Vanderbilt University Political Science - Stats I
Fall 2024 - Prof. Jim Bisbee

21 Lecture 21

21.1 Heteroskedasticity and What to Do About It

- As discussed earlier in the course, we assume homoskedasticity of the errors across all observations in order to vastly simplify our calculation of $Var\left(\widehat{\beta}_j\right)$. By assumiming that $\sigma_i^2 = \sigma^2$ for all i, we can then write

$$
\begin{aligned}
Var\left(\widehat{\beta}_j\right) &= \frac{\sigma^2}{n \cdot var(x_j) \cdot \left(1 - R_j^2\right)} \\
\widehat{Var\left(\widehat{\beta}_j\right)} &= \frac{\widehat{\sigma}^2}{n \cdot var(x_j) \cdot \left(1 - R_j^2\right)}
\end{aligned}
$$

- We also needed the homoskedasticity assumption in order for the Gauss-Markov theorem to hold that OLS is the best linear unbiased estimator of the parameters of a linear population model.

- What to do? There are two approaches:

  - Heteroskedasticity of unknown form (the safe, but ignorant and often inefficient approach)

  - Modeling heteroskedasticity (requires more assumptions, but if assumptions are correct the efficient approach)

21.1.1 Heteroskedasticity of unknown form: use robust ("White") standard errors

- In the simple bivariate case, we of course write the model

$$y_i = \beta_0 + \beta_1 x_{1i} + u_i$$

- The presence of heteroskedasticity means that we can no longer write

$$VAR(u_i|x_i) = \sigma^2.$$

- We of course need to write instead

$$VAR(u_i|x_i) = \sigma_i^2,$$

because the value of σ^2 now depends on the value of x_i.

- Recall that in our final step of deriving the OLS estimator in scalar form we write

$$\widehat{\beta}_1 = \beta_1 + \frac{\sum (x_i - \overline{x}) u_i}{\sum (x_i - \overline{x})^2}.$$

- So now consider

$$\begin{aligned} VAR\left(\widehat{\beta}_1\right) &= VAR\left[\frac{\sum (x_i - \overline{x}) u_i}{\sum (x_i - \overline{x})^2}\right] \\ &= \left[\frac{1}{SST_x}\right]^2 \sum (x_i - \overline{x})^2 VAR(u_i|x_i) \\ &= \frac{\sum (x_i - \overline{x})^2 \sigma_i^2}{SST_x^2}. \end{aligned}$$

- What to do? Well, in 1980 (in the most cited economics paper in the past 35 years), Halbert White showed that a valid estimator for $VAR\left(\widehat{\beta}_1\right)$ in the presence of heteroskedasticity (if the other Gauss-Markov assumptions hold) is

$$\widehat{VAR\left(\widehat{\beta}_1\right)} = \frac{\sum (x_i - \overline{x})^2 \widehat{u}_i^2}{SST_x^2},$$

where $\widehat{u}_i^2$ is simply the squared residual associated with each observation i.

- A similar formula holds in the multiple regression model, where we write

$$\widehat{VAR\left(\widehat{\beta}_1\right)} = \frac{\sum \widehat{r}_{ij}^2 \widehat{u}_i^2}{SSR_j^2},$$

where

- $\widehat{r}_{ij}$ is the residual obtained for observation i when regressing x_j on all the other x's, and
- SSR_j^2 is the sum of squared residuals from this regression.
- Note the similarities to the formula in the bivariate case.

- I hope it is obvious then that the estimated standard error of $\widehat{\beta}_1$ is

$$\sqrt{\frac{\sum \widehat{r}_{ij}^2 \widehat{u}_i^2}{SSR_j^2}}.$$

- These standard errors have lots of different names:
  - "White standard errors"
  - "Huber-White standard errors"
  - "Robust standard errors" (because the se's are "robust" in the presence of heteroskedasticity)
  - "Heteroskedasticity-robust standard errors"
  - These all mean the same thing.

- It is often–but not always–the case that robust standard errors are larger than OLS standard errors.

21.1.2 Testing for heteroskedasticity

- We can blithely report robust standard errors to be sure that our hypothesis tests are correct in the presence of heteroskedasticity.

- But, remember that if heteroskedasticity is present, OLS is no longer the best linear unbiased estimator. As we will see, you can obtain a better estimator when the form of heteroskedasticity is known.

- We are interested in tests that detect error variance that depends on the value of x. We start with the linear model

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u.$$

- Now let's specify the null hypothesis

$$H_0 : VAR(u|x_1, x_2, ...x_k) = \sigma^2$$

- Under the zero condition mean assumption this is equivalent to

$$H_0 : E(u^2|x_1, x_2, ...x_k) = E\left(u^2\right) = \sigma^2.$$

- So how do we test whether u^2 is related to the x's? How about assuming a linear function

$$u^2 = \delta_0 + \delta_1 x_1 + ... + \delta_k x_k + \nu$$

  (Why am I using delta's instead of beta's here?)

- The null hypothesis now becomes

$$H_0 : \delta_0 = \delta_1 = ...\delta_k = 0.$$

- We of course do not have u^2 - these are population values that we never see. But we have estimates of u^2–our squared residuals, the $\widehat{u}^2$. So if we estimate the equation

$$\widehat{u}^2 = \delta_0 + \delta_1 x_1 + ... + \delta_k x_k + error$$

  we now have a test to see the extent to which the errors in the population are related to one or more of the x's.

- One approach would be to see if any of the delta's are statistically significant. But what might be a better way?

- Look at the F-statistic from this regression, which tells us whether the x's are *jointly* significant in explaining the squared residuals. (For once, the dumb F-stat provided by typical

OLS output is helpful here!) You'll recall that we defined the F-statistic as

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$$

- Noting that $SST = SSE + SSR$ and $R^2 = \frac{SSE}{SST}$ and $1 - R^2 = \frac{SSR}{SST}$, we can write $SSR = SST\left(1 - R^2\right)$. Now rewrite F as

$$\begin{aligned} F &\equiv \frac{\left[SST_r\left(1-R_r^2\right) - SST_{ur}\left(1-R_{ur}^2\right)\right]/q}{\left[SST_{ur}\left(1-R_{ur}^2\right)\right]/(n-k-1)} \\ &= \frac{\left[\left(1-R_r^2\right) - \left(1-R_{ur}^2\right)\right]/q}{\left(1-R_{ur}^2\right)/(n-k-1)} \quad [\text{since } SST_r = SST_{ur}] \\ &= \frac{\left(R_{ur}^2 - R_r^2\right)/q}{\left(1-R_{ur}^2\right)/(n-k-1)}. \end{aligned}$$

- In the case where we are testing the joint significance of all the coefficients in a model, the restricted equation is

$$y = \beta_0 + u.$$

- Note that this equation explains none of the variation in y, as there is nothing on the right-hand side that varies except u. Thus $R_r^2 = 0$, and so the F-statistic in this case (and in any case where the test is that all the variables in the model are jointly insignificant) is equal to

$$F = \frac{\left(R_{ur}^2\right)/k}{\left(1-R_{ur}^2\right)/(n-k-1)}.$$

- Note that in this case $q = k$, as we have restricted the values of each of the k x's to be zero.

- This statistic as approximately a $F_{k,n-k-1}$ distribution under H_0, and so the p-value associated with this F-statistic is the probability that we could have obtained the coefficients we see by chance if there were no heteroskedasticity. So where $p < .05$ (or as Stata puts it, "Prob ¿ F" is less than .05, we reject H_0 at the .05 level and decide that heteroskedasticity is present.

- There are lots of other tests for heteroskedasticity. They all follow the same general pattern but with more complexity. Read about them if you like on pages. 271-276 of your text.

21.1.3 Modeling Heteroskedasticity

- We don't have time to cover the ways heteroskedasticity is modeled and corrected for using what is called generalized least squares. (Neal is likely to pick this topic up in Week 1 or Week 2 of Quant II.) The preview is this:
  - Model the heteroskedasticity using versions of the linear model we used above.
  - Determine the extent to which the errors change with each observation.
  - Instead of running OLS, which counts each observation the same when it minimizes the sum of squared residuals...
  - ...run weighted least squares, which *downweights* those observations with a higher error variance when minimizing the sum of squared residuals.
  - If you have modeled the heteroskedasticity correctly, you now have estimators that are BLUE.
  - If you haven't modeled it correctly, you have biased estimates of the βs.

21.1.4 What to do

- Generally you want to be able to say that your results are robust to the threat of heteroskedasticity.
- By presenting robust standard errors, you can assure your reader that the statistical significance of a particular $\widehat{\beta}_k$ is not due to an improperly estimated $VAR\left(\widehat{\beta}_k\right)$.
- Notice that because robust se's are (generally) larger than OLS se's, you're taking the safe route.
- HOWEVER, what if your paper relies on the idea that β_k is zero–a failure to reject the null? Then you'll probably want to venture into modeling heteroskedasticity, because proper modeling yields results that are more efficient–that is, less likely to get a false negative result.
- This is complicated. You'll learn more in Quant II.

21.2 Transformations of Variables

- Go over "Transforming Nonlinearity" from Fox.