THE GAUSS–MARKOV THEOREM

The Gauss–Markov theorem asserts that the ordinary least-squares estimator $\hat{\beta} = (X'X)^{-1}X'y$ of the parameter β in the classical linear regression model $(y; X\beta, \sigma^2 I)$ is the unbiased linear estimator of least dispersion. This dispersion is usually characterised in terms of the variance of an arbitrary linear combination of the elements of $\hat{\beta}$, although it may also be characterised in terms of the determinant of the dispersion matrix $D(\hat{\beta})$. Thus,

The Gauss-Markov Theorem. *If $\hat{\beta}$ is the ordinary least-squares estimator of β in the classical linear regression model $(y; X\beta, \sigma^2 I)$, and if β^* is any other linear unbiased estimator of β, then $V(p'\beta^*) \geq V(p'\hat{\beta})$, where p is any constant vector of the appropriate order.*

Proof. Since $\beta^* = By$ is an unbiased estimator, it follows that $E(\beta^*) = BE(y) = BX\beta = \beta$, which implies that $BX = I$. Now let us write $B = (X'X)^{-1}X' + G$. Then $BX = I$ implies that $GX = 0$. It follows that

$$
\begin{aligned}
D(\beta^*) &= BD(y)B' \\
&= \sigma^2\big\{(X'X)^{-1}X' + G\big\}\big\{X(X'X)^{-1} + G'\big\} \\
&= \sigma^2(X'X)^{-1} + \sigma^2 GG' \\
&= D(\hat{\beta}) + \sigma^2 GG'.
\end{aligned} \tag{1}
$$

Therefore, for any constant vector p of order k, there is the identity

$$
\begin{aligned}
V(p'\beta^*) &= p'D(\hat{\beta})p + \sigma^2 p'GG'p \\
&\geq p'D(\hat{\beta})q = V(p'\hat{\beta});
\end{aligned} \tag{2}
$$

and thus the inequality $V(p'\beta^*) \geq V(p'\hat{\beta})$ is established.

The tactic of taking arbitrary linear combinations of the elements of $\hat{\beta}$ is to avoid the difficulty inherent in the fact that $\hat{\beta}$ is a vector quantity for which there is no uniquely defined measure of dispersion. An alternative approach, which is not much favoured, is to use the generalised variance that is provided by the determinant of the $D(\beta)$. A version of the Gauss–Markov Theorem that uses this measure can also be proved.

It is worthwhile to consider an alternative statement and proof of the theorem, which also considers the variance of an arbitrary linear combination of the elements of $\hat{\beta}$.

The Gauss–Markov Theorem. (An Alternative Statement). *Let $q'y$ be a linear estimator of the scalar function $p'\beta$ of the regression parameters in the model $(y; X\beta, \sigma^2 I)$. Then $q'y$ is an unbiased estimator, such that $E(q'y) = q'E(y) = q'X\beta = p'\beta$ for all β, if and only if $q'X = p'$. Moreover, $q'y$ has the minimum variance in the class of all unbiased linear estimators if and only if*

$$
q'y = q'X(X'X)^{-1}X'y = p'(X'X)^{-1}X'y. \tag{3}
$$

Therefore, since p is arbitrary, it can be said that $\hat{\beta} = (X'X)^{-1}X'y$ is the minimum variance unbiased linear estimator of β.

Proof. It is obvious that $q'X = p'$ is the necessary and sufficient condition for $q'y$ to be an unbiased estimator of $p'\beta$. To find the unbiased estimator of minimum variance, consider the following criterion:

$$\begin{aligned} &\text{Minimise} \quad V(q'y) = q'D(y)q = \sigma^2 q'q \\ &\text{Subject to} \quad E(q'y) = p'\beta \quad \text{or, equivalently} \quad X'q = p. \end{aligned} \tag{4}$$

To evaluate the criterion, we form the following Lagrangean expression:

$$L = q'q - 2\lambda'(X'q - p). \tag{5}$$

Differentiating with respect to q and setting the results to zero gives, after some rearrangement, the condition

$$q' = \lambda'X'. \tag{6}$$

Postmultiplying by X gives $q'X = p' = \lambda'X'X$, whence

$$\lambda' = q'X(X'X)^{-1} = p'(X'X)^{-1}. \tag{7}$$

On postmultiplying (6) by y and on substituting the expression for λ', we get

$$\begin{aligned} q'y &= q'X(X'X)^{-1}X'y \\ &= p'(X'X)^{-1}X'y = p'\hat{\beta}, \end{aligned} \tag{8}$$

which proves the second part of the theorem.