Heteroskedasticity … *and Robust Standard Errors*

- ***Those SLR/MLR conditions: LUEs and BLUE***
- ***… What happens under heteroskedasticity?***
- ***Learning from the Sample Mean Estimator***
- ***Turning to the SLR Model***
- ***Implications: Estimation and Inference***
- ***Back to Basics***
- ***… and White Corrected Standard Errors***
- ***Some Examples Using ", robust"***
- ***Let's Get Practical!***

Those SLR/MLR conditions: LUEs and BLUE

1. You may recall that back at the dawn of time, we rolled out those SLR/MLR conditions:

Those SLR Conditions	**Those MLR Conditions**
SLR.1: Linear Model (DGM) $Y = \beta_0 + \beta_1 X + U$	MLR.1: Linear Model (DGM) $Y = \beta_0 + \sum \beta_j X_j + U$
SLR2: Random sampling	MLR2: Random sampling
SLR3: Sampling variation in the independent variable	MLR3: No perfect collinearity amongst the RHS variables
SLR.4: The U's have zero conditional mean: $E(U \mid X = x) = 0$ for any x	MLR.4: The U's have zero conditional mean: $E(U \mid x_1, \ldots, x_n) = 0$ for any $x_1, \ldots, x_n$
SLR.5: Homoskedastic errors The U's have constant conditional variance $Var(U \mid X = x) = \sigma^2$ for all x	MLR.5: Homoskedastic errors The U's have constant conditional variance: $Var(U \mid x_1, \ldots, x_n) = \sigma^2$ for all $x_1, \ldots, x_n$

1) At that time, you saw that given SLR.1-.4 and MLR.1-.4, OLS estimators were linear unbiased estimators, LUE's. But you also saw that those OLS estimators were just one of about a gazillion LUEs.

2) In fact, in the context of SLR models…

   i) Any weighted average of the slopes of the lines connecting the data points to the samples means will also be a LUE (conditional on the x's) of the slope parameter.

   Here are a gazillion LUEs: $\sum \alpha_i \left(\frac{Y_i - \bar{Y}}{X_i - \bar{X}} \right)$, where $\sum \alpha_i = 1$

2. And so while it is nice to know that OLS estimators are LUEs (given those SLR/MLR conditions), that alone doesn't distinguish them from a crowd of other perhaps equally attractive linear unbiased estimators.

3. But when you add SLR.5/MLR.5 (constant conditional variance of the U's) into the mix, OLS estimators stand alone as Best Linear Unbiased Estimators (BLUE)… so that in the class of LUE's, OLS estimators have the smallest variance. That's the Gauss-Markov Theorem, which connects OLS and BLUE.

4. We now turn to the question: What if the U's are in fact heteroskedastic? What if SLR.5 and MLR.5 do not hold? What changes? How does that impact things?

… What happens under heteroskedasticity?

5. ***OLS estimators remain unbiased.*** If you have SLR.1-.4/MLR1.-.4, then OLS estimators are still LUEs, as that result does not require homoskedasticity. But of course, there are another gazillion LUEs, so maybe you should not be so impressed in this regard by the least squares estimators!

6. ***But, OLS estimators are no longer BLUE***.

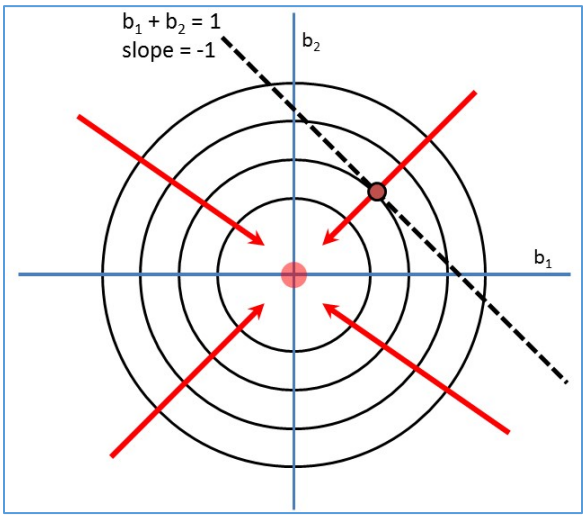
If you lose homoskedasticity, then OLS estimators will no longer be *Best Linear Unbiased Estimators*.

a. Recall that under OLS, you derive parameter estimates by minimizing SSRs, where, in a sense, each residual^2 receives equal weight in the summing up process. As you'll see below, under heteroskedasticity, the BLUE estimator will now have you minimizing weighted SSRs, where each residual^2 is weighted by the inverse of the variance of the respective observation. And so observations with higher variances will receive relatively less weight in summing up the residual^{2}s.

7. ***Some Intuition:*** Observations from distributions having larger variances are less reliable, or put differently, come with more noise attached (*noisier*). So while you don't want to completely ignore them, you do want to pay them less attention than the more reliable observations from distributions with smaller variances.

Learning from the Sample Mean Estimator

8. Also back at the dawn of time, you saw that with random sampling and given a certain set of assumptions, including homoskedasticity, the ***Sample Mean*** was a ***Best Linear Unbiased Estimator (BLUE)*** of the unknown mean of a distribution.



9. Recall our analysis at that dawn in time:
   a. ***Linear unbiased estimators***:

      $W = b_1Y_1 + b_2Y_2 + ... + b_nY_n$, where

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

   b. ***The BLUE challenge***:

      $\min\ Var(W) = \sum b_i{}^2\sigma^2 = \sigma^2 \sum b_i{}^2$

      subject to $\sum_{i=1}^{n} b_i = 1$

      Note that under homoskedasticity, you could take the σ^2 outside the summation, and so the particular value of σ^2 does not affect the b_i^*'s that solve the optimization problem.

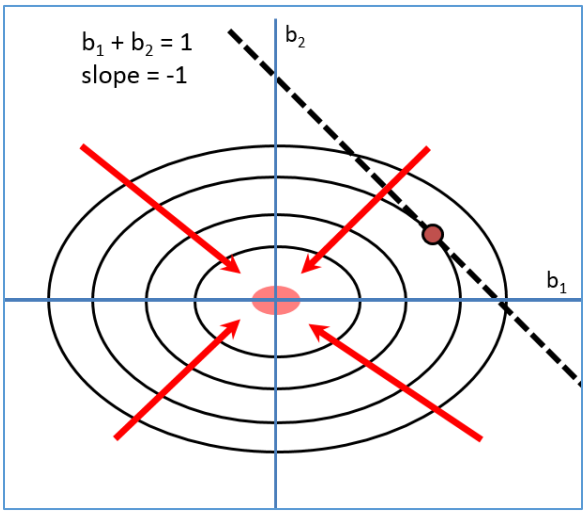
   c. ***The BLUE estimator***: $b_i^* = \frac{1}{n}$, so that $W = \bar{Y} = \frac{1}{n}\sum Y_i$.

10. Now, with heteroskedasticity, you have:
    a. ***Linear unbiased estimators***: Unchanged; as above.
    b. ***The BLUE challenge***:

       $\min\ Var(W) = \sum b_i{}^2\sigma_i^2$ subject to

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

       Note that under heteroskedasticity, you cannot simplify the expression to be minimized by taking σ^2 outside the summation. This complicates the minimization problem.



       Under homoskedasticity, the level curves of $Var(W)$ were circles, as in the figure above. Now, they are ellipses (or ellipsoids in higher dimensions; see right)… as they are defined by $b_1^2\sigma_1^2 + b_2^2\sigma_2^2 = constant$, for some *constant* level.

c. ***The BLUE estimator***:

$$\frac{\partial Var(W)}{\partial b_i} = \frac{\partial Var(W)}{\partial b_j} \Rightarrow 2\sigma_i^2 b_i^* = 2\sigma_j^2 b_j^* \text{ and so } \frac{b_i^*}{b_j^*} = \frac{\sigma_j^2}{\sigma_i^2}, \text{ or}$$

$$b_i^* = \frac{1/\sigma_i^2}{K}, \text{ where } K = \sum \frac{1}{\sigma_j^2}.$$

d. And so in this simple example, the BLUE will be a weighted average of the sampled values, where the weights are proportional to the inverse of the respective variances (the noise/uncertainty attached to each observation).

11. But wait! You almost never know the variances of the observations, the σ_i^2. So while all of this may make sense in theory, where are those weights coming from? That's a challenge, to which we'll return below.

Turning to the SLR Model

12. So what happens if SLR.5 or MLR.5 is violated and you have heteroskedastic errors?

13. The BLUE challenge remains the same… but has a new solution:

a. ***Linear unbiased estimators***:

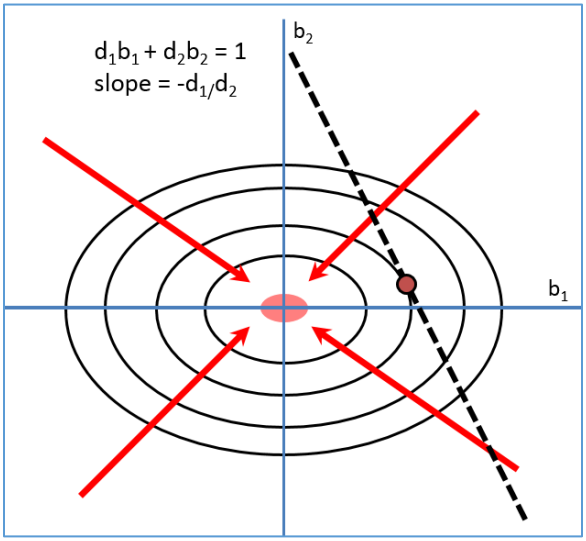
$W = b_1 Y_1 + b_2 Y_2 + ... + b_n Y_n$, where

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

b. ***The BLUE challenge***:

min $Var\left[\sum b_i Y_i\right] = \sum b_i^2 \sigma_i^2$ subject to

$\sum b_i = 0$ and $\sum b_i x_i = 1$.



14. We'll skip the details, but not surprisingly (given the results above), the BLUE estimator under these conditions will be a weighted least squares estimator, where each residual^2 is now weighted by the inverse of the variance of the observation:

a. min $wgtSSR = \sum \frac{1}{\sigma_i^2}(y_i - \beta_0 - \beta_1 x_i)^2$

b. If you know the variances, the σ_i^2's, then as easy way to implement this is to divide all the variables (including the constant variable, 1), by σ_i and run OLS, since

$$\sum\left(\frac{y_i}{\sigma_i}-\beta_0\frac{1}{\sigma_i}-\beta_1\frac{x_i}{\sigma_i}\right)^2=\sum\frac{1}{\sigma_i^2}(y_i-\beta_0-\beta_1x_i)^2$$

Note that in the OLS model, the simple constant variable has been replaced by a non-constant $\frac{1}{\sigma_i}$… so to run this, the RHS would have two types of variables, the $\frac{1}{\sigma_i}$'s and the $\frac{x_i}{\sigma_i}$'s, and no constant term.

c. Of course, and as above, you rarely know the variances, the σ_i^2's… which does make it rather difficult to run weighted least squares, eh? As promised, we'll come back to this.

Implications: Estimation and Inference

15. Let's continue with the SLR model. Recall that under SLR.1-SLR.5, you have:

a. SLR.5: $Var(U \mid X = x) = \sigma^2$ for all x.

b. Variance and standard deviation of the OLS slope estimator:

$$Var(B_1)=\frac{\sigma^2}{\sum(x_i-\bar{x})^2}=\frac{\sigma^2}{(n-1)S_{xx}} \text{ and } sd(B_1)=\frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}}=\frac{\sigma}{S_x\sqrt{n-1}}$$

c. $MSE=\frac{SSR}{n-2}$ is an unbiased estimator of σ^2: $E(MSE)=\sigma^2$

d. Standard error of the slope estimator:

$$se(B_1)=\frac{RMSE}{\sqrt{\sum(x_i-\bar{x})^2}}=\frac{RMSE}{S_x\sqrt{n-1}}, \text{ where } RMSE=\sqrt{MSE}.$$

16. And of course, once you have the standard error, you can generate t stats, p values, and confidence intervals, and assess statistical significance… which is to say: ***You can do inference!***

17. ***But without SLR.5/MLR.5, you lose all this***… made obvious by the fact that you no longer have a single variance, σ^2, to estimate. In fact, you have different σ_i^2's for the different observations. But no one told the statistical software that, and so those packages will continue to report $MSE=\frac{SSR}{n-2}$ and $se(B_1)=\frac{RMSE}{\sqrt{\sum(x_i-\bar{x})^2}}$, even though those reported figures are no longer as meaningful or useful as they were under homoskedasticity.

18. So what's an econometrician to do? Let's get back to basics, and assume SLR.1-SLR.4.

Back to Basics…

19. Assume that $Y_i = \beta_0 + \beta_1 x_i + U_i$ for *i=1, …, n* observations… so you are conditioning on the x's, and assume that the U_i's are independent with $E(U_i \mid x's) = 0$, $Var(U_i \mid x's) = \sigma_i^2$, and $Cov(U_i, U_j \mid x's) = 0 \;\; i \neq j$. And as above, assume that you know the variances, the σ_i^2's.

20. Since $B_1 = \dfrac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2} = \dfrac{\sum (x_i - \bar{x}) Y_i}{(n-1) S_{xx}}$, $Var(B_1) = Var\left(\sum \dfrac{(x_i - \bar{x})}{(n-1) S_{xx}} Y_i \right) = \sum \dfrac{(x_i - \bar{x})^2}{\left[(n-1) S_{xx} \right]^2} \sigma_i^2$

$= \dfrac{1}{\sum (x_i - \bar{x})^2} \sum \dfrac{(x_i - \bar{x})^2}{(n-1) S_{xx}} \sigma_i^2$.

21. So if you have heteroskedasticity, then

$$Var(B_1) = \frac{1}{\sum (x_i - \bar{x})^2} \sum w_i \, \sigma_i^2,$$

where $w_i = \dfrac{(x_i - \bar{x})^2}{(n-1) S_{xx}}$ are non-negative weights summing to 1.

22. So the variance of the slope estimator, B_1, is $\dfrac{1}{\sum (x_i - \bar{x})^2}$ times a ***weighted*** average of the σ_i^2's, $\sum w_i \, \sigma_i^2$, where the weights are proportional to the square of the x-distances from the mean.

    a. ***Check.*** If you have homoskedasticity, then you have the usual formula:

    $$Var(B_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}.$$

… and White Corrected Standard Errors

23. Various authors have proposed different ways in which to correct reported standard errors for heteroskedasticity. In what is I believe the most cited economics paper ever, Hal White suggested in 1980 that you use the squared residuals, the $\hat{u}_i^2$ from the SLR regression, to estimate the σ_i^2's.[1]

24. Following this suggestion:

    a. Run the OLS regression and capture the residuals, $\hat{u}_i = y_i - \beta_1 x_i$, *i=1, …, n.*[2]

    b. If you use $\hat{u}_i^2$ to estimate σ_i^2, and you make a $\left(\frac{n}{n-2}\right)$ degrees of freedom adjustment to the formula, you can estimate $Var(B_1)$ using the previous formula:
    $\left(\frac{n}{n-2}\right)\frac{1}{\sum(x_i-\overline{x})^2}\sum w_i\ \hat{u}_i^2$.

    c. The square root of this will be the standard error:
    $$se^*(B_1) = \sqrt{\left(\frac{n}{n-2}\right)\frac{1}{\sum(x_i-\overline{x})^2}\sum w_i\ \hat{u}_i^2}$$
    This gives us White's *heteroskedasticity-corrected standard error*, sometimes called the *robust standard error*.

25. If you take an ***unweighted*** average of the $\hat{u}_i^2$, then you have the calculated OLS variance, which you have seen before:
    $$=\left(\frac{n}{n-2}\right)\frac{1}{\sum(x_i-\overline{x})^2}\sum\frac{1}{n}\hat{u}_i^2 = \frac{SSR/(n-2)}{\sum(x_i-\overline{x})^2}.$$

26. So the OLS variance (calculated assuming homoskedasticity) is driven by an ***unweighted*** average of the $\hat{u}_i^2$, whereas the heteroscedasticity corrected variance is driven by a ***weighted*** average of the $\hat{u}_i^2$, where the weights are proportional to the square of the x-distances from $\overline{x}$, the $(x_i-\overline{x})^2$'s.

27. It is often believed that robust standard errors are always larger than OLS reported standard errors. Indeed, the word *robust* does suggest that, eh? But that is not correct! Robust standard errors are sometimes larger than reported OLS standard errors, and sometimes smaller. It all depends on the relationship between the $\hat{u}_i^2$ and the $(x_i-\overline{x})^2$'s. If observations with above average $\hat{u}_i^2$ receive higher weights (have larger x-distances from the

[1] White, Halbert (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48 (4): 817–838.
[2] Note that by construction, these will have mean 0.

mean, $(x_i - \overline{x})^2$'s) then the robust standard errors will be larger than the reported OLS standard errors, and vice –versa. (See simulation figures and discussion below.)

28. It is also often believed that heteroskedasticity alone causes robust standard errors to differ from the OLS reported standard errors. That is also not correct. As shown in the following example, if there is no systematic relationship between the $\hat{u}_i^2$ and the $(x_i - \overline{x})^2$'s , then the robust standard errors will be very similar to the reported OLS standard errors. But reported standard errors will change under robust estimation of the there is some systematic relationship between the $\hat{u}_i^2$ and the $(x_i - \overline{x})^2$'s . If that relationship is positive, then larger residuals will receive greater weight, and robust standard errors will be larger (and t stats smaller) than the reported OLS se's … and is the relationship is negative, then the opposite occurs.
29. To repeat: Differences between reported OLS standard errors and robust se's, are not driven so much by heteroskedasticity alone, but rather by the extent to which that heteroskedasticity is correlated with deviations in the RHS variable from its mean.

Some Examples Using ", robust"

30. To generate robust standard errors in Stata, just add ***", robust"*** to the end of your regression command. The estimated coefficients will be unchanged, since under SLR.1-SLR.4, OLS remains a LUE. But the reported standard errors will change, and with that change, you'll see changes in t stats, p values, confidence intervals and perhaps, statistical significance.

    Here are two examples.

31. ***Example 1 – Sovereign Debt***: Here's an example using the *sovdebt* dataset (standard errors are in parentheses):

	(1) NSRate	(2) NSRate (robust)
corrupt	0.562***	0.562***
	(0.0369)	(0.0388)
lngdp	0.343***	0.343***
	(0.0380)	(0.0366)
inflation	-0.0498**	-0.0498*
	(0.0180)	(0.0192)
deficit_gdp	-0.0439**	-0.0439***
	(0.0130)	(0.00909)
debt_gdp	-0.00884***	-0.00884**
	(0.00227)	(0.00279)
_cons	2.593***	2.593***
	(0.244)	(0.182)
N	108	108

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Note that not surprisingly, some standard errors have increased with the *robust* specification (*corrupt*, *inflation*, *debt_gdp*)… while others (*lngdp* and *deficit_gdp*) have decreased. The reported coefficients are, as expected, unchanged.

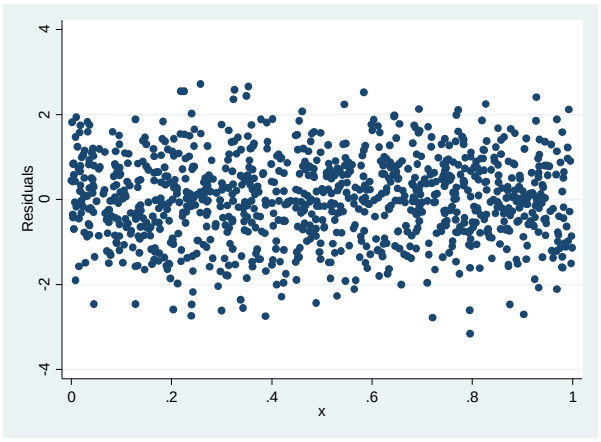
32. ***Example 2 – Simulation***: This example shows the importance of some sort of systematic relationship between the $\hat{u}_i^2$ and the $(x_i - \bar{x})^2$'s in driving differences between OLS reported standard errors and robust standard errors. In this example, the x's are uniformly distributed on [0,1], and the y's are generated by the equation $Y_i = x_i + U_i$, where the U's have different specifications to illustrate the impact of heteroskedasticity:

    a. Case I: $U_i \sim N(0,1)$

    b. Case II: $U_i \sim 2 \cdot |x - .5| \cdot N(0,1)$

    c. Case III: $U_i \sim \frac{1}{16 \cdot |x - .5|} N(0,1)$ (drop the observation if $U_i > 1$)

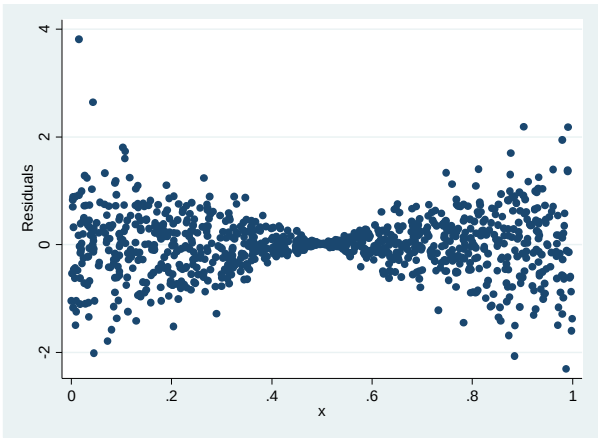
33. *Case I – Standard Errors Unchanged*: No systematic relationship between the $\hat{u}_i^2$ and the $(x_i - \bar{x})^2$'s



	(1) y0	(2) y0 robust
x	1.083***	1.083***
	(0.108)	**(0.106)**
_cons	-0.0447	-0.0447
	(0.0623)	(0.0620)
N	1000	1000

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

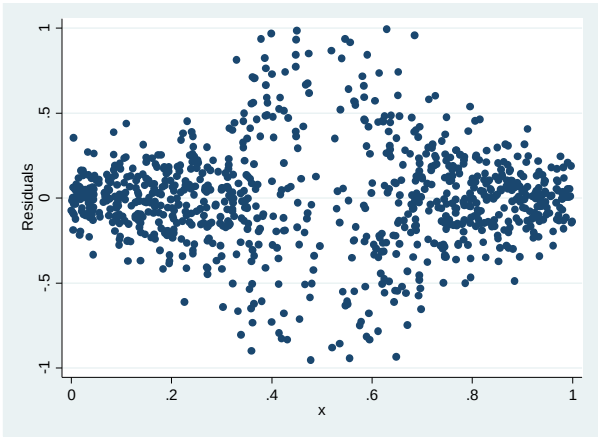
34. *Case II – Standard Errors Increase*: Positive systematic relationship between the $\hat{u}_i^2$ and the $(x_i - \bar{x})^2$'s



	(1) y2	(2) y2 robust
x	0.976***	0.976***
	(0.0606)	**(0.0801)**
_cons	-0.00325	-0.00325
	(0.0348)	(0.0437)
N	1000	1000

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

35. Case III – *Standard Errors Decrease*: Negative systematic relationship between the $\hat{u}_i^2$ and the $(x_i - \bar{x})^2$'s



	(1) y1	(2) y1 robust
x	0.990***	0.990***
	(0.0326)	**(0.0201)**
_cons	0.00959	0.00959
	(0.0190)	(0.0136)
N	903	903

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Let's Get Practical!

36. So what to make of all this? The lessons are really quite simple!

a. Run weighted least squares if you can… but no sweat if you can't, as OLS estimators are still LUEs, with either homoskedasticity or heteroskedasticity.

b. If you want to give weighted least squares a try, think about using proxy variables, which might be correlated with the variances of the observations. Don't hold back: Try different weighting schemes and see if the weights matter much. And if they do, then deal with it!

c. And irrespective of the weighting, add ***, robust*** to generate robust standard errors. It's still OK to do this even if you do not have heteroskedasticity. And while there are tests for heteroskedasticity, there is no need for them at all. Just run ***, robust*** and move on with (a *robust*) life!