Proof of the Gauss-Markov Theorem

The Gauss-Markov Theorem

Under the Gauss-Markov Linear Model, the OLS estimator $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ of an estimable linear function $\boldsymbol{c}'\boldsymbol{\beta}$ is the unique *Best Linear Unbiased Estimator* (BLUE) in the sense that $\text{Var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}})$ is strictly less than the variance of any other linear unbiased estimator of $\boldsymbol{c}'\boldsymbol{\beta}$.

Unbiased Linear Estimators of $\boldsymbol{c}'\boldsymbol{\beta}$

- If $\boldsymbol{a}$ is a fixed vector, then $\boldsymbol{a}'\boldsymbol{y}$ is a *linear function of* $\boldsymbol{y}$.
- An estimator that is a linear function of $\boldsymbol{y}$ is said to be a *linear estimator*.
- A linear estimator $\boldsymbol{a}'\boldsymbol{y}$ is an unbiased estimator of $\boldsymbol{c}'\boldsymbol{\beta}$ if and only if

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{c}'\boldsymbol{\beta} \;\forall\; \boldsymbol{\beta} \in \mathbb{R}^{\mathrm{p}} \quad &\Longleftrightarrow \quad \boldsymbol{a}'\mathrm{E}(\boldsymbol{y}) = \boldsymbol{c}'\boldsymbol{\beta} \;\forall\; \boldsymbol{\beta} \in \mathbb{R}^{\mathrm{p}} \\
&\Longleftrightarrow \quad \boldsymbol{a}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{c}'\boldsymbol{\beta} \;\forall\; \boldsymbol{\beta} \in \mathbb{R}^{p} \\
&\Longleftrightarrow \quad \boldsymbol{a}'\boldsymbol{X} = \boldsymbol{c}'.
\end{aligned}
$$

The OLS Estimator of $\boldsymbol{c}'\boldsymbol{\beta}$ is a Linear Estimator

- We have previously defined the Ordinary Least Squares (OLS) estimator of an estimable $\boldsymbol{c}'\boldsymbol{\beta}$ by $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is any solution to the normal equations $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}$.

- We have previously shown that $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is the same for any $\hat{\boldsymbol{\beta}}$ that is a solution to the normal equations.

- We have previously shown that $(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$ is a solution to the normal equations for any generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$ denoted by $(\boldsymbol{X}'\boldsymbol{X})^{-}$.

- Thus, $\boldsymbol{c}'\hat{\boldsymbol{\beta}} = \boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{\ell}'\boldsymbol{y}$ (where $\boldsymbol{\ell}' = \boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$) so that $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is a linear estimator.

$\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is an Unbiased Estimator of an Estimable $\boldsymbol{c}'\boldsymbol{\beta}$

- By definition, $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable if and only if there exists a linear unbiased estimator of $\boldsymbol{c}'\boldsymbol{\beta}$.

- It follows from slide 3 that $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable if and only if $\boldsymbol{c}' = \boldsymbol{a}'\boldsymbol{X}$ for some vector $\boldsymbol{a}$.

- If $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable, then

$$\boldsymbol{\ell}'\boldsymbol{X} = \boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{a}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{a}'\boldsymbol{P_X}\boldsymbol{X} = \boldsymbol{a}'\boldsymbol{X} = \boldsymbol{c}'.$$

- Thus, by slide 3, $\boldsymbol{c}'\hat{\boldsymbol{\beta}} = \boldsymbol{\ell}'\boldsymbol{y}$ is an unbiased estimator of $\boldsymbol{c}'\boldsymbol{\beta}$ whenever $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable.

Proof of the Gauss-Markov Theorem

- Suppose $\boldsymbol{d}'\boldsymbol{y}$ is any linear unbiased estimator other than the OLS estimator $\boldsymbol{c}'\hat{\boldsymbol{\beta}} = \boldsymbol{\ell}'\boldsymbol{y}$.
- Then we know the following:
  1. $\boldsymbol{d} \neq \boldsymbol{\ell} \iff ||\boldsymbol{d} - \boldsymbol{\ell}||^2 = (\boldsymbol{d} - \boldsymbol{\ell})'(\boldsymbol{d} - \boldsymbol{\ell}) > 0$, and
  2. $\boldsymbol{d}'\boldsymbol{X} = \boldsymbol{\ell}'\boldsymbol{X} = \boldsymbol{c}' \implies \boldsymbol{d}'\boldsymbol{X} - \boldsymbol{\ell}'\boldsymbol{X} = \boldsymbol{0}' \implies (\boldsymbol{d} - \boldsymbol{\ell})'\boldsymbol{X} = \boldsymbol{0}'$.
- We need to show $\text{Var}(\boldsymbol{d}'\boldsymbol{y}) > \text{Var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}})$.

Proof of the Gauss-Markov Theorem

$$\begin{aligned}
\text{Var}(\boldsymbol{d}'\boldsymbol{y}) &= \text{Var}(\boldsymbol{d}'\boldsymbol{y} - \boldsymbol{c}'\hat{\boldsymbol{\beta}} + \boldsymbol{c}'\hat{\boldsymbol{\beta}}) \\
&= \text{Var}(\boldsymbol{d}'\boldsymbol{y} - \boldsymbol{c}'\hat{\boldsymbol{\beta}}) + \text{Var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) + 2\text{Cov}(\boldsymbol{d}'\boldsymbol{y} - \boldsymbol{c}'\hat{\boldsymbol{\beta}}, \boldsymbol{c}'\hat{\boldsymbol{\beta}}).
\end{aligned}$$

$$\begin{aligned}
\text{Var}(\boldsymbol{d}'\boldsymbol{y} - \boldsymbol{c}'\hat{\boldsymbol{\beta}}) &= \text{Var}(\boldsymbol{d}'\boldsymbol{y} - \boldsymbol{\ell}'\boldsymbol{y}) = \text{Var}((\boldsymbol{d}' - \boldsymbol{\ell}')\boldsymbol{y}) = \text{Var}((\boldsymbol{d} - \boldsymbol{\ell})'\boldsymbol{y}) \\
&= (\boldsymbol{d} - \boldsymbol{\ell})'\text{Var}(\boldsymbol{y})(\boldsymbol{d} - \boldsymbol{\ell}) = (\boldsymbol{d} - \boldsymbol{\ell})'(\sigma^2\boldsymbol{I})(\boldsymbol{d} - \boldsymbol{\ell}) \\
&= \sigma^2(\boldsymbol{d} - \boldsymbol{\ell})'\boldsymbol{I}(\boldsymbol{d} - \boldsymbol{\ell}) = \sigma^2(\boldsymbol{d} - \boldsymbol{\ell})'(\boldsymbol{d} - \boldsymbol{\ell}) > 0 \text{ by (1).}
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(\boldsymbol{d}'\boldsymbol{y} - \boldsymbol{c}'\hat{\boldsymbol{\beta}}, \boldsymbol{c}'\hat{\boldsymbol{\beta}}) &= \text{Cov}(\boldsymbol{d}'\boldsymbol{y} - \boldsymbol{\ell}'\boldsymbol{y}, \boldsymbol{\ell}'\boldsymbol{y}) = \text{Cov}((\boldsymbol{d} - \boldsymbol{\ell})'\boldsymbol{y}, \boldsymbol{\ell}'\boldsymbol{y}) \\
&= (\boldsymbol{d} - \boldsymbol{\ell})'\text{Var}(\boldsymbol{y})\boldsymbol{\ell} = \sigma^2(\boldsymbol{d} - \boldsymbol{\ell})'\boldsymbol{\ell} \\
&= \sigma^2(\boldsymbol{d} - \boldsymbol{\ell})'\boldsymbol{X}[(\boldsymbol{X}'\boldsymbol{X})^-]'\boldsymbol{c} = 0 \text{ by (2).}
\end{aligned}$$

Proof of the Gauss-Markov Theorem

It follows that

$$\begin{aligned}\text{Var}(\boldsymbol{d}'\boldsymbol{y}) &= \text{Var}(\boldsymbol{d}'\boldsymbol{y} - \boldsymbol{c}'\hat{\boldsymbol{\beta}}) + \text{Var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) \\ &> \text{Var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}). \qquad \square\end{aligned}$$