**Gauss-Markov Assumptions**

| Assumption | Expression & Explanation | Potential Violations | Impact of Violations on Estimates |
|---|---|---|---|
| **Model is linear in parameters** **MLR 1** | $\beta$, not $\ln(\beta)$ or $\beta^2$ in the model.<br>$X_i$ can be transformed to be nonlinear, but all $\beta$s are linear | If actual relationship is not linear, involves quadratic, etc. | |
| **Samples are independent and random** **MLR 2** | Each individual equally as likely to be a part of our sample used for estimation of the $\beta$, and they all come from the population of interest. Implies no serial correlation in the errors | Non-random sampling, panel data | Increasing your sampling variance for $\beta$-hat. Will still be unbiased, but will be more **efficient** if serial correlation is in the *data.* |
| **No perfect collinearity in regressors** **MLR 3** | $R^2 = 1$<br>This would imply that all of y is perfectly explained by $x_1$ | High collinearity (e.g. 0.9)<br>As R2 $\rightarrow \infty$, Var($\beta$-hat) $\rightarrow \infty$<br>Including same var in different units | Increases the Var($\beta$-hat), meaning that your estimates are no longer **efficient.** |
| **Zero conditional mean of error term** **MLR 4** | $E(u \mid x) = 0$<br>$Cov(u_i, x_i) = 0$<br>Knowing x tells me nothing about the error term. Being at some relative point in the x-range does not imply that my error will be positive or negative. | Omitted variable bias<br>Reverse Causality<br><br>Measurement Error in Indep Var (bad proxy variables $\rightarrow$ more ME $\rightarrow$ more biased $\beta$-hat) | Estimates of $\beta$ will be **biased**, not centered around the true value of $\beta$<br><br>Said that x is **endogenous**, and impacting the bias of the estimate of $\beta$ |
| **Homoskedasticity of the errors ($u_i$)** **MLR 5** | $Var(u_i \mid x_i) = \sigma^2$<br>Error term has constant variance (ties in with zero conditional mean). Variance doesn't change dependent upon x-value. | Heteroskedasticity (variance in errors) is a function of x. Can arise from grouping data or aggregating data (error depends on the size of the groupings; avg errors no longer equal to one another) | Estimates are unbiased, but not **efficient**. Have a higher sampling variance for estimates of $\beta$. Fix with weighted least squares. Zero conditional mean is still preserved.<br><br>Do not need homoskedasticity for *consistency.* |
| **Errors are normally distributed** **MLR 6** | | | |
| **No serial correlation in the errors** | Implied by random sampling, but if we have no random sample, this assumption has to hold on its own. Important with panel data.<br><br>$E(y_i, y_j) = 0$, replace with the model relationship with $x_i$ to derive $Cov(u_i, u_j)=0$ | Occurs in panel data (efficiency)<br><br>Functional misspecification (bias)<br><br>Omitted variables (bias)<br><br>Measurement variable (bias) | Increasing your sampling variance for $\beta$-hat. Will still be unbiased, but will be more **efficient** if the serial correlation is in the *data.* Can be unbiased w/ right model. Can also be symptomatic of model specification, which also $\rightarrow$ **bias**. |

| | | | |
|---|---|---|---|
| **Non-zero sample variance in x**<br><br>**Var(x) ≠ 0** | X cannot be so closely clustered so as to not have any variance, and x cannot be a constant. You need variance to pick up an effect.<br>*This is a technical assumption; not much focus.* | If x is a constant | |

If these assumptions are met, the estimators are said to be **BLUE**

**B**est (efficient)
**L**inear (MLR1)
**U**nbiased (centered around the true value of $\beta$)
**E**stimators

BLUE estimators have the characteristics of being:

1. Unbiased
   a. Distribution of $\beta$-hat estimates are centered around true $\beta$
2. Consistent
   a. As sample size increases, the $\beta$-hat estimates get closer to true $\beta$
   b. Distribution starts wide and non-normal, then narrows and becomes normal as n increases
3. Efficient
   a. $\beta$-hat distribution has the lowest sampling variance ($\sigma^2$) of all possible $\beta$-hat estimates
   b. Narrower, taller distribution making it more efficient in that it is *more* centered around true $\beta$ and the estimates of $\beta$ from that distribution have less variance