These consist of the three LS assumptions, plus two more:

1. $E(u|X = x) = 0$.
2. $(X_i, Y_i), i = 1, \ldots, n$, are i.i.d.
3. $E[Y^4] < \infty, E[X^4] < \infty$.
4. $var(u|X) = \sigma_u^2$.
5. $u \sim N(0, \sigma^2)$.

Three Gauss-Markov Conditions:

1. $E[u_i|X_1, \ldots, X_n] = 0$
2. $var[u_i|X_1, \ldots, X_n] = \sigma_u^2$, $0 < \sigma_u^2 < \infty$
3. $E[u_i u_j|X_1, \ldots, X_n] = 0$.

   If the model satisfies the GM conditions, $\hat{\beta}_1^{OLS}$ is BLUE.

$\hat{\beta}_1$ has the **smallest variance among" all linear estimators*** ("estimators that are linear functions of " Y_1,..., Y_n)." This is" "the " Gauss-Markov theorem. (BLUE = Best Linear Unbiased Estimator)

Proof: SW Appendix 5.2

Draw backs of OLS

1. The GM theorem isn't that compelling:
   1. Homoskedasticity often doesn't hold \
   2. Result is only for linear estimators – only a small subset of estimators (more on this)
2. The strongest optimality result ("part II") requires homoskedastic normal errors – not plausible in applications (hourly earnings data)
3. OLS is very sensitive to outliers.

   In estimating population mean, **if there are big outliers, median is preferred to mean because median is less sensitive to outliers**.

A Real-World Example for Heteroskedasticity {-}

Think about the economic value of education: if there were no expected economic value-added to receiving university education, you probably would not be reading this script right now. A starting point to empirically verify such a relation is to have data on working individuals. More precisely, we need data on wages and education of workers in order to estimate a model like

$wage_i = \beta_0 + \beta_1 \cdot education_i + u_i$.