Gauss-Markov Theorem

The Gauss-Markov Theorem is given in the following regression model and assumptions:

The regression model

$$y_i = \beta_1 + \beta_2 x_i + u_i, \qquad i = 1, \cdots, n \tag{1}$$

Assumptions (A) or **Assumptions (B)**:

$$\textbf{Assumptions (A)} \begin{cases} \mathrm{E}u_i & = & 0 & \text{for all } i \\ \mathrm{Var}(u_i) & = & \sigma^2 & \text{for all } i \\ & & & \text{(homoscedasticity)} \\ \mathrm{Cov}(u_i, u_j) & = & 0 & \text{for all } \ i \neq j \\ x_i & & \text{is} & \text{nonstochastic constant} \end{cases}$$

The text book uses Assumptions (B) (see p.588 of the text):

$$\textbf{Assumptions (B)} \begin{cases} \mathrm{E}(u_i | x_1, \cdots, x_n) & = & 0 & \text{for all } i \\ \mathrm{Var}(u_i | x_1, \cdots, x_n)) & = & \sigma^2 & \text{for all } i \\ & & & \text{(homoscedasticity)} \\ \mathrm{Cov}(u_i, u_j | x_1, \cdots, x_n) & = & 0 & \text{for all } \ i \neq j \end{cases}$$

If we use Assumptions (B), we need to use the law of iterated expectations in proving the BLUE. With Assumptions (B), the BLUE is given **conditionally on** $x_1, \cdots, x_n$.

Let us use Assumptions (A). The Gauss-Markov Theorem is stated in the boxed statement below:

Gauss Markov Theorem

> Under Assumptions (A), the OLS estimators, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are the **Best Linear Unbiased Estimator (BLUE)**, that is
>
> 1. **Unbias**: $\mathrm{E}\widehat{\beta}_1 = \beta_1$ and $\mathrm{E}\widehat{\beta}_2 = \beta_2$
>
> 2. **Best**: $\widehat{\beta}_1$ and $\widehat{\beta}_2$ have the smallest variances **among the class of all linear unbiased estimators.**

Real data seldomly satisfy Assumptions (A) or Assumptions (B). Accordingly we should think that the Gauss-Markov theorem only holds in the never-never land. However, it is important to understand the Gauss-Markov theorem on two grounds:

1. We may treat the world of the Gauss-Markov theorem as equivalent to the world of perfect competition in micro economic theory.

2. The mathematical exercises are good for your souls.

We shall prove the Gauss-Markov theorem using the simple regression model of equation (1). We can prove the Gauss-Markov theorem using the multiple regression model

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i, \quad i = 1, \cdots, n. \tag{2}$$

To do so, however, we need to use vector and matrix language (linear algebra.) Actually, once you learn linear algebra the proof of Gauss-Markov theorem is far more straight forward than the proof for the simple regression model of (1).

In the text book the Gauss-Markov theorem is discussed on the following pages:

$$\begin{cases} 127 \\ 590 - -591 \\ 604 - -605 \\ 622 - -623 \\ 641 - -642 \end{cases}$$

You should take a look at these pages.

Proving the Gauss-Markov Theorem

The unbiasedness of $\widehat{\beta}_1$ and of $\widehat{\beta}_2$ are given in the **Comments on the Midterm Examination** and the answers to Assignment #5. So, we prove here the minimum variance properties. There are generally two ways to prove bestness: (i) using linear algebra, and (ii) using calculus. We prove bestness using linear algebra first, and we leave the proof using calculus to the **Appendix**. First we prove that $\widehat{\beta}_1$ has the smallest variance among all other linear estimators of β_1.

Proof that $\widehat{\beta}_1$ is best

We need to re-express $\widehat{\beta}_1$ first.

$$\begin{aligned}
\widehat{\beta}_1 &= \bar{y} - \widehat{\beta}_2 \bar{x} = \frac{1}{n}\sum y_i - \left(\sum \frac{(x_i - \bar{x})y_i}{s_{xx}}\,\bar{x}\right) \\
&= \sum \frac{1}{n}\, y_i - \sum \frac{(x_i - \bar{x})\bar{x}}{s_{xx}}\, y_i \\
&= \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{s_{xx}}\right) y_i, \quad \text{where} \;\; s_{xx} = \sum x_i^2 - n\,\bar{x}^2 \\
&= \sum w_i y_i, \quad \text{where} \;\; w_i = \frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{s_{xx}}.
\end{aligned}$$

The **BLUE** only looks at linear estimators of β_1. The linear estimators are defined by

$$\tilde{\beta}_1 = \sum_{i=1}^{n} a_i y_i.$$

In passing we notice that if

$$a_i = w_i, \quad \text{for all} \;\; i = 1, \cdots, n$$

then $\tilde{\beta}_1 = \widehat{\beta}_1$.

We have to make $\tilde{\beta}_1$ unbiased. To take expectation of $\tilde{\beta}_1$ we first substitute equation (1): $y_i = \beta_1 + \beta_2 x_i + u_i$ for y_i:

$$\begin{aligned}
\tilde{\beta}_1 &= \sum_{i=1}^{n} a_i y_i = \sum_{i=1}^{n} a_i(\beta_1 + \beta_2 x_i + u_i) = \beta_1 \sum a_i + \beta_2 \sum a_i x_i + \sum a_i u_i. \\
\mathrm{E}\tilde{\beta}_1 &= \beta_1 \sum a_i + \beta_2 \sum a_i x_i + \sum a_i \mathrm{E} u_i = \beta_1 \sum a_i + \beta_2 \sum a_i x_i,
\end{aligned}$$

since $\mathrm{E}u_i = 0$ for all i. We see that

$$\mathrm{E}\tilde{\beta}_1 = \beta_1 \iff \sum a_i = 1 \quad \sum a_i x_i = 0$$

($\iff$) means "if and only if."

We take variance of $\tilde{\beta}_1$:

$$\begin{aligned}
\mathrm{Var}(\tilde{\beta}_1) &\equiv \mathrm{E}(\tilde{\beta}_1 - \mathrm{E}\tilde{\beta}_1)^2 = \mathrm{E}(\tilde{\beta}_1 - \beta_1)^2 \quad \text{since } \mathrm{E}\tilde{\beta}_1 = \beta_1 \\
&= \mathrm{E}(\sum a_i u_i)^2 \\
&= (a_1^2 \mathrm{E}u_1^2 + \cdots + a_n^2 \mathrm{E}u_n^2 + 2a_1 a_2 \mathrm{E}u_1 u_2 + \cdots + 2a_{n-1}a_n \mathrm{E}u_{n-1}u_n) \\
&= \sigma^2 (a_1^2 + \cdots + a_n^2) = \sigma^2 \sum a_i^2,
\end{aligned}$$

since $\mathrm{E}u_i^2 = \sigma^2$ and $\mathrm{E}u_i u_j = 0,\ \ i \neq\ j$. The variance of the OLS estimator, $\mathrm{Var}(\widehat{\beta}_1)$ is given by

$$\mathrm{Var}(\widehat{\beta}_1) = \sigma^2 \sum w_i^2.$$

We see

$$\mathrm{Var}(\tilde{\beta}_1) \ \geq\ \mathrm{Var}(\widehat{\beta}_1) \iff \sum_{i=1}^{n} a_i^2 \ \geq\ \sum_{i=1}^{n} w_i^2.$$

Since a_i is an arbitrary nonstochastic constant we can rewrite a_i as

$$a_i = w_i + d_i.$$

Earlier we saw that $\tilde{\beta}_1$ is unbiased if and only if $\sum a_i = 1$ and $\sum a_i x_i = 1$. So,

$$\begin{aligned}
\sum a_i &= \sum w_i + \sum d_i = 1 \\
\sum a_i x_i &= \sum w_i x_i + \sum d_i x_i = 0.
\end{aligned}$$

But

$$\begin{aligned}
\sum w_i &= \sum \left(\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{s_{xx}} \right) = 1 - \frac{\bar{x}}{s_{xx}} \sum (x_i - \bar{x}) = 1 \\
\sum w_i x_i &= \sum \left(\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{s_{xx}} \right) x_i = \frac{1}{n} \sum x_i - \frac{\sum x_i^2 - n\,\bar{x}^2}{s_{xx}}\, \bar{x} = \bar{x} - \bar{x} = 0.
\end{aligned}$$

Hence

$$\sum d_i = 0 \quad \text{and} \quad \sum d_i x_i = 0.$$

We square a_i^2 and sum with respect to $i = 1, \cdots, n$:

$$\sum a_i^2 = \sum (w_i + d_i)^2 = \sum w_i^2 + \sum d_i^2 + 2 \sum w_i d_i = \sum w_i^2 + \sum d_i^2$$

since the cross product term is zero:

$$\sum w_i d_i = \sum \left(\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{s_{xx}} \right) d_i = \frac{1}{n} \sum d_i - \frac{1}{s_{xx}} (\sum d_i x_i - \bar{x} \sum d_i) = 0$$

Hence

$$\sum a_i^2 = \sum w_i^2 + \sum d_i^2 \geq \sum w_i^2,$$

and this concludes the proof.

Proof that $\widehat{\beta}_2$ is best

$$\widehat{\beta}_2 = \frac{\sum (x_i - \bar{x}) y_i}{s_{xx}} = \sum \left(\frac{x_i - \bar{x}}{s_{xx}} \right) y_i = \sum v_i y_i$$

where

$$v_i = \frac{x_i - \bar{x}}{s_{xx}}$$

We shall use the fact that

$$\sum v_i = 0 \quad \text{and} \quad \sum v_i^2 = s_{xx}.$$

The variance of $\widehat{\beta}_2$, $\mathrm{Var}(\widehat{\beta}_2)$, is given by

$$\mathrm{Var}(\widehat{\beta}_2) = \sigma^2 \sum v_i^2.$$

Let $\tilde{\beta}_2$ be a linear estimator of β_2:

$$\tilde{\beta}_2 = \sum b_i y_i.$$

We need to find the conditions that make $\tilde{\beta}_2$ unbiased. Taking expectation we have

$$\mathrm{E}\tilde{\beta}_2 = \mathrm{E} \sum b_i (\beta_1 + \beta_2 x_i + u_i) = \beta_1 \sum b_i + \beta_2 \sum b_i x_i$$

and thus

$$\mathrm{E}(\tilde{\beta}_2) = \beta_2 \iff \sum b_i = 0, \ \sum b_i x_i = 1.$$

The variance of $\tilde{\beta}_2$, $\mathrm{Var}(\tilde{\beta}_2)$, is

$$\mathrm{Var}(\tilde{\beta}_2) = \sigma^2 \sum b_i^2.$$

Let

$$b_i = v_i + c_i$$

then

$$\begin{aligned}\sum b_i &= \sum v_i + \sum c_i = \sum c_i = 0\\ \sum b_i x_i &= \sum v_i x_i + \sum c_i x_i \implies \sum c_i x_i = 0\end{aligned}$$

since

$$\sum v_i x_i = \frac{1}{s_{xx}} \sum (x_i - \bar{x}) x_i = \frac{\sum x_i^2 - n\,\bar{x}^2}{s_{xx}} = 1.$$

So the variance of $\tilde{\beta}_2$ becomes

$$\mathrm{Var}(\tilde{\beta}_2) = \sigma^2 \sum b_i^2 \;=\; \sigma^2 \;, \sum (v_i + c_i)^2$$

$$= \;\; \sigma^2 (\sum v_i^2 + \sum c_i^2 + 2 \sum v_i c_i) \;\geq\; \sigma^2 \sum v_i^2 + \sigma^2 \sum c_i^2 \;\geq\; \mathrm{Var}(\widehat{\beta}_2)$$

since

$$\sum v_i c_i = \frac{1}{s_{xx}} \sum (x_i - \bar{x}) c_i = \frac{1}{s_{xx}} (\sum x_i c_i - \bar{x} \sum c_i) = 0.$$

Appendix: Proving Bestness using calculus

Another way to prove that the OLS estimators, $\widehat{\beta}_1$ and $\widehat{\beta}_2$, are best is to use calculus to find the minimum variance. Since variance is a quadratic function, it is twice differentiable and thus we may use calculus to find the minimum.

Proving that $\widehat{\beta}_1$ is best

The variance of a linear unbiased estimator is given by

$$\sigma^2 \sum a_i^2$$

with two linear constraints

$$\sum a_i = 1 \quad \text{and} \quad \sum a_i x_i = 0.$$

Hence we may form the following minimization problem subject to the linear constraints:

$$\min_{a_1, \cdots, a_n} \sigma^2 \sum a_i^2$$

$$\text{subject to} \qquad \begin{cases} \sum a_i &= \;\; 1 \\ \sum a_i x_i &= \;\; 0 \end{cases}$$

We form the Lagrangian

$$\Lambda = \sigma^2 \sum a_i^2 - \lambda_1(\sum a_i - 1) - \lambda_2 \sum a_i x_i.$$

The first order conditions are

$$\begin{aligned}
\frac{\partial \Lambda}{\partial a_1} &= 2\sigma^2 a_1 - \lambda_1 - x_1\lambda_2 = 0 \qquad (1)\\
\frac{\partial \Lambda}{\partial a_2} &= 2\sigma^2 a_2 - \lambda_1 - x_2\lambda_2 = 0 \qquad (2)\\
\vdots \quad &\vdots \quad \vdots\\
\frac{\partial \Lambda}{\partial a_n} &= 2\sigma^2 a_n - \lambda_1 - x_n\lambda_2 = 0 \qquad (\text{n})\\
\frac{\partial \Lambda}{\partial \lambda_1} &= -\sum a_i + 1 = 0 \qquad (\text{n+1})\\
\frac{\partial \Lambda}{\partial \lambda_2} &= -\sum a_i x_i = 0 \qquad (\text{n+2})
\end{aligned}$$

Adding the left hand and right hand sides of equations (1)–(n) we have

$$2\,\sigma^2 \sum a_i - n\,\lambda_1 - \lambda_2 \sum x_i = 0.$$

Since $\sum a_i = 1$

$$2\sigma^2 - n\,\lambda_1 - \lambda_2 n\,\bar{x} = 0 \qquad (*)$$

Multiplying the left hand and right hand sides of equations (1)–(n) by $x_1,\, x_2, \cdots x_n$ respectively and adding up we have

$$2\,\sigma^2 \sum a_i x_i - \lambda \sum x_i - \lambda_2 \sum x_i^2 = 0$$

Since $\sum a_i x_i = 0$ we have

$$-n\,\bar{x}\,\lambda_1 - \lambda_2 \sum x_i^2 = 0 \qquad (**)$$

Equations (*) and (**) form a linear equation system in λ_1 and in λ_2 :

$$\begin{cases} n\,\lambda_1 + \lambda_2 n\,\bar{x} &= 2\,\sigma^2 \\ n\,\bar{x}\lambda_1 + \lambda_2 \sum x_i^2 &= 0 \end{cases}$$

Solving for λ_1 and for λ_2 we have

$$\lambda_1 = \frac{2\,\sigma^2}{n\,s_{xx}} \sum x_i^2, \quad \text{and} \quad \lambda_2 = -\frac{2\sigma^2 \bar{x}}{s_{xx}}.$$

From equations (1)–(n)we have

$$2\,\sigma^2 a_i = \lambda_1 + \lambda_2 x_i, \quad i = 1, \cdots, n.$$

Substituting for λ_1 and for λ_2 we obtain

$$
\begin{aligned}
2\,\sigma^2 a_i &= \frac{2\,\sigma^2}{n\,s_{xx}}\sum x_i^2 - \frac{2\,\sigma^2 \bar{x} x_i}{s_{xx}} \qquad i = 1, \cdots, n \\
\text{or} & \\
a_i &= \frac{\sum x_i^2}{n\,s_{xx}} - \frac{\bar{x}\,x_i}{s_{xx}} = \frac{s_{xx} + n\,\bar{x}^2}{n\,s_{xx}} - \frac{\bar{x}\,x_i}{s_{xx}} \\
& \qquad \text{since} \quad \sum x_i^2 = s_{xx} + n\,\bar{x}^2, \\
&= \frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{s_{xx}} = w_i, \quad i = 1, \cdots, n.
\end{aligned}
$$

w_i is for the OLS estimator of β_1, $\widehat{\beta}_1$.

The second order conditions are

$$\frac{\partial^2 \Lambda}{\partial\, a_1^2} = 2\,\sigma^2,\ \cdots,\ \frac{\partial^2 \Lambda}{\partial\, a_n^2} = 2\,\sigma^2, \quad \frac{\partial^2 \Lambda}{\partial\, \lambda_1^2} = 0,\ \ \frac{\partial^2 \Lambda}{\partial\, \lambda_2^2} = 0,$$

and the cross-derivatives

$$\frac{\partial^2 \Lambda}{\partial a_1 \partial a_2} = 0,\ \cdots, ;\ \frac{\partial^2 \Lambda}{\partial a_n \partial a_{n-1}} = 0$$

$$\frac{\partial^2 \Lambda}{\partial \lambda_1 \partial a_i} = -1, \quad i = 1, \cdots, n$$

$$\frac{\partial^2 \Lambda}{\partial \lambda_2 \partial a_i} = -x_i, \quad i = 1, \cdots, n.$$

Hence the bordered Hessian becomes

$$H = \begin{bmatrix}
\frac{\partial^2 \Lambda}{\partial a_1^2} & \frac{\partial^2 \Lambda}{\partial a_1 \partial a_2} & \cdots & \frac{\partial^2 \Lambda}{\partial a_1 \partial a_n} & \frac{\partial^2 \Lambda}{\partial a_1 \partial \lambda_1} & \frac{\partial^2 \Lambda}{\partial a_1 \partial \lambda_2} \\
\vdots & \vdots & & \vdots & \vdots & \vdots \\
\frac{\partial^2 \Lambda}{\partial a_n \partial a_1} & \frac{\partial^2 \Lambda}{\partial a_n \partial a_2} & \cdots & \frac{\partial^2 \Lambda}{\partial a_n^2} & \frac{\partial^2 \Lambda}{\partial a_n \partial \lambda_1} & \frac{\partial^2 \Lambda}{\partial a_n \partial \lambda_2} \\
\frac{\partial^2 \Lambda}{\partial \lambda_1 \partial a_1} & \frac{\partial^2 \Lambda}{\partial \lambda_1 \partial a_2} & \cdots & \frac{\partial^2 \Lambda}{\partial \lambda_1 \partial a_n} & \frac{\partial^2 \Lambda}{\partial \lambda_1^2} & \frac{\partial^2 \Lambda}{\partial \lambda_1 \partial \lambda_2} \\
\frac{\partial^2 \Lambda}{\partial \lambda 2 \partial a_1} & \frac{\partial^2 \Lambda}{\partial \lambda_2 \partial a_2} & \cdots & \frac{\partial^2 \Lambda}{\partial \lambda 2 \partial a_n} & \frac{\partial^2 \Lambda}{\partial \lambda_2 \partial \lambda_1} & \frac{\partial^2 \Lambda}{\partial \lambda_2^2}
\end{bmatrix}.$$

This becomes

$$H = \begin{bmatrix} 2\sigma^2 & 0 & 0 & \cdots & 0 & -1 & -x_1 \\ 0 & 2\sigma^2 & 0 & \cdots & 0 & -1 & -x_2 \\ \vdots & \vdots & \ddots & & \vdots & \vdots & \vdots \\ & & & \ddots & & & \\ 0 & 0 & 0 & \cdots & 2\sigma^2 & -1 & -x_n \\ -1 & -1 & -1 & \cdots & -1 & 0 & 0 \\ -x_1 & -x_2 & -x_3 & \cdots & -x_n & 0 & 0 \end{bmatrix}.$$

and it can be proved that H is negative definite, and hence the solutions

$$a_1 = w_1,\ a_2 = w_2, \cdots, a_n = w_n$$

yield the minimum variance.

Proving that $\widehat{\beta}_2$ is best

The constrained minimization problem becomes

$$\min_{b_1,\cdots,b_n} \sigma^2 \sum b_i^2$$

$$\text{subject to} \qquad \begin{cases} \sum b_i & = & 0 \\ \sum b_i x_i & = & 1 \end{cases}$$

The first order conditions are

$$\begin{aligned}
\frac{\partial\Lambda}{\partial b_1} &= 2\sigma^2 b_1 - \lambda_1 - x_1\lambda_2 = 0 && (1) \\
\frac{\partial\Lambda}{\partial b_2} &= 2\sigma^2 b_2 - \lambda_1 - x_2\lambda_2 = 0 && (2) \\
\vdots \quad & \vdots \quad \vdots \\
\frac{\partial\Lambda}{\partial b_n} &= 2\sigma^2 b_n - \lambda_1 - x_n\lambda_2 = 0 && (n) \\
\frac{\partial\Lambda}{\partial \lambda_1} &= -\sum b_i = 0 && (n+1) \\
\frac{\partial\Lambda}{\partial \lambda_2} &= -\sum b_i x_i + 1 = 0 && (n+2)
\end{aligned}$$

We proceed just in the same way as we did before and obtain

$$\begin{cases} \lambda_1 + \bar{x}\,\lambda_2 & = & 0 \\ n\,\bar{x}\,\lambda_1 + (\sum x_i^2)\lambda_2 & = & 2\,\sigma^2 \end{cases}$$

Solving for λ_1 and for λ_2 we obtain

$$\begin{aligned}\lambda_1 &= -\frac{2\sigma^2\bar{x}}{s_{xx}}\\ \lambda_2 &= \frac{2\sigma^2}{s_{xx}}\end{aligned}$$

Substituting for λ_1 and for λ_2 we obtain

$$\begin{aligned}2\,\sigma^2 b_i &= \lambda_1 + \lambda_2\,x_i = -\frac{2\sigma^2\bar{x}}{s_{xx}} + \frac{2\sigma^2 x_i}{s_{xx}}\\ &= 2\,\sigma^2\left(\frac{x_i - \bar{x}}{s_{xx}}\right)\end{aligned}$$

Hence

$$b_i = \frac{x_i - \bar{x}}{s_{xx}} = v_i, \quad i = 1, \cdots, n.$$

The second order conditions are obtained in a similar way and the bordered Hessian is negative definite.