TA Session for Econometrics I 2025

10

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

June 21, 2025

Motivation

OLS:
Assume that $X \in \mathbb{R}^{n \times k}$ is a random matrix and $u \in \mathbb{R}^n$ is a random vector. We consider the following linear model:

$$y = X\beta + u$$

where $\beta \in \mathbb{R}^k$ is a parameter vector. We assume that the error term follows a multivariate normal distribution:

$$u|X \sim N_{\mathbb{R}^n}(\mathbf{0}, \sigma^2 I_n)$$

In the standard OLS framework, we impose a rather strong assumption on the error term u, namely:

- Homoskedasticity (constant variance)
- No correlation between errors

However, in many real-world situations, the error term exhibits **heteroskedasticity** and/or **correlation**, violating these assumptions.

Limitation of OLS

When the error term u is not homoskedastic or uncorrelated, the OLS estimator is still unbiased but no longer efficient. That is, it does not have the minimum variance among all linear unbiased estimators.

Generalised Least Squares (GLS) relaxes the strict assumptions of OLS by allowing a more flexible structure of the error variance:

$$u|X \sim N_{\mathbb{R}^n}(\mathbf{0}, \Sigma)$$

$\Sigma \in \mathbb{R}^{n \times n}$ is symmetric and positive definite.

- Σ represents the variance-covariance matrix of the error term.
- It allows for **heteroskedasticity** and **correlation** among errors.
- GLS provides efficient estimates when Σ is known or can be estimated.

Considering OLS Estimator under heteroskedasticity assumption

Assume that $X \in \mathbb{R}^{n \times k}$ is a random matrix and $u \in \mathbb{R}^n$ is a random vector. We consider the following linear model:

$$y = X\beta + u$$

where $\beta \in \mathbb{R}^k$ is a parameter vector. We assume that the error term follows a multivariate normal distribution:

$$u|X \sim N_{\mathbb{R}^n}(\mathbf{0}, \Sigma)$$

Even if we assume heteroskedasticity structure, we have the estimator remains the same:

$$\hat{\beta} = \arg\min_{\beta}(y - X\beta)^\top (y - X\beta)$$

F.O.C.:

$$-2X^\top (y - X\hat{\beta}) = 0$$

We have:

$$\hat{\beta}_{OLS} = (X^\top X)^{-1} X^\top y$$

Properties of $\hat{\beta}$

$$\begin{aligned} E[\hat{\beta}|X] &= E[(X^\top X)^{-1} X^\top y|X] \\ &= E[(X^\top X)^{-1} X^\top X\beta|X] + E[(X^\top X)^{-1} X^\top u|X] \\ &= \beta + (X^\top X)^{-1} X^\top E[u|X] \\ &= \beta \end{aligned}$$

Since we assume $E[u|X] = 0$, we can conclude that the unbiasedness still holds.

$$
\begin{aligned}
Var[\hat{\beta}|X] &= Var(\beta + (X^\top X)^{-1} X^\top u|X) \\
&= (X^\top X)^{-1} X^\top Var(u|X) X (X^\top X)^{-1} \\
&= (X^\top X)^{-1} X^\top \Sigma X (X^\top X)^{-1}
\end{aligned}
$$

GLS Model

Since Σ is by construction symmetric and positive definite, there exists a nonsingular $n \times n$ matrix, P, such that:

$$\Sigma = P^\top P.$$

By the linearity assumption, we can make the original model as follows:

$$P^{-1}y = P^{-1}X\beta + P^{-1}u.$$

We rewrite this model as

$$y^* = X^*\beta + u^*$$

where $u^*|X \sim N_{\mathbb{R}^n}(\mathbf{0}, I_n)$.

$$\hat{\beta}_{GLS} = \arg\min_{\beta} u^{*\top} u^*$$

$$\hat{\beta} = (X^{*\top} X^*)^{-1} X^{*\top} y^* = (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} y$$

$$\begin{aligned}
E[\hat{\beta}_{GLS}|X^*] &= E[\hat{\beta}_{GLS}|X] \\
&= E[(X^{*\top} X^*)^{-1} X^{*\top} y^*|X] \\
&= E[(X^{*\top} X^*)^{-1} X^{*\top} X^* \beta|X] + E[(X^{*\top} X^*)^{-1} X^{*\top} u^*|X] \\
&= \beta + (X^{*\top} X^*)^{-1} X^{*\top} E[u^*|X] \\
&= \beta
\end{aligned}$$

Since we assume $E[u|X] = 0$, we can conclude that the unbiasedness holds.

Remember that X and X^* have same information.

$$
\begin{aligned}
Var[\hat{\beta}_{GLS}|X] &= Var(\beta + (X^{*\top}X^*)^{-1}X^{*\top}u^*|X) \\
&= (X^{*\top}X^*)^{-1}X^{*\top}Var(u^*|X)X^*(X^{*\top}X^*)^{-1} \\
&= (X^{*\top}X^*)^{-1} \\
&= (X^\top(P^{-1})^\top P^{-1}X)^{-1} \\
&= (X^\top\Sigma^{-1}X)^{-1}
\end{aligned}
$$

Comparison of variances

Since the OLS estimator is still unbiased under heteroskedasticity, we have:

$$E[\hat{\beta}_{OLS}|X] = AE[y|X] = AX\beta = \beta$$

where AX must satisfy I_k. $(A = (X^\top X)^{-1} X^\top)$

Let A_{GLS} denote $(X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1}$, then

$$\hat{\beta}_{OLS} - \hat{\beta}_{GLS} = (A - A_{GLS})y$$

Now we consider the variance of this.

$$Var(\hat{\beta}_{OLS} - \hat{\beta}_{GLS}) = (A - A_{GLS})\Sigma(A - A_{GLS})^\top > 0$$

because Σ is positive (semi) definite matrix.

Therefore, we have

$$\mathrm{Var}(\hat{\beta}_{OLS}) \succeq \mathrm{Var}(\hat{\beta}_{GLS})$$

in the matrix sense. This proves that the GLS estimator is more efficient than any other linear unbiased estimator under the given assumptions.

$$\mathrm{Var}(\hat{\beta}_{OLS} - \hat{\beta}_{GLS}) = (A - A_{GLS})\Sigma(A - A_{GLS})^\top \succeq 0$$

because Σ is a positive semi-definite matrix.

Expanding the left-hand side gives:

$$\mathrm{Var}(\hat{\beta}_{OLS} - \hat{\beta}_{GLS}) = \mathrm{Var}(\hat{\beta}_{OLS}) - \mathrm{Var}(\hat{\beta}_{GLS}) + \text{cross terms}$$

However, since both estimators are linear and unbiased, and A_{GLS} minimises the variance among all such estimators, the cross terms vanish or do not affect the semi-definite inequality.

Therefore, we conclude that:

$$\mathrm{Var}(\hat{\beta}_{OLS}) \succeq \mathrm{Var}(\hat{\beta}_{GLS})$$

in the matrix sense. This confirms that the GLS estimator is more efficient than any other linear unbiased estimator under the given assumptions.

Setup and Objective

We consider the linear model:

$$y = X\beta + u, \quad \mathbb{E}[u] = 0, \quad \mathbb{E}[uu^\top] = \Sigma$$

Let $\tilde{\beta} = Ay$ be a linear estimator. The unbiasedness condition is:

$$\mathbb{E}[\tilde{\beta}] = AX\beta = \beta \quad \Rightarrow \quad AX = I_k$$

Goal: Show that

$$A\Sigma A^\top \succeq (X^\top \Sigma^{-1} X)^{-1}$$

for any matrix A satisfying $AX = I_k$.

Minimising the Variance: Lagrangian Method

We minimise the trace of the variance:

$$\min_A \operatorname{tr}(A\Sigma A^\top) \quad \text{subject to } AX = I_k$$

Set up the Lagrangian:

$$\mathcal{L}(A, \Lambda) = \operatorname{tr}(A\Sigma A^\top) + \operatorname{tr}\left[\Lambda^\top (I_k - AX)\right]$$

First-order condition:

$$\frac{\partial \mathcal{L}}{\partial A} = 2A\Sigma - \Lambda X^\top = 0 \Rightarrow A\Sigma = \frac{1}{2}\Lambda X^\top$$

Use the constraint $AX = I_k$ to solve for A.

Deriving the GLS Estimator and Efficiency

From $A\Sigma = \frac{1}{2}\Lambda X^\top$ and $AX = I_k$, we obtain:

$$A = (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} \equiv A_{GLS}$$

Thus, A_{GLS} minimises $\mathrm{tr}(A\Sigma A^\top)$ among all linear unbiased estimators.

For any A such that $AX = I_k$, we have:

$$A\Sigma A^\top = A_{GLS}\Sigma A_{GLS}^\top + (A - A_{GLS})\Sigma(A - A_{GLS})^\top \succeq A_{GLS}\Sigma A_{GLS}^\top$$

Conclusion: The GLS estimator achieves the smallest variance among all linear unbiased estimators.