OLS Efficiency

- Consistency is important, but it only tells us that we're right in expectation.
    - Every sample and computation of coefficients results in some amount of error.
- But how close are the coefficients to the true values?
    - Without an answer, there is no sense of scale or meaningfulness of the estimates.
    - Suppose we compute a slope of 2; could it be 3 if we repeated the experiment? Or 300? Are we convinced the real value is not 0?
    - Our coefficients are random variables because x_{i} and y_{i} are all random variables.

OLS Efficiency (cont.)

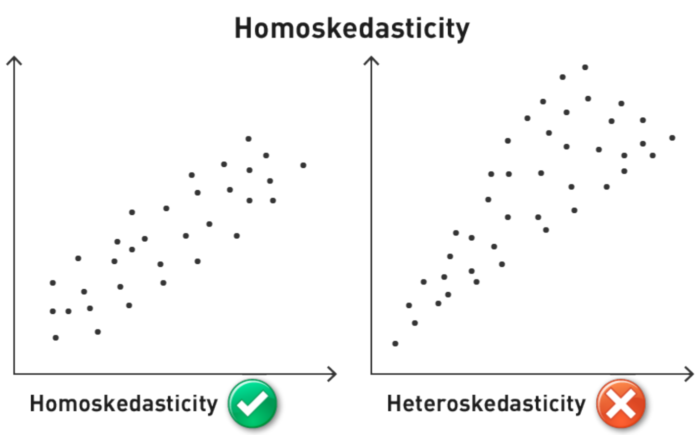
- The next step is to state how much our estimates will vary between samples.
- This is captured by the variance of each coefficient, $\mathrm{var}\left(\hat{\beta}_j\right)$.
- Computing this is important for a variety of statistical inference tasks.
- To do this, we need to add one more important assumption.

Homoskedasticity

Fifth and final Gauss-Markov assumption

- Assumption MLR.5 (Homoskedasticity): The variance of the error term is constant.
  - $\mathrm{Var}(u_i \mid x_1, x_2, \ldots, x_k) = \sigma^2$
- I.e., the error term cannot vary more for some values of *x*'s than others.
  - If considering error as including all unobserved factors, then factors vary equally for all values of *x*'s
- Explanatory variable values must contain no information about variability of the error.
- This is a strong assumption and is unrealistic for many real datasets.

Homoskedastic Plots



- Residuals are estimates of the error, so see if they have constant variance.
- For a fixed *x*, imagine taking a vertical slice through these plots.
- The thickness of the band indicates the variance; it should be the same for all *x*'s.

Sampling Variance of OLS Estimators

$$\mathrm{Var}\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\mathrm{SST}_j\,(1 - R_j^2)}, j = 1, ..., k$$

- Under assumptions MLR.1–MLR.5, we can compute an exact formula for variance of the slope coefficient.
- σ^2 is the variance of the error term.
  - The more the error varies, the more noise exists to throw off the estimates, so variance increases.
- SST_j is the total sample variation in explanatory variable x_j.
  - $\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$
  - The more variation in x_j with which to work, the more precise the estimate.

Sampling Variance of OLS Estimators (cont.)

$$\mathrm{Var}\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\mathrm{SST}_j\,(1 - R_j^2)}, j = 1, ..., k$$

- R-squared is from a regression of x_j on all other independent variables.
- This is the fraction of variation in x_j that cannot be explained by the other variables.
- It is only the unique variation in x_j that is left in the denominator.
- If there is multicollinearity, the unique variation will be small and precision is lost.

Multicollinearity

$$\text{avgscore} = \beta_0 + \beta_1 \text{ teachexp} + \beta_2 \text{ matexp} + \beta_3 \text{ othexp} + u$$

- In this example we are modeling test scores as a function of expenditure on teachers, expenditure on instructional materials, and other expenditures.
- Different expenditure categories will be strongly correlated because if a school has many resources, it will spend a lot on everything.
    - In most schools, all expenditures will tend to be high or all will tend to be low.
- We need information about situations in which a given category changes differently from other categories.
- As a result, sampling variance of the estimated effects will be large.
- We may decide to lump categories together or collect more data, depending on goals.

The Gauss-Markov Theorem

- Question: Though we have a formula for the variance of OLS coefficients, could we achieve less variance?
- Answer: Use **Theorem 3.4 (Gauss-Markov theorem).**
    - Under assumptions MLR.1–MLR.5, the OLS estimators are the best linear unbiased estimators (BLUEs) of the regression coefficients.
- We already know what linear unbiased estimators are.
- "Best" is defined as OLS coefficients having the smallest possible variance.

The Gauss-Markov Theorem (cont.)

- For any other linear unbiased estimator with coefficients, $\tilde{\beta}_j$
    - $\mathrm{Var}\left(\hat{\beta}_j\right) \leq \mathrm{Var}(\tilde{\beta}_j), j = 1, ..., k$
- The theorem provides a theoretical reason to use OLS.
    - It is the most famous benchmark for the performance of OLS.

More About BLUE

- Every letter in BLUE is necessary.
- OLS will only be Best in the class of Linear Unbiased Estimators.
- If we look at estimators that are not linear, even lower variance can be achieved.
    - E.g., when the variance of *y* changes with *x*, a technique called weighted least squares can be used to gain efficiency.

More About BLUE (cont.)

- If we look at estimators that are biased, lower variance can also be achieved.
  - E.g., we could use the estimators $\hat{\beta}_j = 0$
  - These are terrible estimators, but they have zero variance.
  - There are sometimes biased estimators that are still consistent and can outperform OLS.
    - This is sometimes the case for maximum likelihood estimators.