

Gauss Markov Assumptions

Assumptions

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

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i$$

- Assumption 2: The sample data $\{x_{i1}, \dots, x_{ik}, y_i\}$ is a random sample.
- Assumption 3: The error term has a zero conditional mean such that

$$E[u|x_{i1}, \dots, x_{ik}] = 0$$

- Assumption 4: The error term is homoskedastic such that

$$Var(u|x_{i1}, \dots, x_{ik}) = \sigma_u^2$$

- 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).

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 $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i$

$$0 = cov(y_i, y_j) = cov(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i, \beta_0 + \beta_1 x_{j1} + \dots + \beta_k x_{jk} + u_j)$$

$$= \sum_{s=1}^k \sum_{t=1}^k \beta_s \beta_t cov(x_{is}, x_{jt}) + \sum_{s=1}^k \beta_s cov(x_{is}, u_j) + \sum_{t=1}^m \beta_t cov(x_{jt}, u_i) + cov(u_i, u_j)$$

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

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

Also, by the exogeneity assumption,

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

Leaving us with

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

Gauss Markov Theorem Proof

Derivation of the Slope Estimators

To derive the slope estimators, start by regressing x_ℓ on all of the other regressors in the model, which takes the form

$$x_\ell = \gamma_0 + \gamma_1 x_1 + \cdots + \gamma_{\ell-1} x_{\ell-1} + \gamma_{\ell+1} x_{\ell+1} + \cdots + \gamma_k x_k + r_{i\ell}$$

If we denote the residuals of the model as $\hat{r}_{i\ell}$ and the predicted values as $\hat{x}_{i\ell}$, then

$$x_{i\ell} = \hat{x}_{i\ell} + \hat{r}_{i\ell}$$

Plugging this derivation into the $(\ell + 1)^{th}$ first order condition:

$$\begin{aligned} \frac{\partial SSR}{\partial \beta_\ell} &= \sum_{i=1}^n -2x_{i\ell} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0 \\ &\rightarrow \sum_{i=1}^n (\hat{x}_{i\ell} + \hat{r}_{i\ell}) (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0 \\ &\rightarrow \sum_{i=1}^n \hat{x}_{i\ell} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) + \sum_{i=1}^n \hat{r}_{i\ell} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0 \\ &\rightarrow \sum_{i=1}^n \hat{x}_{i\ell} \hat{u}_i + \sum_{i=1}^n \hat{r}_{i\ell} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0 \end{aligned}$$

Because $\hat{x}_{i\ell}$ is simply a linear combination of all the other regressors in the model (i.e., $\hat{x}_{i\ell} = \hat{\gamma}_0 + \hat{\gamma}_1 x_{i1} + \cdots + \hat{\gamma}_{\ell-1} x_{i\ell-1} + \gamma_{\ell+1} x_{i\ell+1} + \cdots + \gamma_k x_{ik}$), it follows that $\sum_{i=1}^n \hat{x}_{i\ell} \hat{u}_i = 0$.

$$\begin{aligned} &\rightarrow \sum_{i=1}^n \hat{r}_{i\ell} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0 \\ &\rightarrow \sum_{i=1}^n \hat{r}_{i\ell} (y_i - \hat{\beta}_\ell x_{i\ell}) + \sum_{i=1}^n \hat{r}_{i\ell} (-\hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_{\ell-1} x_{i\ell-1} - \hat{\beta}_{\ell+1} x_{i\ell+1} - \cdots - \hat{\beta}_k x_{ik}) = 0 \end{aligned}$$

Because $\hat{r}_{i\ell}$ are the residuals from regressing x_ℓ on all the other regressors, $\sum_{i=1}^n x_{ij} \hat{r}_{i\ell} = 0 \forall j \neq \ell$.

$$\begin{aligned} &\rightarrow \sum_{i=1}^n \hat{r}_{i\ell} (y_i - \hat{\beta}_\ell x_{i\ell}) = 0 \\ &\rightarrow \sum_{i=1}^n \hat{r}_{i\ell} (y_i - \hat{\beta}_\ell (\hat{x}_{i\ell} + \hat{r}_{i\ell})) = 0 \\ &\rightarrow \sum_{i=1}^n \hat{r}_{i\ell} (y_i - \hat{\beta}_\ell \hat{r}_{i\ell}) - \hat{\beta}_\ell \sum_{i=1}^n \hat{r}_{i\ell} \hat{x}_{i\ell} = 0 \end{aligned}$$

Since $\sum_{i=1}^n \hat{r}_{i\ell} \hat{x}_{i\ell} = 0$ (This can be thought of as $\sum_{i=1}^n \hat{r}_{i\ell} \hat{x}_{i\ell} = 0 \rightarrow \sum_{i=1}^n \hat{u}_i \hat{y}_i = 0 \rightarrow \hat{\beta}_0 \sum_{i=1}^n \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^n x_{i1} \hat{u}_i = 0 + 0 = 0$ in the SLR case),

$$\begin{aligned} &\rightarrow \sum_{i=1}^n \hat{r}_{i\ell} (y_i - \hat{\beta}_\ell \hat{r}_{i\ell}) = 0 \\ &\rightarrow \hat{\beta}_\ell \sum_{i=1}^n \hat{r}_{i\ell}^2 = \sum_{i=1}^n \hat{r}_{i\ell} y_i \\ &\rightarrow \hat{\beta}_\ell = \frac{\sum_{i=1}^n \hat{r}_{i\ell} y_i}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \end{aligned}$$

Proof of Unbiasedness

Using the previous derivation,

$$\begin{aligned}
\hat{\beta}_\ell &= \frac{\sum_{i=1}^n \hat{r}_{i\ell} y_i}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \\
&= \frac{\sum_{i=1}^n \hat{r}_{i\ell} (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i)}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \\
&= \frac{\beta_0 \sum_{i=1}^n \hat{r}_{i\ell} + \beta_1 \sum_{i=1}^n x_{i1} \hat{r}_{i\ell} + \cdots + \beta_k \sum_{i=1}^n x_{ik} \hat{r}_{i\ell} + \sum_{i=1}^n u_i \hat{r}_{i\ell}}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \\
&= \frac{\beta_\ell \sum_{i=1}^n x_{i\ell} \hat{r}_{i\ell} + \sum_{i=1}^n u_i \hat{r}_{i\ell}}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \\
&= \frac{\beta_\ell \sum_{i=1}^n (\hat{x}_{i\ell} + \hat{r}_{i\ell}) \hat{r}_{i\ell} + \sum_{i=1}^n u_i \hat{r}_{i\ell}}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \\
&= \frac{\beta_\ell \sum_{i=1}^n \hat{x}_{i\ell} \hat{r}_{i\ell} + \beta_\ell \sum_{i=1}^n \hat{r}_{i\ell}^2 + \sum_{i=1}^n u_i \hat{r}_{i\ell}}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \\
&= \frac{\beta_\ell \sum_{i=1}^n (\hat{\gamma}_0 + \hat{\gamma}_1 x_{i1} + \cdots + \hat{\gamma}_{\ell-1} x_{i\ell-1} + \gamma_{\ell+1} x_{i\ell+1} + \cdots + \gamma_k x_{ik}) \hat{r}_{i\ell} + \beta_\ell \sum_{i=1}^n \hat{r}_{i\ell}^2 + \sum_{i=1}^n \hat{r}_{i\ell} u_i}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \\
&= \frac{0 + \beta_\ell \sum_{i=1}^n \hat{r}_{i\ell}^2 + \sum_{i=1}^n \hat{r}_{i\ell} u_i}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \\
&= \beta_\ell + \frac{\sum_{i=1}^n \hat{r}_{i\ell} u_i}{\sum_{i=1}^n \hat{r}_{i\ell}^2}
\end{aligned}$$

Taking the expectation of this value (conditional on x):

$$\begin{aligned}
E[\hat{\beta}_\ell | x] &= E\left[\beta_\ell + \frac{\sum_{i=1}^n \hat{r}_{i\ell} u_i}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \middle| x\right] \\
&= \beta_\ell + \frac{1}{\sum_{i=1}^n \hat{r}_{i\ell}^2} E\left[\sum_{i=1}^n \hat{r}_{i\ell} u_i \middle| x\right] \\
&= \beta_\ell + \frac{1}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \sum_{i=1}^n E[\hat{r}_{i\ell} u_i | x] \\
&= \beta_\ell + \frac{1}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \sum_{i=1}^n \hat{r}_{i\ell} E[u_i | x] \\
&= \beta_\ell + \frac{1}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \sum_{i=1}^n \hat{r}_{i\ell} \cdot 0 = \beta_\ell
\end{aligned}$$

Sampling Variance of OLS Estimators

Again using the previous derivation that

$$\hat{\beta}_\ell = \beta_\ell + \frac{\sum_{i=1}^n \hat{r}_{i\ell} u_i}{\sum_{i=1}^n \hat{r}_{i\ell}^2}$$

We can derive the sampling variance of the OLS estimator $\hat{\beta}_\ell$ as:

$$\begin{aligned} \text{Var}(\hat{\beta}_\ell | x) &= \text{Var}\left(\beta_\ell + \frac{\sum_{i=1}^n \hat{r}_{i\ell} u_i}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \middle| x\right) \\ &= \frac{1}{\left(\sum_{i=1}^n \hat{r}_{i\ell}^2\right)^2} \text{Var}\left(\sum_{i=1}^n \hat{r}_{i\ell} u_i \middle| x\right) \\ &= \frac{1}{\left(\sum_{i=1}^n \hat{r}_{i\ell}^2\right)^2} \sum_{i=1}^n \text{Var}(\hat{r}_{i\ell} u_i | x) \\ &= \frac{1}{\left(\sum_{i=1}^n \hat{r}_{i\ell}^2\right)^2} \sum_{i=1}^n \hat{r}_{i\ell}^2 \text{Var}(u_i | x) \\ &= \frac{1}{\left(\sum_{i=1}^n \hat{r}_{i\ell}^2\right)^2} \sum_{i=1}^n \hat{r}_{i\ell}^2 \sigma_u^2 \\ &= \frac{\sigma_u^2 \sum_{i=1}^n \hat{r}_{i\ell}^2}{\left(\sum_{i=1}^n \hat{r}_{i\ell}^2\right)^2} \\ &= \frac{\sigma_u^2}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \\ &= \frac{\sigma_u^2}{SST_{x_\ell} - SSE_{x_\ell}} \\ &= \frac{\sigma_u^2}{SST_{x_\ell} (1 - R_\ell^2)} \end{aligned}$$

where R_ℓ^2 is the R -squared from regressing x_ℓ on all other independent variables.

Proof OLS Estimators are BLUE Under the Gauss-Markov Assumptions

Restricting the scope to linear unbiased estimators, we can define an estimator

$$\tilde{\beta}_\ell = \sum_{i=1}^n w_{i\ell} y_i,$$

where each $w_{i\ell}$ can be a function of the sample values of all the independent variables and

$$E \left[\tilde{\beta}_\ell \middle| x \right] = \beta_\ell$$

$$E \left[\sum_{i=1}^n w_{i\ell} y_i \middle| x \right] = \beta_\ell$$

$$E \left[\sum_{i=1}^n w_{i\ell} (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i) \middle| x \right] = \beta_\ell$$

$$\beta_0 \sum_{i=1}^n w_{i\ell} + \beta_1 \sum_{i=1}^n x_{i1} w_{i\ell} + \cdots + \beta_k \sum_{i=1}^n x_{ik} w_{i\ell} + \sum_{i=1}^n w_{i\ell} E[u_i | x] = \beta_\ell$$

$$E \left[\beta_0 \sum_{i=1}^n w_{i\ell} + \beta_1 \sum_{i=1}^n x_{i1} w_{i\ell} + \cdots + \beta_\ell \sum_{i=1}^n x_{i\ell} w_{i\ell} + \cdots + \beta_k \sum_{i=1}^n x_{ik} w_{i\ell} + \sum_{i=1}^n u_i w_{i\ell} \middle| x \right] = \beta_\ell,$$

where $\sum_{i=1}^n w_{i\ell} E[u_i | x] = 0$ by the zero conditional mean assumption. Thus, it must be the case that

$$\sum_{i=1}^n w_{i\ell} = 0$$

$$\sum_{i=1}^n w_{i\ell} x_{ij} = 0 \quad \forall j \neq \ell$$

$$\sum_{i=1}^n w_{i\ell} x_{i\ell} = 1$$

Using this fact, we can derive the variance of $\tilde{\beta}_\ell$ as

$$\begin{aligned} \text{Var}(\tilde{\beta}_\ell | x) &= \text{Var} \left(\beta_0 \sum_{i=1}^n w_{i\ell} + \beta_1 \sum_{i=1}^n x_{i1} w_{i\ell} + \cdots + \beta_\ell \sum_{i=1}^n x_{i\ell} w_{i\ell} + \cdots + \beta_k \sum_{i=1}^n x_{ik} w_{i\ell} + \sum_{i=1}^n u_i w_{i\ell} \middle| x \right) \\ &= \text{Var} \left(\beta_\ell + \sum_{i=1}^n u_i w_{i\ell} \middle| x \right) \\ &= \sum_{i=1}^n \text{Var}(u_i w_{i\ell} | x) \\ &= \sum_{i=1}^n w_{i\ell}^2 \text{Var}(u_i | x) \\ &= \sigma_u^2 \sum_{i=1}^n w_{i\ell}^2 \end{aligned}$$

Suppose $w_{i\ell} = v_{i\ell} + c_{i\ell}$ where $v_{i\ell} = \frac{\hat{r}_{i\ell}}{\sum_{i=1}^n \hat{r}_{i\ell}^2}$ (the OLS weights) and $c_{i\ell}$ is some arbitrary difference from the least squares weights, then

$$\begin{aligned}
\text{Var}(\tilde{\beta}_\ell | x) &= \sigma_u^2 \sum_{i=1}^n w_{i\ell}^2 = \sigma_u^2 \sum_{i=1}^n (v_{i\ell} + c_{i\ell})^2 \\
&= \sigma_u^2 \sum_{i=1}^n v_{i\ell}^2 + \sigma_u^2 \sum_{i=1}^n c_{i\ell}^2 + 2\sigma_u^2 \sum_{i=1}^n v_{i\ell} c_{i\ell} \\
&= \sigma_u^2 \sum_{i=1}^n \left(\frac{\hat{r}_{i\ell}}{\sum_{i=1}^n \hat{r}_{i\ell}^2} \right)^2 + \sigma_u^2 \sum_{i=1}^n c_{i\ell}^2 + 2\sigma_u^2 \sum_{i=1}^n v_{i\ell} c_{i\ell} \\
&= \sigma_u^2 \frac{\sum_{i=1}^n \hat{r}_{i\ell}^2}{\left(\sum_{i=1}^n \hat{r}_{i\ell}^2 \right)^2} + \sigma_u^2 \sum_{i=1}^n c_{i\ell}^2 + 2\sigma_u^2 \sum_{i=1}^n v_{i\ell} c_{i\ell} \\
&= \frac{\sigma_u^2}{\sum_{i=1}^n \hat{r}_{i\ell}^2} + \sigma_u^2 \sum_{i=1}^n c_{i\ell}^2 + 2\sigma_u^2 \sum_{i=1}^n v_{i\ell} c_{i\ell} \\
&= \text{Var}(\hat{\beta}_\ell) + \sigma_u^2 \sum_{i=1}^n c_{i\ell}^2 + 2\sigma_u^2 \sum_{i=1}^n v_{i\ell} c_{i\ell}
\end{aligned}$$

Hence, to show OLS estimators are BLUE under the G-M assumptions, it remains to show that

$$\begin{aligned}
\text{Var}(\hat{\beta}_\ell) &\leq \text{Var}(\hat{\beta}_\ell) + \sigma_u^2 \sum_{i=1}^n c_{i\ell}^2 + 2\sigma_u^2 \sum_{i=1}^n v_{i\ell} c_{i\ell} \\
&\rightarrow 0 \leq \sigma_u^2 \sum_{i=1}^n c_{i\ell}^2 + 2\sigma_u^2 \sum_{i=1}^n v_{i\ell} c_{i\ell},
\end{aligned}$$

which just requires showing $2\sigma_u^2 \sum_{i=1}^n v_{i\ell} c_{i\ell} \geq 0$ since $\sigma_u^2 \sum_{i=1}^n c_{i\ell}^2 \geq 0$. Finalizing the proof,

$$\begin{aligned}
2\sigma_u^2 \sum_{i=1}^n v_{i\ell} c_{i\ell} &\geq 0 \\
&\rightarrow \sum_{i=1}^n v_{i\ell} c_{i\ell} \geq 0
\end{aligned}$$

Returning to the previous derivations that $\sum_{i=1}^n w_{i\ell} = 0$, $\sum_{i=1}^n w_{i\ell} x_{ij} = 0 \ \forall \ j \neq \ell$, and $\sum_{i=1}^n w_{i\ell} x_{i\ell} = 1$,

$$\begin{aligned}
0 &= \sum_{i=1}^n w_{i\ell} = \sum_{i=1}^n (v_{i\ell} + c_{i\ell}) = \sum_{i=1}^n v_{i\ell} + \sum_{i=1}^n c_{i\ell} = 0 + \sum_{i=1}^n c_{i\ell} \\
&\rightarrow \sum_{i=1}^n c_{i\ell} = 0 \\
0 &= \sum_{i=1}^n w_{i\ell} x_{ij} = \sum_{i=1}^n (v_{i\ell} + c_{i\ell}) x_{ij} = \sum_{i=1}^n v_{i\ell} x_{ij} + \sum_{i=1}^n c_{i\ell} x_{ij} = 0 + \sum_{i=1}^n c_{i\ell} x_{ij} \\
&\rightarrow \sum_{i=1}^n c_{i\ell} x_{ij} = 0 \ \forall \ j \neq \ell
\end{aligned}$$

$$\begin{aligned}
1 &= \sum_{i=1}^n w_{il} x_{il} = \sum_{i=1}^n (v_{il} + c_{il}) x_{il} = \sum_{i=1}^n v_{il} x_{il} + \sum_{i=1}^n c_{il} x_{il} = 1 + \sum_{i=1}^n c_{il} x_{il} \\
&\rightarrow \sum_{i=1}^n c_{il} x_{il} = 0
\end{aligned}$$

Thus,

$$\begin{aligned}
\sum_{i=1}^n v_{il} c_{il} &= \sum_{i=1}^n \frac{\hat{r}_{il} c_{il}}{\sum_{i=1}^n \hat{r}_{il}^2} = \frac{\sum_{i=1}^n \hat{r}_{il} c_{il}}{\sum_{i=1}^n \hat{r}_{il}^2} \\
&= \frac{\sum_{i=1}^n (x_{il} - \hat{x}_{il}) c_{il}}{\sum_{i=1}^n \hat{r}_{il}^2} \\
&= \frac{\sum_{i=1}^n x_{il} c_{il} - \sum_{i=1}^n \hat{x}_{il} c_{il}}{\sum_{i=1}^n \hat{r}_{il}^2} \\
&= \frac{0 - \sum_{i=1}^n (\hat{\gamma}_0 + \hat{\gamma}_1 x_{i1} + \cdots + \hat{\gamma}_{\ell-1} x_{i\ell-1} + \hat{\gamma}_{\ell+1} x_{i\ell+1} + \cdots + \hat{\gamma}_k x_{ik}) c_{il}}{\sum_{i=1}^n \hat{r}_{il}^2} \\
&= - \frac{\hat{\gamma}_0 \sum_{i=1}^n c_{il} + \hat{\gamma}_1 \sum_{i=1}^n x_{i1} c_{il} + \cdots + \hat{\gamma}_{\ell-1} \sum_{i=1}^n x_{i\ell-1} c_{il} + \hat{\gamma}_{\ell+1} \sum_{i=1}^n x_{i\ell+1} c_{il} + \cdots + \hat{\gamma}_k \sum_{i=1}^n x_{ik} c_{il}}{\sum_{i=1}^n \hat{r}_{il}^2} \\
&= - \frac{\hat{\gamma}_0 \cdot 0 + \hat{\gamma}_1 \cdot 0 + \cdots + \hat{\gamma}_{\ell-1} \cdot 0 + \hat{\gamma}_{\ell+1} \cdot 0 + \cdots + \hat{\gamma}_k \cdot 0}{\sum_{i=1}^n \hat{r}_{il}^2} = 0
\end{aligned}$$