A Simple Gauss-Markov Theorem

THOMAS WIEMANN
University of Chicago

May 18, 2021

Motivation of the Gauss-Markov Theorem

Under assumptions, the Gauss-Markov Theorem guides the choice of estimator among a class of estimators.

Assumptions:

- The data is generated by a linear model.
- The explanatory variables are exogeneous.
- The disturbances are homoskedastic.
- The observations are *iid*.

Scope:

- The estimator is linear.
- The estimator is unbiased.

Linear Estimator

Definition (Linear Estimator)

An estimator $\hat{\theta}$ is a *linear estimator* if it is a linear function of the dependent variable – that is,

$$\hat{\theta} = \sum_{i=1}^{N} c_i y_i. \tag{1}$$

Typically, the constants c_i are functions of the explanatory variables x_i.

Example (Simple Least Squares)

The simple least squares estimators for β_0 and β_1 are linear:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})y_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \sum_{i=1}^{N}\left(\frac{(x_i - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right) y_i, \tag{2}$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1 = \sum_{i=1}^{N}\left(\frac{1}{N} - \frac{\bar{x}(x_i - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right) y_i \tag{3}$$

A Simple Gauss-Markov Theorem

Theorem (Simple Gauss-Markov)

Consider the linear model given by

$$Y = X\beta_1 + U, \tag{4}$$

where $U \sim P$ are random disturbances such that $E[U|X] = 0$ and $Var(U|X) = \sigma^2$. Suppose we observe a sample $(y_i, x_i) \overset{iid}{\sim} (Y, X)$ of size N. Then the least squares estimator for β_1, given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}, \tag{5}$$

is the estimator with the lowest conditional variance among all linear unbiased estimators. Formally,

$$Var\left(\hat{\beta}_1 | \{x_i\}_{i=1}^N\right) \leq Var\left(\tilde{\beta}_1 | \{x_i\}_{i=1}^N\right), \tag{6}$$

$\forall \tilde{\beta}_1 \in \left\{ b \middle| \, b = \sum_{i=1}^N g_i y_i, \; E\left[b | \{x_i\}_{i=1}^N \right] = \beta_1 \right\}$.

A Simple Gauss-Markov Theorem (Contd.)

Proof.

1. $\tilde{\beta}_1 = \sum g_i y_i$, $\hat{\beta}_1 = \sum c_i y_i$ where $c_i = \frac{x_i}{\sum x_i^2}$

$w_i = c_i - g_i \Rightarrow g_i = c_i - w_i$

$$\tilde{\beta}_1 = \sum (c_i - w_i) y_i \overset{\text{Linear Model}}{=} \underbrace{\sum c_i x_i}_{=1} \beta_1 - \sum w_i x_i \beta_1 + \sum (c_i - w_i) u_i = \beta_1 - \sum w_i x_i \beta_1 + \sum (c_i - w_i) u_i$$

2. $$E[\tilde{\beta}_1 | X] = \beta_1 - \sum w_i x_i \beta_1 + \sum (c_i - w_i) \underbrace{E[u_i | X]}_{=0} = \beta_1 - (\sum w_i x_i) \beta_1 \overset{\text{Unbiased}}{=} \beta_1$$

$\Rightarrow \sum w_i x_i = 0$

3. $$Var(\tilde{\beta}_1 | X) = E[(\sum (c_i - w_i) u_i)^2 | X] \overset{iid}{=} E[\sum (c_i - w_i)^2 u_i^2 | X] = \sum (c_i^2 - 2 c_i w_i + w_i^2) \underbrace{E[u_i^2 | X]}_{= \sigma^2}$$

$$= \frac{(\sum x_i^2)}{(\sum x_i^2)^2} \sigma^2 - 2 \underbrace{\left(\frac{\sum w_i x_i}{\sum x_i^2} \right)}_{=0} \sigma^2 + \sum w_i^2 \sigma^2 = \underbrace{\left(\frac{1}{\sum x_i^2} \right) \sigma^2}_{= Var(\hat{\beta}_1 | X)} + \underbrace{\sum w_i^2 \sigma^2}_{\geq 0}$$

□

On the Choice of Estimators

The Gauss-Markov Theorem has limited practical relevance in modern data analysis. It's assumptions and scope are too restrictive. Yet, it delivers crucial insight on the choice of estimators.

The idea of the Gauss-Markov Theorem is to guide the choice of estimator under the desideratum of low mean-squared error:

$$E[(\hat{\theta} - \theta)^2] = E\left[\left(\hat{\theta} - E[\hat{\theta}]\right)^2\right] + \left(E[\hat{\theta}] - \theta\right)^2 \tag{7}$$

This remains incredibly relevant and is the motivation for highly active research in econometrics, statistics, and machine learning.