# The Gauss-Markov Theorem

Raymond Han

14.32/14.320 Recitation 6

## The Gauss-Markov Theorem

The Gauss-Markov Theorem states that under classical assumptions, the OLS estimator, $\hat{\beta}$ has the lowest sampling variance among all unbiased estimators of the regression slope $\beta$ that are linear, i.e. can be written as a linear combination of the $Y_i$. Because of this, we say that OLS is BLUE (which stands for best linear unbiased estimator).

Recall our classical regression assumptions:

1. The CEF of $Y_i$ given $X_i$ is linear, in which case we can write:

$$E\left[Y_i \mid X_i\right] = \alpha + \beta X_i,$$

   where $\alpha$ and $\beta$ are as defined above

2. Define the regression residual $\varepsilon_i = Y_i - [\alpha + \beta X_i]$. Classicists assume:
   (a) $E\left[\varepsilon_i \varepsilon_j\right] = 0; i \neq j$ (random sampling)
   (b) $E\left[\varepsilon_i^2 \mid X_i\right] = E\left[\varepsilon_i^2\right] = \sigma_\varepsilon^2$ (homoskedasticity)
   (c) $\varepsilon_i$ is normally distributed (normality)

3. $X_i$ is fixed in repeated samples

Our proof of the Gauss-Markov Theorem relies only on three of these assumptions: random sampling, homoskedasticity, and fixed $X_i$. That is, OLS is BLUE even when the CEF is nonlinear, or $\varepsilon_i$ is not distributed normally. (Although we don't do so here, it's also possible to show OLS is BLUE when $X_i$ is random under the additional assumption that the regression residual is conditional mean zero, i.e. $\mathrm{E}[\varepsilon_i|X_i] = 0$.)

## Proof

Consider a regression of the form $Y_i = \beta X_i + \alpha + \varepsilon_i$.

We want to show that $\mathrm{Var}\left(\hat{\beta}^{OLS}\right) \leq \mathrm{Var}\left(\hat{\beta}^{alt}\right)$, where $\hat{\beta}^{alt}$ is any alternative linear unbiased estimator of the slope $\beta$. Recall that we can write the OLS estimator as a linear combination of the $Y_i$, i.e. $\hat{\beta}^{OLS} = \sum_i a_i Y_i$, with weights $a_i \equiv \tilde{X}_i / \sum_i \tilde{X}_i^2$, where $\tilde{X}_i = X_i - \overline{X}_i$.

The sampling variance of OLS is then

$$\begin{aligned}
\mathrm{Var}\left(\hat{\beta}^{OLS}\right) &= \mathrm{Var}\left(\sum_i a_i Y_i\right) \\
&= \mathrm{Var}\left(\sum_i a_i(\beta X_i + \alpha + \varepsilon_i)\right) \\
&= \mathrm{Var}\left(\sum_i a_i \varepsilon_i\right) && \text{(fixed } X_i\text{)} \\
&= \sum_i \mathrm{Var}(a_i \varepsilon_i) && \text{(random sampling)} \\
&= \sigma^2 \sum_i a_i^2 && \text{(homoskedasticity)}
\end{aligned}$$

Since $\hat{\beta}^{alt}$ is linear, we can write it as $\hat{\beta}^{alt} = \sum_i b_i Y_i$, where $b_i$ is some other set of weights. Then let $c_i \equiv b_i - a_i$, the difference between the alternative weights and the OLS weights. We can now write the sampling variance of $\hat{\beta}^{alt}$ as $\mathrm{Var}(\sum_i b_i Y_i) = \mathrm{Var}(\sum_i (a_i + c_i) Y_i)$. Then by the same argument as above,

$$\begin{aligned}
\mathrm{Var}\left(\hat{\beta}^{alt}\right) &= \sigma^2 \sum_i (a_i + c_i)^2 \\
&= \sigma^2 \sum_i (a_i^2 + c_i^2 + 2a_i c_i) \\
&= \mathrm{Var}\left(\hat{\beta}^{OLS}\right) + \sigma^2 \sum_i c_i^2 + 2\sigma^2 \sum_i a_i c_i
\end{aligned}$$

The second term is always (weakly) positive since it's proportional to a sum of squares. Therefore, if we can show that $\sum_i a_i c_i$ is zero (or greater), the proof is complete. The latter must be true if $\hat{\beta}^{alt}$ is unbiased. To see this, note that

$$\begin{aligned}
\mathrm{E}\left[\hat{\beta}^{alt}\right] &= \mathrm{E}\left[\sum_i (a_i + c_i) Y_i\right] \\
&= \beta + \mathrm{E}\left[\sum_i c_i(\beta X_i + \alpha + \epsilon_i)\right] \\
&= \beta + \beta \sum_i c_i X_i + \alpha \sum c_i
\end{aligned}$$

So for $\mathrm{E}\left[\hat{\beta}^{alt}\right]$ to equal $\beta$ for arbitrary $(\alpha, \beta)$, it must be the case that $\sum_i c_i X_i = 0$ and $\sum_i c_i = 0$. As a result of the former,

$$\begin{aligned}
\sum_i a_i c_i &= \sum_i \frac{\tilde{X}_i c_i}{\sum_i \tilde{X}_i^2} \\
&= \sum_i \frac{(X_i - \overline{X}) c_i}{\sum_i (X_i - \overline{X})^2} \\
&= \frac{1}{\sum_i \tilde{X}_i^2} \left( \sum_i c_i X_i - \overline{X} \sum_i c_i \right) \\
&= 0.
\end{aligned}$$

Applying this fact to our expression for the sampling variance of $\hat{\beta}^{alt}$ implies

$$\begin{aligned}
\mathrm{Var}\left(\hat{\beta}^{alt}\right) &= \mathrm{Var}\left(\hat{\beta}^{OLS}\right) + \sigma^2 \sum_i c_i^2 \\
&\geq \mathrm{Var}\left(\hat{\beta}^{OLS}\right).
\end{aligned}$$