plim}_{n\to\infty} \beta_{GLS} = \beta. \tag{2.10}$$

(iii) **efficiency** The variance of the GLS estimator is the minimum one in the class of linear unbiased estimator.

Proof. We can derive these properties via a similar calculation in the derivation of the OLS estimator.

(i) **unbiasedness** This property is shown above (in Remark 2.1).

(ii) **consistency** By taking the probability limit on both sides of Eq. (2.4), we have

$$
\begin{aligned}
\operatorname*{plim}_{n\to\infty} \beta_{GLS} &= \operatorname*{plim}_{n\to\infty} \left[\beta + \left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}u\right)\right] \\
&= \beta + \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^{n} X_i^{*\prime} X_i^{*}\right)^{-1} \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^{n} X_i^{*\prime} u_i^{*}\right) \\
&= \beta + \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1} \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}u_i\right). \qquad (2.11)
\end{aligned}
$$

Here we apply the **convergence of the product of random variables in probability**. From the **weak law of large numbers** (WLLN),

$$
\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}X_i \xrightarrow{p} \mathbb{E}\left[X_i'\Omega^{-1}X_i\right] < \infty; \qquad (2.12)
$$

$$
\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}u_i \xrightarrow{p} \mathbb{E}\left[X_i'\Omega^{-1}u_i\right] = \mathbf{0}(\in \mathbb{R}^k). \qquad (2.13)
$$

We can prove $\mathbb{E}[X_i'\Omega^{-1}u_i] = \mathbf{0}$ by using the **vec operator** with the **orthogonal condition** with respect to X and u:

$$
\text{vec } \mathbb{E}\left[X_i'\Omega^{-1}u_i\right] = \mathbb{E}\left[\text{vec }\left(X_i'\Omega^{-1}u_i\right)\right] = \mathbb{E}\left[\left(u_i \otimes X_i'\right)\text{vec } \Omega^{-1}\right] = \mathbf{0}.
$$

In addition,

$$
\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1} = \mathbb{E}\left[X_i'\Omega^{-1}X_i\right]^{-1} \qquad (2.14)
$$

holds from the **continuous mapping theorem**. Thus, substituting Eq. (2.12) and (2.13) into Eq. (2.11) results in

$$
\operatorname*{plim}_{n\to\infty} \beta_{GLS} = \beta + \mathbb{E}\left[X_i'\Omega^{-1}X_i\right]^{-1}\mathbf{0} = \beta,
$$

which proves that $\beta_{GLS} \xrightarrow{p} \beta$.

(iii) **efficiency** As for the efficiency of the GLS estimator, the **Gauss–Markov theorem** supports the efficiency. Note that β_{GLS} is efficient than β_{OLS}(, which is shown in the Appendix B).

□

2.3 GLS Estimator: Asymptotic Normality

In this section, we derive the asymptotic distribution of the GLS estimator to observe how the distribution changes as $n \to \infty$.

Theorem 2.3 (Asymptotic Normality of an GLS Estimator)**.** Let β_{GLS} be the GLS estimator obtained under the assumption [**GH1–GH3**]. Suppose

GH5: $\mathbf{B} = \mathbb{E}[X_i'\Omega^{-1}u_iu_i\Omega^{-1}X_i]$ exists;

as well as [**GH1–GH4**]. Then, the GLS estimator asymptotically follows a normal distribution as follows:

$$\sqrt{n}(\beta_{GLS} - \beta) \xrightarrow{d} N_{\mathbb{R}^k}\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right).$$

Proof. From Eq. (2.5), we have

$$\begin{aligned}\beta_{GLS} &= \beta + \left(X'\Omega^{-1}X\right)^{-1}X'u \\ &= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}u_i\right).\end{aligned}$$

Therefore, rewriting results in

$$\sqrt{n}(\beta_{GLS} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i'\Omega^{-1}u_i\right). \tag{2.15}$$

From the **Lindeberg–Feller central limit theorem** (Lindeberg–Feller CLT) as well as the **weak law of large numbers** (WLLN) and **continuous mapping theorem**, we have

$$\begin{aligned}&\left(\frac{1}{n}X_i'\Omega^{-1}X_i\right)^{-1} \xrightarrow{\mathbb{P}} \mathbb{E}\left[X_i'\Omega^{-1}X_i\right]^{-1} =: \mathbf{A}; \\ &\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i'\Omega^{-1}u_i\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}u_i - \mathbf{0}\right) \xrightarrow{d} N_{\mathbb{R}^k}\left(\mathbf{0}, \mathbb{V}[X_i'\Omega^{-1}u_i]\right),\end{aligned}$$

since from the orthogonal condition,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}u_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i'\Omega^{-1}u_i] = \mathbf{0}.$$

Then,

$$\mathbb{V}[X_i'\Omega^{-1}u_i] = \mathbb{E}\left[X_i'\Omega^{-1}u_iu_i\Omega^{-1}X_i\right] = \mathbf{B} < \infty.$$

Therefore, from Eq. (2.15) and the **Slutsky's theorem**,

$$\sqrt{n}(\beta_{GLS} - \beta) \xrightarrow{d} \mathbf{A}^{-1}\mathbf{Z},$$

where

$$\mathbf{Z} \sim N_{\mathbb{R}^k}(\mathbf{0}, \mathbf{B}).$$

From the following relation:

$$\mathbf{Z} \sim N_{\mathbb{R}^k}(\mathbf{0}, \mathbf{B}) \Longrightarrow \mathbf{A}^{-1}\mathbf{B} \sim N_{\mathbb{R}^k}\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right),$$

we obtain

$$\sqrt{n}(\beta_{GLS} - \beta) \xrightarrow{d} N_{\mathbb{R}^k}(\mathbf{0}, \mathbf{B}) \Longrightarrow \mathbf{A}^{-1}\mathbf{B} \sim N_{\mathbb{R}^k}\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right).$$

□

Appendix

A Gauss–Markov Theorem for a Multiple Regression Model

Here we will obtain a general result for the class of linear unbiased estimators of β. It can be conducted via a direct method.

Theorem A.1 (Gauss–Markov Theorem for a Multiple Regression Model)**.** Under the assumption [**H1**–**H4**], the OLS estimater β_{OLS} of the multiple regression model

$$y_i = X_i\beta + u_i, \tag{A.1}$$

for all $i \in \{1, \ldots, n\}$ is of minimum variance among the class of linear unbiased estimator.

Proof. Let us assume another unbiased linear estimator of β, say $\tilde{\beta}$. Thus, there exists a matrix $A \in \mathbb{R}^{k\times n}$ such that $\tilde{\beta} = Ay$. Since $\tilde{\beta}$ is an unbiased estimator,

$$\mathbb{E}[\tilde{\beta}] = \beta \tag{A.2}$$

holds, which yields

$$\mathbb{E}\left[A\left\{X\beta + u\right\}\right] = \beta \iff AX\beta = \beta. \tag{A.3}$$

Therefore, $AX = I_k$ must be satisfied. Moreover, from the equation:

$$\tilde{\beta} - \mathbb{E}[\tilde{\beta}] = A\left\{y - X\beta\right\} = Au. \tag{A.4}$$

The variance $\mathbb{V}[\tilde{\beta}]$ becomes

$$\mathbb{V}[\tilde{\beta}] = \mathbb{V}\left[Au\right] = A\mathbb{V}\left[u\right]A' = A\left(\sigma^2 I_n\right)A' = \sigma^2 AA', \tag{A.5}$$

from the assumption $\mathbb{V}\left[u\right] = \sigma^2 I_n$. Using the **projection matrix**:

$$\mathcal{M}_X := I_n - X\left(X'X\right)^{-1}X' \left(\iff I_n = \mathcal{M}_X + X\left(X'X\right)^{-1}X'\right), \tag{A.6}$$

we can rewrite Eq. (A.5) as follows:

$$\begin{aligned}
\mathbb{V}[\tilde{\beta}] &= A\left(\sigma^2 I_n\right)A' \\
&= \sigma^2 A\left(X\left(X'X\right)^{-1}X' + \mathcal{M}_X\right)A' \\
&= \sigma^2\left(AX\left(X'X\right)^{-1}X'A' + A\mathcal{M}_X A'\right).
\end{aligned}$$

Substituting $AX = I_k$ and $\mathbb{V}[\beta_{OLS}] = \sigma^2\left(X'X\right)^{-1}$ into the above equation results in

$$\mathbb{V}[\tilde{\beta}] = \mathbb{V}[\beta_{OLS}] + \sigma^2 A\mathcal{M}_X A' \iff \mathbb{V}[\tilde{\beta}] - \mathbb{V}[\beta_{OLS}] = \sigma^2 A\mathcal{M}_X A'.$$

Hence, the difference of ith diagonal elements of variance–covariance matrices becomes

$$\mathbb{V}[\tilde{\beta}]_{ii} - \mathbb{V}[\beta_{OLS}]_{ii} = a_i'\mathcal{M}a_i \geq 0$$

for any column vector a_i in A for $i \in \{1, \ldots, k\}$, which proves the theorem. □

B Comparison of the OLS and GLS Estimator

We compare the efficiency between OLS and GLS estimators. Using the results of Section 2, under the assumption [**GH1–GH3**], we have

$$\begin{aligned}\mathbb{V}[\beta_{OLS}|\underline{x}] &= \left(\underline{x}'\underline{x}\right)^{-1}\underline{x}'\Omega\underline{x}\left(\underline{x}'\underline{x}\right)^{-1};\\ \mathbb{V}[\beta_{GLS}|\underline{x}] &= \left(\underline{x}'\Omega^{-1}\underline{x}\right)^{-1}.\end{aligned}$$

Then, subtracting $\mathbb{V}[\beta_{GLS}|\underline{x}]$ from $\mathbb{V}[\beta_{OLS}|\underline{x}]$ results in

$$\begin{aligned}\mathbb{V}[\beta_{OLS}|\underline{x}] - \mathbb{V}[\beta_{GLS}|\underline{x}] &= \left(\underline{x}'\underline{x}\right)^{-1}\underline{x}'\Omega\underline{x}\left(\underline{x}'\underline{x}\right)^{-1} - \left(\underline{x}'\Omega^{-1}\underline{x}\right)^{-1}\\ &= \left(\underline{x}'\underline{x}\right)^{-1}\underline{x}'\Omega\underline{x}\left(\underline{x}'\underline{x}\right)^{-1} - \left(\underline{x}'\Omega^{-1}\underline{x}\right)^{-1}\underline{x}'\Omega^{-1}\Omega\Omega^{-1}\underline{x}\left(\underline{x}'\Omega^{-1}\underline{x}\right)^{-1}\\ &= \left\{\left(\underline{x}'\underline{x}\right)^{-1}\underline{x}' - \left(\underline{x}'\Omega^{-1}\underline{x}\right)^{-1}\underline{x}'\Omega^{-1}\right\}\Omega\left\{\underline{x}\left(\underline{x}'\underline{x}\right)^{-1} - \Omega^{-1}\underline{x}\left(\underline{x}'\Omega^{-1}\underline{x}\right)^{-1}\right\}\\ &= \left\{\left(\underline{x}'\underline{x}\right)^{-1}\underline{x}' - \left(\underline{x}'\Omega^{-1}\underline{x}\right)^{-1}\underline{x}'\Omega^{-1}\right\}\Omega\left\{\left(\underline{x}'\underline{x}\right)^{-1}\underline{x}' - \left(\underline{x}'\Omega^{-1}\underline{x}\right)^{-1}\underline{x}'\Omega^{-1}\right\}'\\ &=: A\Omega A',\end{aligned}$$

where Ω is positive definite. Therefore, if $\Omega \neq I_n$, then $A\Omega A'$ also becomes positive definite. As a consequence,

$$\mathbb{V}[\beta_{OLS}|\underline{x}]_{ii} - \mathbb{V}[\beta_{GLS}|\underline{x}]_{ii} > 0,$$

for all $i \in \{1, \ldots, n\}$. These results infer that β_{GLS} is more efficient than β_{OLS}.

References

[1] Greene, W. H., *Econometric analysis Seventh Edition*. 2012, Pearson.

[2] Wooldridge, J. M., *Econometric analysis of cross section and panel data*. 2010, MIT Press.