

BSM946 - Workshop 2: Heteroskedasticity

Open a do-file and run all your commands through it. This will allow you to keep a log of what you have done. If you want to add notes to the do file so you have explanations of the commands or results just type % and the line will not run as you execute the file. It is suggested to keep notes from the workshops to help you when you revisit this material.

Open the STATA dataset we will be using:

bcuse attend

We are going to look at the following regression to see how attendance influences grades:

regress final priGPA attend frosh soph hwrt

where final is the students final exam score in the module, priGPA is the students grade point average (Grade) across all modules up to this point attend is the students attendance, frosh and soph are dummy variables which show which year of study the student is in and hwrt is the percentage of homework the student has turned in for the module.

HETEROSKEDASTICITY

1) Lets start by graphically testing for heteroskedasticity.

Use the command **rvfplot** to plot the residuals against the fitted values

Use the command **rvpplot x** to test plot the residuals against any independent variable X.

Do you think there is heteroskedasticity in the sample?

2) As noted in earlier class, the **predict** command allows you to generate new variables containing predicted values such as residuals and standard errors.

To obtain the predicted errors of the regression simply run the regression and then obtain the estimated residual values through using the predict command.

predict yhat (provides the predicted y values for the regression)

predict uhat, resid (provides the predicted residuals. The predicted residuals are used in the calculation of many of the tests for heteroskedasticity)

Try to run the Goldfeld-Quandt test with the independent variable attend using the commands seen in the lecture:

estat sum x (To find the mean of the X variable of interest)

gen groupA=X< Mean from above

gen uhatsq = uhat^2

sdtest uhatsq, by(groupA)

3) Does this Goldfeld-Quant Test suggest there is heteroskedasticity?

4) Run the Breusch-Pagan test for heteroskedasticity using the command **hettest**. The null hypothesis for this test is homoskedasticity (constant error variance.)

Can you explain what this command does? Does the result suggest that there is any heteroskedasticity?

Correcting for Heteroskedasticity

To obtain the White corrected robust variance estimates, simply type in the **robust** option at the end of the regression command.

regress final priGPA attend frosh soph hwrtte, robust

How do these results compare to the original regression results you estimated?