

# Proof of the Gauss Markov Theorem

By: Philip Nye

## Contents

<b>1</b>	<b>Gauss Markov Assumptions</b>	
1.1	Assumptions . . . . .	
1.2	Proof that Random Sampling Implies Zero Serial Correlation with Cross-Sectional Data . . . . .	
<b>2</b>	<b>Gauss Markov Theorem Proofs</b>	
2.1	Proof of Unbiasedness . . . . .	
2.2	Proof of Greatest Efficiency . . . . .	

# 1 Gauss Markov Assumptions

## 1.1 Assumptions

- Assumption 1: The population model is linear in parameters such that

$$y_i = \alpha + \beta_1 x_{i,1} + \dots + \beta_m x_{i,m} + u_i$$

- Assumption 2: The sample data  $\{x_{i,1}, \dots, x_{i,m}, y_i\}$  is a random sample.
- Assumption 3: The error term has a zero conditional mean such that

$$E[u|x_{i,1}, \dots, x_{i,m}] = 0$$

- Assumption 4: The error term is homoskedastic such that

$$Var(u|x_{i,1}, \dots, x_{i,m}) = \sigma_u^2$$

- Assumption 5: None of the regressors exhibit perfect collinearity with one another.
- Assumption 6: There exists no serial correlation between the error terms such that

$$cov(u_i, u_j) = 0 \forall i \neq j$$

Note: Assumption 2 is sufficient to satisfy assumption 6 in the case of cross-sectional data (see below for proof).

## 1.2 Proof that Random Sampling Implies Zero Serial Correlation with Cross-Sectional Data

Under assumption 2, we know the sample data is independent and identically distributed (i.i.d), meaning drawing one observation does not make drawing another observation any more or less likely and that the observations come from the same distribution. Thus, for some observation  $i \neq j$ ,

$$cov(y_i, y_j) = E[y_i y_j] - E[y_i]E[y_j] = E[y_i]E[y_j] - E[y_i]E[y_j] = 0$$

Similarly, by our first assumption that  $y_i = \alpha + \beta_1 x_{i,1} + \dots + \beta_m x_{i,m} + u_i$

$$\begin{aligned} 0 &= cov(y_i, y_j) = cov(\alpha + \beta_1 x_{i,1} + \dots + \beta_m x_{i,m} + u_i, \alpha + \beta_1 x_{j,1} + \dots + \beta_m x_{j,m} + u_j) \\ &= \sum_{s=1}^m \sum_{t=1}^m \beta_s \beta_t cov(x_{i,s}, x_{j,t}) + \sum_{s=1}^m \beta_s cov(x_{i,s}, u_j) + \sum_{t=1}^m \beta_t cov(x_{j,t}, u_i) + cov(u_i, u_j) \end{aligned}$$

Using the fact that the sample is i.i.d.,

$$\sum_{s=1}^m \sum_{t=1}^m \beta_s \beta_t cov(x_{i,s}, x_{j,t}) = 0$$

Also, by the exogeneity assumption,

$$\sum_{s=1}^m \beta_s cov(x_{i,s}, u_j) + \sum_{t=1}^m \beta_t cov(x_{j,t}, u_i) = 0$$

Leaving us with

$$cov(u_i, u_j) = 0$$

## 2 Gauss Markov Theorem Proofs

### 2.1 Proof of Unbiasedness

Assuming a simple bivariate population model in which

$$y_i = \alpha + \beta^P x_i + u_i$$

we've shown, using first order conditions,

$$\hat{\beta}_{LS} = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sum_{i=1}^n v_i y_i$$

where  $v_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$ .

Note:  $\sum_{i=1}^n v_i = 0$  since the numerator of  $v_i$  comes out to  $\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n\bar{x} - n\bar{x} = 0$

However,  $\sum_{i=1}^n v_i^2 = \left( \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^2 = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} > 0$

Using these results, we know

$$\begin{aligned} E[\hat{\beta}_{LS}] &= E\left[\sum_{i=1}^n v_i y_i\right] = \sum_{i=1}^n E[v_i y_i] = \sum_{i=1}^n E[\alpha v_i + \beta^P x_i v_i + u_i v_i] \\ &= \alpha \sum_{i=1}^n E[v_i] + \beta^P \sum_{i=1}^n E[x_i v_i] + \sum_{i=1}^n E[u_i v_i] \\ &= \alpha \sum_{i=1}^n 0 + \beta^P \sum_{i=1}^n E\left[\frac{(x_i - \bar{x}) x_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + \sum_{i=1}^n E[u_i v_i] \\ &= \beta^P + \sum_{i=1}^n E[u_i v_i] \end{aligned}$$

Using the exogeneity assumption, we know  $E[u_i v_i] = E[u_i] E[v_i] = 0 \cdot E[v_i] = 0$ , and thus

$$E[\hat{\beta}_{LS}] = \beta^P + \sum_{i=1}^n 0 = \beta^P$$

$\therefore$  Under the Gauss Markov assumptions, we know the least squares estimator  $\hat{\beta}_{LS}$  is unbiased as  $E[\hat{\beta}_{LS}] = \beta^P$ .

## 2.2 Proof of Greatest Efficiency

$$\begin{aligned}\hat{\beta}_{LS} &= \sum_{i=1}^n y_i v_i = \sum_{i=1}^n \alpha v_i + \beta^P x_i v_i + u_i v_i = \beta^P + \sum_{i=1}^n u_i v_i \\ &\rightarrow \text{Var}(\hat{\beta}_{LS}) = \text{Var}(\beta^P + \sum_{i=1}^n u_i v_i) = \text{Var}(\sum_{i=1}^n u_i v_i)\end{aligned}$$

Using the exogeneity and zero serial correlation assumptions,

$$\text{Var}(\sum_{i=1}^n u_i v_i) = \sum_{i=1}^n \text{Var}(u_i v_i) = \sum_{i=1}^n v_i^2 \text{Var}(u_i)$$

Under homoskedasticity,  $\text{Var}(u_i) = \sigma_u^2$ , a constant, so

$$\sum_{i=1}^n v_i^2 \text{Var}(u_i) = \sigma_u^2 \sum_{i=1}^n v_i^2 = \frac{\sigma_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2} > 0$$

Now, take another arbitrary unbiased estimator of  $\beta^P, \tilde{\beta}$ . By the linearity constraint,

$$\tilde{\beta} = \sum_{i=1}^n b_i y_i = \alpha \sum_{i=1}^n b_i + \beta^P \sum_{i=1}^n b_i x_i + \sum_{i=1}^n b_i u_i$$

Also, under the unbiased constraint,

$$E[\tilde{\beta}] = \beta^P$$

So,

$$E[\tilde{\beta}] = E[\alpha \sum_{i=1}^n b_i + \beta^P \sum_{i=1}^n b_i x_i + \sum_{i=1}^n b_i u_i] = \alpha \sum_{i=1}^n b_i + \beta^P \sum_{i=1}^n b_i x_i$$

as  $b_i$  implicitly contains  $x_i$ , which, by assumption, is exogenous from the error term  $u_i$ .

Thus, it must be the case that

$$\sum_{i=1}^n b_i = 0 \tag{1}$$

$$\sum_{i=1}^n b_i x_i = 1 \tag{2}$$

$$\begin{aligned}&\rightarrow \text{Var}(\tilde{\beta}) = \text{Var}(\sum_{i=1}^n b_i y_i) = \text{Var}(\sum_{i=1}^n \alpha b_i + \beta^P b_i x_i + b_i u_i) \\ &= \text{Var}(\alpha \sum_{i=1}^n b_i + \beta^P \sum_{i=1}^n b_i x_i + \sum_{i=1}^n b_i u_i) = \text{Var}(\beta^P + \sum_{i=1}^n b_i u_i) = \text{Var}(\sum_{i=1}^n b_i u_i)\end{aligned}$$

Under the homoskedasticity and random sampling assumptions,

$$\text{Var}(\sum_{i=1}^n b_i u_i) = \sum_{i=1}^n \text{Var}(b_i u_i) = \sum_{i=1}^n b_i^2 \text{Var}(u_i) = \sigma_u^2 \sum_{i=1}^n b_i^2$$

Now, suppose  $b_i = v_i + c_i$  where  $v_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$  and  $c_i$  equals some arbitrary difference from the least squares weights. Thus,

$$\begin{aligned} Var(\tilde{\beta}) &= \sigma_u^2 \sum_{i=1}^n b_i^2 = \sigma_u^2 \sum_{i=1}^n v_i^2 + 2\sigma_u^2 \sum_{i=1}^n v_i c_i + \sigma_u^2 \sum_{i=1}^n c_i^2 \\ &= Var(\hat{\beta}_{LS}) + 2\sigma_u^2 \sum_{i=1}^n v_i c_i + \sigma_u^2 \sum_{i=1}^n c_i^2 \end{aligned}$$

Remembering the conditions that  $\sum_{i=1}^n v_i = 0$  and  $\sum_{i=1}^n b_i = 0$ .

$$\begin{aligned} \sum_{i=1}^n b_i &= \sum_{i=1}^n v_i + \sum_{i=1}^n c_i = 0 + \sum_{i=1}^n c_i = \sum_{i=1}^n c_i \\ &\rightarrow \sum_{i=1}^n c_i = 0 \end{aligned}$$

Similarly, since  $\sum_{i=1}^n v_i x_i = 1$  and  $\sum_{i=1}^n b_i x_i = 1$ ,

$$\begin{aligned} \sum_{i=1}^n b_i x_i &= 1 = \sum_{i=1}^n (v_i + c_i) x_i = \sum_{i=1}^n v_i x_i + \sum_{i=1}^n c_i x_i = 1 + \sum_{i=1}^n c_i x_i \\ &\rightarrow \sum_{i=1}^n c_i x_i = 0 \end{aligned}$$

Finally, returning to the variance of  $\tilde{\beta}$ ,

$$Var(\tilde{\beta}) = Var(\hat{\beta}_{LS}) + 2\sigma_u^2 \sum_{i=1}^n v_i c_i + \sigma_u^2 \sum_{i=1}^n c_i^2$$

where

$$2\sigma_u^2 \sum_{i=1}^n v_i c_i = 2\sigma_u^2 \sum_{i=1}^n \frac{(x_i - \bar{x})c_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = 2\sigma_u^2 \frac{\sum_{i=1}^n c_i x_i - \bar{x} \sum_{i=1}^n c_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = 0$$

since  $\sum_{i=1}^n c_i = 0$  and  $\sum_{i=1}^n c_i x_i = 0$ .

Thus,

$$Var(\tilde{\beta}) = Var(\hat{\beta}_{LS}) + \sigma_u^2 \sum_{i=1}^n c_i^2 \geq Var(\hat{\beta}_{LS})$$

as  $\sigma_u^2 \geq 0$  and  $\sum_{i=1}^n c_i^2 \geq 0$ .

$\therefore \hat{\beta}_{LS}$  is the best linear unbiased estimator of the population parameter  $\beta^P$  because any other linear unbiased estimator has at least as much variance as the least square estimator under the Gauss Markov assumptions.