Econometrics I
TA Session 8

Jukina HATAKEYAMA *

June 4, 2024

Contents

*E-mail: u868710a@ecs.osaka-u.ac.jp

1 Matrix Transformation

First, we review a matrix decomposition, particularly in the case of a symmetric positive definite matrix, which plays a powerful role in obtaining the "Generalized Least Squares estimator", discussed in a further section. Second, we introduce the *vec* operator and its useful property related to the Kronecker product of matrices.

1.1 Diagonal Reduction of a Symmetric Positive Definite Matrix

Theorem 1.1 (Decomposition of a Symmetric Positive Definite Matrix)**.** Let $W = (w_1, \ldots, w_n) \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix. Then, there exists a matrix $W^{-1/2} \in \mathbb{R}^{n\times n}$ such that

$$W^{-1/2} W W^{-1/2'} = I_n,$$

where $I_n \in \mathbb{R}^{n\times n}$ is an identity matrix. This matrix also satisfies

$$W^{-1/2'} W^{-1/2} = W^{-1}.$$

1.2 Decomposition of a Variance–Covariance Matrix

From the definition of the variance–covariance matrix,

$$\Omega := \mathbb{V}[u \big| X] = \begin{bmatrix} \mathbb{V}[u_1 \big| X] & \mathrm{Cov}[u_1, u_2 \big| X] & \cdots & \mathrm{Cov}[u_1, u_n \big| X] \\ \mathrm{Cov}[u_2, u_1 \big| X] & \mathbb{V}[u_2 \big| X] & \cdots & \mathrm{Cov}[u_2, u_n \big| X] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[u_n, u_1 \big| X] & \mathrm{Cov}[u_n, u_2 \big| X] & \cdots & \mathbb{V}[u_n \big| X] \end{bmatrix},$$

the variance–covariance matrix for $u = (u_1, \ldots, u_n)$ is a symmetric matrix. Thus, if we assume the variance–covariance matrix is positive definite, then we can apply the above theorem: there exists a matrix $\Omega^{-1/2}$ such that

$$\Omega^{-1/2} \Omega \Omega^{-1/2'} = I_n \quad \text{and} \quad \Omega^{-1/2'} \Omega^{-1/2} = \Omega^{-1}.$$

1.3 Vec Operator

The definition of *vec operator* is given as follows.

Definition 1.1 (vec operator)**.** The vec operator creates a column vector from a matrix $A \in \mathbb{R}^{m\times n}$ by stacking the the column vectors of $A = (a_1, \ldots, a_n)$ below one another:

$$\mathrm{vec}\ (A) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^{mn\times 1}$$

This dedfinition gives rise to the following useful property.

Theorem 1.2. For matrices $A \in \mathbb{R}^{i\times l}$, $X \in \mathbb{R}^{l\times m}$, and $B \in \mathbb{R}^{m\times n}$,

$$\text{vec }(AXB) = (B' \otimes A)\,\text{vec }(X). \tag{1}$$

Proof. We denote the matrices A, X, and B as follows:

$$A = (a_1, \dots, a_n) \in \mathbb{R}^{i\times l}, \quad X = (x_1, \dots, x_l) \in \mathbb{R}^{l\times m}, \quad B = (b_1, \dots, b_k) \in \mathbb{R}^{m\times n}.$$

Then, the j^{th} column of the matrix $AXB \in \mathbb{R}^{m\times k}$, denoted as $(AXB)_k$, becomes

$$\begin{aligned}
(AXB)_k &= AXb_k \\
&= A\sum_{i=1}^{m} x_i b_{i,k} \\
&= \begin{pmatrix} b_{1,k}A & b_{2,k}A & \dots & b_{m,k}A \end{pmatrix} \text{vec }(X) \\
&= \left(\begin{pmatrix} b_{1,k} & b_{2,k} & \dots & b_{m,k} \end{pmatrix} \otimes A \right) \text{vec }(X) \\
&= \left(b_k' \otimes A \right) \text{vec }(X)
\end{aligned}$$

Stacking the columns together yields

$$\text{vec }(AXB) = \begin{pmatrix} (AXB)_1 \\ (AXB)_2 \\ \dots \\ (AXB)_n \end{pmatrix} = \begin{pmatrix} (b_1' \otimes A) \\ (b_2' \otimes A) \\ \dots \\ (b_k' \otimes A) \end{pmatrix} \text{vec }(X) = (B' \otimes A)\,\text{vec }(X),$$

which proves (1). $\square$

Ex.)

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} c & d \\ d & c \end{pmatrix} \begin{pmatrix} e & f \\ f & e \end{pmatrix} = \begin{pmatrix} ace + bde + adf + bcf & acf + bdf + ade + bce \\ cbe + ade + bdf + acf & cbf + adf + bde + ace \end{pmatrix}$$

$$\text{vec}\left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} c & d \\ d & c \end{pmatrix} \begin{pmatrix} e & f \\ f & e \end{pmatrix} \right\} = \begin{pmatrix} ace + bde + adf + bcf \\ cbe + ade + bdf + acf \\ acf + bdf + ade + bce \\ cbf + adf + bde + ace \end{pmatrix}$$

$$\begin{pmatrix} e & f \\ f & e \end{pmatrix} \otimes \begin{pmatrix} a & b \\ b & a \end{pmatrix} \text{vec}\left\{ \begin{pmatrix} c & d \\ d & c \end{pmatrix} \right\} = \begin{pmatrix} ae & be & af & bf \\ be & ae & bf & af \\ af & bf & ae & be \\ bf & af & be & ae \end{pmatrix} \begin{pmatrix} c \\ d \\ d \\ c \end{pmatrix}$$

2 Generalized Least Squares Estimator

In this section, we 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 heteroscedasticity with respect to the errors for the linear regression model:

$$y = Xb + u. \tag{2}$$

The definition is given as follows.

Definition 2.1 (A Linear Heteroscedastic Model)**.** We call *a linear heteroscedastic model* a model where the random vector y linearly depends on k explanatory variables X as (2) 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 OLS Estimator for a Linear Heteroscedastic Model

Here we consider the properties of the OLS estimator derived under the assumption [**H1**–**H3**].

Proposition 2.1. The OLS estimator

$$\hat{b} = \left(X'X\right)^{-1} X'y \tag{3}$$

becomes an unbiased estimator under [**GH1**–**GH3**].

Proof.

$$\mathbb{E}[\hat{b}|X] = \mathbb{E}\left[\left(X'X\right)^{-1}\left(X'y\right)\Big|X\right] = b + \mathbb{E}\left[\left(X'X\right)^{-1}X'u\Big|X\right] = b + \left(X'X\right)^{-1}X'\mathbb{E}\left[u\big|X\right] = b,$$

which shows that the OLS estimator (or (3)) is an unbiased estimator. □

Proposition 2.2 (Variance of the OLS estimator for a Linear Heteroscedastic Model)**.** Under the assumption [**H1**–**H3**], the variance of the OLS estimator $\mathbb{V}[\hat{b}]$ becomes

$$\mathbb{V}[\hat{b}|X] = \left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1}. \tag{4}$$

Proof. The variance of the OLS estimator is calculated as follows:

$$\begin{aligned}
\mathbb{V}[\hat{b}|X] &= \mathbb{E}\left[\left(\hat{b}-\mathbb{E}[\hat{b}|X]\right)\left(\hat{b}-\mathbb{E}[\hat{b}|X]\right)'\Big|X\right] \\
&= \mathbb{E}\left[\left(X'X\right)^{-1}X'uu'X\left(X'X\right)^{-1}\big|X\right] \\
&= \left(X'X\right)^{-1}X'\mathbb{E}\left[uu'\big|X\right]X\left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1}.
\end{aligned}$$

This proves (4). □

2.2 Derivation of the GLS Estimator

In this subsection, we derive the *Generalized Least Squares* (: GLS) estimator, which is defined as follows.

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

$$\hat{b}_{GLS} = \arg\min_b \|y - Xb\|^2_{\Omega^{-1}} = \arg\min_b \left(y - Xb\right)'\Omega^{-1}\left(y - Xb\right).$$

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

Theorem 2.1 (Generalized Least Squares Estimator)**.** Suppose [**H1**–**H3**] holds. Then the GLS estimator $\hat{b}$ exists uniquely and satisfies

$$\hat{b}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}y. \tag{5}$$

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

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

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

$$y^* = X^*b + u^*, \tag{6}$$

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^*:

$$\left(X^{*'}X^*\right)^{-1}\left(X^{*'}y^*\right) = \left(X\Omega^{-1/2'}\Omega^{-1/2}X'\right)^{-1}\left(X\Omega^{-1/2'}\Omega^{-1/2}y\right) = \left(X\Omega^{-1}X'\right)^{-1}X\Omega^{-1}y = \hat{b}_{GLS}, \tag{7}$$

which proves (5). □

To obtain the GLS estimator, we have another method, as in the derivation of the OLS estimator, by confirming the first and second order condition for the minimization problem of the following loss function $S(b)$:

$$\arg\min_b \|y - Xb\|_{\Omega^{-1}}^2 =: \arg\min_b S(b).$$

The first order condition becomes

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

Of course you can apply the chain rule here. Recall that $y'\Omega^{-1}Xb \in \mathbb{R}$ and thereby $y'\Omega^{-1}Xb = (y'\Omega^{-1}Xb)' = b'X'\Omega^{-1}y (\in \mathbb{R})$. The OLS estimator, denoted as $\hat{b}_{GLS}$, satisfies this equation, and hence

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

From the assumption **GH3**, the inverse matrix $(X'\Omega^{-1}X)^{-1}$ 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 (5). The second order condition becomes

$$\nabla_{bb'}^2 \|y - Xb\|_2^2 = 2X'\Omega^{-1}X.$$

By assumption **GH3**, $X'\Omega^{-1}X$ is a positive definite matrix. This shows that the loss function $S(b)$ has a minimum at the GLS estimator $\hat{b}$.

From the Theorem 2.1, we can confirm that the GLS estimator expressed as (5) is a random variable since we can rewite it as follows:

$$\hat{b} = b + \left(X'\Omega^{-1}X \right)^{-1} X'\Omega^{-1}u. \tag{8}$$

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 2.3 (Mean of the GLS Estimator)**.** Under the assumption [**GH1–GH3**], the conditional expectation of the GLS estimator $\hat{b}_{GLS}$ becomes

$$\mathbb{E}[\hat{b}_{GLS}|X] = b. \tag{9}$$

Proof. Calculating the expectation of $\hat{b}_{GLS}$ yields

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

which proves (9). □

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

$$\mathbb{E}[\hat{b}_{GLS}] = b.$$

from the *law of iterated expectation*:

$$\mathbb{E}[\hat{b}_{GLS}] = \mathbb{E}\left[\mathbb{E}[\hat{b}_{GLS}|X]\right] = \mathbb{E}[b] = b.$$

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

Proposition 2.4 (Variance of the GLS Estimator)**.** Under the assumption [**GH1**–**GH3**], the conditional variance of the OLS estimator $\hat{b}$ becomes

$$\mathbb{V}[\hat{b}_{GLS}|X] = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1}, \tag{10}$$

and the unconditional variance becomes

$$\mathbb{V}[\hat{b}_{GLS}] = \mathbb{E}\left[\left(X'\Omega^{-1}X\right)^{-1}\right]. \tag{11}$$

Proof. From the Eq, (8) and Eq, (9),

$$\hat{b}_{GLS} - \mathbb{E}[\hat{b}_{GLS}|X] = \hat{b}_{GLS} - b = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}u.$$

Therefore,

$$\begin{aligned}
\mathbb{V}[\hat{b}_{GLS}|X] &= \mathbb{E}\left[\left(\hat{b} - \mathbb{E}[\hat{b}|X]\right)\left(\hat{b} - \mathbb{E}[\hat{b}|X]\right)' \middle| X\right] \\
&= \mathbb{E}\left[\left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}uu'\Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1} |X\right] \\
&= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1} \underbrace{\mathbb{E}\left[uu'|X\right]}_{=\Omega} \Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1} \\
&= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1} \\
&= \left(X'\Omega^{-1}X\right)^{-1}.
\end{aligned}$$

This implies (10) holds. Thus,

$$\mathbb{V}[\hat{b}_{GLS}] = \mathbb{E}\left[\mathbb{V}[\hat{b}_{GLS}|X]\right] + \mathbb{V}\left[\mathbb{E}[\hat{b}_{GLS}|X]\right] = \mathbb{E}\left[\left(X'\Omega^{-1}X\right)^{-1}\right] + \mathbb{V}[b] = \mathbb{E}\left[\left(X'\Omega^{-1}X\right)^{-1}\right],$$

which porves (11). □

2.3 Properties of the GLS Estimator

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 $\hat{b}$ becomes an unbiased estimator:

$$\mathbb{E}[\hat{b}_{GLS}] = b; \tag{12}$$

(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 $\hat{b}_{GLS}$ satisfies

$$\hat{b}_{GLS} \xrightarrow[n\to\infty]{p} b; \tag{13}$$

(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 (7), we have

$$\begin{aligned}
\operatorname*{plim}_{n\to\infty} \hat{b}_{GLS} &= \operatorname*{plim}_{n\to\infty} \left[b + \left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}u\right)\right] \\
&= b + \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^{*\prime}X_i^*\right)^{-1} \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^{*\prime}u_i^*\right) \\
&= b + \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1} \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}u_i\right). \qquad (14)
\end{aligned}$$

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

$$\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}X_i \xrightarrow[n\to\infty]{p} \mathbb{E}\left[X_i'\Omega^{-1}X_i\right] < \infty; \tag{15}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}u_i \xrightarrow[n\to\infty]{p} \mathbb{E}\left[X_i'\Omega^{-1}u_i\right] = \mathbf{0}(\in \mathbb{R}^k). \tag{16}$$

We can prove $\mathbb{E}\left[X_i'\Omega^{-1}u_i\right] = \mathbf{0}$ 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,

$$\operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1} = \mathbb{E}\left[X_i'\Omega^{-1}X_i\right]^{-1} \tag{17}$$

holds from the *continuous mapping theorem*. Thus, substituting (15) and (16) into (14) results in

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

which indicates that $\hat{b} \xrightarrow[n\to\infty]{p} b$.

(iii) **Efficiency**: As for the efficiency of the GLS estimator, the following *Gauss–Markov theorem* supports the efficiency.

□

3 Gauss–Markov Theorem for the Generalized Least Squares Estimator

Here we will obtain a general result for the class of linear unbiased estimators of $\mathbf{b}$. It can be conducted via a direct method as we have seen in the derivation of the Gauss–Markov Theorem for the OLS estimator.

Theorem 3.1 (Gauss–Markov Theorem for the GLS Estimator)**.** Under the assumption [**GH1–GH3**], the GLS estimater $\hat{b}_{GLS}$ of the following regression model

$$y_i^* = X_i^* b + u_i^*, \tag{18}$$

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

Here we have used the following notations:

$$y^* := \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{pmatrix} \in \mathbb{R}^n, \quad X^* := \begin{pmatrix} X_1 & X_2 & \cdots & X_k \end{pmatrix} \in \mathcal{M}_{n\times k}(\mathbb{R}), \quad \text{and} \quad u^* := \begin{pmatrix} u_1^* \\ u_2^* \\ \vdots \\ u_n^* \end{pmatrix} \in \mathbb{R}^n.$$

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

$$\mathbb{E}[\tilde{b}] = b \tag{19}$$

holds, which yields

$$\mathbb{E}\left[A\left\{X^{*}b+u^{*}\right\}\right]=b \iff AX^{*}b=b. \tag{20}$$

Therefore, $AX^{*}=I_k$ must hold. Moreover, from the equation:

$$\tilde{b}-\mathbb{E}[\tilde{b}]=A\left\{y^{*}-X^{*}b\right\}=Au^{*}. \tag{21}$$

the variance $\mathbb{V}[\tilde{b}]$ becomes

$$\mathbb{V}[\tilde{b}]=\mathbb{V}\left[Au^{*}\right]=A\mathbb{V}\left[u^{*}\right]A'=AI_nA'=AA', \tag{22}$$

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

$$\mathcal{M}_{X^{*}}:=I_n-X\left(X^{*\prime}X^{*}\right)^{-1}X^{*\prime}\left(\iff I_n=\mathcal{M}_{X^{*}}+X^{*}\left(X^{*\prime}X^{*}\right)^{-1}X^{*\prime}\right), \tag{23}$$

we can rewrite (22) as follows:

$$\begin{aligned}\mathbb{V}[\tilde{b}]&=AI_nA'\\&=A\left(X^{*}\left(X^{*\prime}X^{*}\right)^{-1}X^{*\prime}+\mathcal{M}_{X^{*}}\right)A'\\&=AX^{*}\left(X^{*\prime}X^{*}\right)^{-1}X^{*\prime}A'+A\mathcal{M}_{X^{*}}A'.\end{aligned}$$

Substituting $AX^{*}(=X^{*\prime}A')=I_k$ and $\mathbb{V}[\hat{b}]=\left(X^{*\prime}X^{*}\right)^{-1}$ into the above equation results in

$$\mathbb{V}[\tilde{b}]=\mathbb{V}[\hat{b}]+A\mathcal{M}_{X^{*}}A' \iff \mathbb{V}[\tilde{b}]-\mathbb{V}[\hat{b}]=\sigma^2A\mathcal{M}_{X^{*}}A'.$$

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

$$\mathbb{V}[\tilde{b}]-\mathbb{V}[\hat{b}]=a_i'\mathcal{M}_{X^{*}}a_i\geq 0$$

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

4 Comparison of the OLS and GLS estimator

Here we compare the efficiency between OLS and GLS estimators. Accroding to the proceedding sections, under the assumption [**GH1**–**GH3**], we have

$$\begin{aligned}\mathbb{V}[\hat{b}_{OLS}|X]&=\left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1};\\\mathbb{V}[\hat{b}_{GLS}|X]&=\left(X'\Omega^{-1}X\right)^{-1}.\end{aligned}$$

Then, subtracting $\mathbb{V}[\hat{b}_{GLS}|X]$ from $\mathbb{V}[\hat{b}_{OLS}|X]$ results in

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

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

$$\mathbb{V}[\hat{b}_{OLS}|X]_{ii} - \mathbb{V}[\hat{b}_{GLS}|X]_{ii} > 0,$$

for all $i \in \{1, \dots, n\}$. These results infer that $\hat{b}_{GLS}$ is more efficient than $\hat{b}_{OLS}$.

5 Asymptotic Normality for the GLS Estimator

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

Theorem 5.1 (Asymptotic Normality of an GLS Estimator)**.** Let $\hat{b}_{GLS}$ be the GLS estimator obtained under the assumption [**GH1–GH3**]. Suppose

GH5: $b = \mathbb{E}[X_i'\Omega^{-1}u_i u_i \Omega^{-1} X_i]$ exists;

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

$$\sqrt{n}(\hat{b} - b) \xrightarrow[n\to\infty]{d} N_{\mathbb{R}^k}\left(\mathbf{0}, \mathbf{A}^{-1} b \mathbf{A}^{-1}\right).$$

Proof. From (8), we have

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

Therefore, rewriting results in

$$\sqrt{n}(\hat{b} - b) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i'\Omega^{-1}u_i\right). \tag{24}$$

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

$$\left(\frac{1}{n} X_i'\Omega^{-1}X_i\right)^{-1} \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}\left[X_i'\Omega^{-1}X_i\right]^{-1} = \mathbf{A};$$

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

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,

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

Therefore, from (24) and the **Slutsky's theorem**,

$$\sqrt{n}(\hat{b} - b) \xrightarrow[n\to\infty]{d} \mathbf{A}^{-1}\mathbf{Z},$$

where

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

From the following relation:

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

we obtain

$$\sqrt{n}(\hat{b} - b) \xrightarrow[n\to\infty]{d} N_{\mathbb{R}^k}\left(\mathbf{0}, b\right) \Longrightarrow \mathbf{A}^{-1}b \sim N_{\mathbb{R}^k}\left(\mathbf{0}, \mathbf{A}^{-1}b\mathbf{A}^{-1}\right).$$

$\square$