Gauss-Markov Theorem (Matrix Form)

Hanxiao Du

Theorem 1 (Gauss-Markov theorem)**.** *Suppose we have a linear model $y = X\beta + \epsilon$, where ϵ is a column vector containing all residuals ϵ_i's.*
Under the Gauss-Markov assumptions,

1. *The residuals ϵ_i have mean zero, i.e. $\mathbb{E}[\epsilon_i] = 0$ or $\mathbb{E}[\epsilon] = \vec{0}$.*

2. *They are homoscedastic, i.e. they all have the same finite variance:*

$$Var(\epsilon_i) = \sigma^2 < \infty$$

3. *Distinct error terms are uncorrelated, i.e. $Cov(\epsilon_i, \epsilon_j) = 0\ \forall i \neq j$.*

or equivalently to assumption 2 and 3, $Var(\epsilon) = \sigma^2 I$ where $\sigma^2 < \infty$, we have that the ordinary least squares (OLS) of the parameter:

$$\hat{\beta}_{OLS} = (X^TX)^{-1}X^Ty$$

is the BLUE (Best Linear Unbiased Estimator), i.e. it is the most efficient estimator (has lowest variance) among all linear unbiased estimators.

Proof.

$$\begin{aligned}
\hat{\beta}_{OLS} &= (X^TX)^{-1}X^Ty && (1)\\
&= (X^TX)^{-1}X^T(X\beta + \epsilon) && (2)\\
&= (X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^T\epsilon && (3)\\
&= \beta + (X^TX)^{-1}X^T\epsilon && (4)
\end{aligned}$$

Thus, the unbiasedness of $\hat{\beta}_{OLS}$ gives:

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_{OLS}] &= \mathbb{E}[\beta + (X^TX)^{-1}X^T\epsilon] && (5)\\
&= \beta + (X^TX)^{-1}X^T\mathbb{E}[\epsilon] && (6)\\
&= \beta, \text{ since by assumption 1, } \mathbb{E}[\epsilon] = \vec{0}. && (7)
\end{aligned}$$

Also by assumption 1,

$$Var(\epsilon) = \mathbb{E}[(\epsilon - \mathbb{E}[\epsilon])(\epsilon - \mathbb{E}[\epsilon])^T] = \mathbb{E}[\epsilon\epsilon^T] = \sigma^2 I \quad (8)$$

Thus,

$$Var(y) = Var(X\beta + \epsilon) = Var(\epsilon) = \sigma^2 I \tag{9}$$

Now, we calculate the variance of the $\hat{\beta}_{OLS}$:

$$
\begin{aligned}
Var(\hat{\beta}_{OLS}) &= \mathbb{E}[(\hat{\beta}_{OLS} - \beta)(\hat{\beta}_{OLS} - \beta)^T] && (10)\\
&= \mathbb{E}[((X^TX)^{-1}X^T\epsilon)((X^TX)^{-1}X^T\epsilon)^T] && (11)\\
&= \mathbb{E}[(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1}] && (12)\\
&= (X^TX)^{-1}X^T\mathbb{E}[\epsilon\epsilon^T]X(X^TX)^{-1} && (13)\\
&= (X^TX)^{-1}X^T\sigma^2IX(X^TX)^{-1} \text{ by (8)} && (14)\\
&= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} && (15)\\
&= \sigma^2(X^TX)^{-1} && (16)
\end{aligned}
$$

Now, for any linear unbiased estimator $\hat{\beta}$ for β, it must can be expressed in linear form:

$$\hat{\beta} = Cy$$

for some constant matrix C.
Also, let D be the matrix such that $C = (X^TX)^{-1}X^T + D$.
Additionally, by the unbiasedness assumption of the estimator $\hat{\beta}$:

$$
\begin{aligned}
\mathbb{E}[\hat{\beta}] &= \mathbb{E}[Cy] && (17)\\
&= \mathbb{E}[((X^TX)^{-1}X^T + D)(X\beta + \epsilon)] && (18)\\
&= \beta + DX\beta = \beta && (19)\\
\implies DX\beta &= \underset{\sim}{0} \implies DX = \underset{\sim}{0} && (20)
\end{aligned}
$$

where $\underset{\sim}{0}$ is the zero matrix.
Another fact from linear algebra:
Finally, calculate the variance of $\hat{\beta}$:

$$
\begin{aligned}
Var(\hat{\beta}) &= Var[Cy] && (21)\\
&= CVar(y)C^T && (22)\\
&= \sigma^2CC^T \text{ by (9)} && (23)\\
&= \sigma^2[(X^TX)^{-1}X^T + D][X(X^TX)^{-1} + D^T] && (24)\\
&= \sigma^2[(X^TX)^{-1}X^TX(X^TX)^{-1} + (X^TX)^{-1}(DX)^t] && (25)\\
&+ DX(X^TX)^{-1} + DD^T] && (26)\\
&= \sigma^2(X^TX)^{-1} + \sigma^2DD^T \text{ by (20)} && (27)\\
&= Var(\hat{\beta}_{OLS}) + \sigma^2DD^T && (28)
\end{aligned}
$$

notice that $\sigma^2 > 0$ and DD^T is definite symmetric.
Therefore, $\hat{\beta}_{OLS}$ has the lowest variance among all linear unbiased estimator so that it is the BLUE (Best Linear Unbiased Estimator).

■