Causal Inference: Problem Set 12

Bezirgen Veliyev (Aarhus University)

November 30, 2020

Based on Chapter 8 and 12 problems of Wooldridge.

1. Which of the following are consequences of heteroskedasticity?
   (i) The OLS estimators are biased and inconsistent.
   (ii) The usual F-statistic no longer has an F-distribution.
   (iii) The OLS estimators are no longer BLUE.
2. True or False: WLS is preferred to OLS when an important variable has been omitted from the model.
3. During lectures, we covered Breusch-Pagan (BP) and White tests for heteroskedasticity. It turns out that the White test can be made more efficient. A special case of the White test regresses $\hat{u}_i^2$ on $\hat{y}_i$ and ${\hat{y}_i}^2$, where the $\hat{u}_i$ are the OLS residuals and the $\hat{y}_i$ are the OLS fitted values. In this problem, we propose a hybrid test that combines BP and the special case of White test, by running the regression $\hat{u}_i^2$ on every regressor and ${\hat{y}_i}^2$. Then, we would test the joint significance of the whole model. (Of course, we always include an intercept in this regression.)
   (i) What are the df associated with the proposed hybrid F-test for heteroskedasticity?
   (ii) Explain why the R-squared from the regression above will always be at least as large as the R-squareds for the BP regression and the special case of the White test.
   (iii) Does part (ii) imply that the new test always delivers a smaller p-value than either the Breusch-Pagan or the special case of the White statistic? Explain.
   (iv) Suppose someone suggests also adding $\hat{y}_i$ to the newly proposed test. What do you think of this idea?
4. Use the data set GPA1 for this exercise.
   (i) Use OLS to estimate a model relating colGPA to hsGPA, ACT, skipped, and PC, where these relate to the individual's college GPA, highschool GPA, score on a prior achievement test, number of classes skipped, and whether they have a PC (Yes, this is an old dataset!). Test if there is heteroskedasticity.

(ii) Obtain the OLS residuals from part (i) and run the regression of $\log(\hat{u}^2)$ on all of the original regressors. Denote the exponentiated fitted values with $\hat{h}_i$.

(iii) Obtain the weighted least squares estimates using weights $1/\sqrt{\hat{h}}$. Compare the weighted least squares estimates for the effect of skipping lectures and the effect of PC ownership with the corresponding OLS estimates. What about their statistical significance?

(iv) Test to see if there is still heteroskedasticity in the WLS model. Should we also include robust standard errors in our output?

5. When the errors in a simple bivariate regression model have AR(1) serial correlation, why do you think the OLS standard errors tend to underestimate the standard errors? Is it always true that the OLS standard errors are too small?

6. Use the data in TRAFFIC2 for this exercise. It includes information about traffic accidents in the 1980s in the US.

   (i) Run an OLS regression of prcfat (percentage of fatal accidents out of all accidents) on a linear time trend, monthly dummy variables, and the variables wkends (number of weekends in the month), unem (unemployment rate), spdlaw (whether there is a law on the maximum speed limit), and beltlaw (whether there is a law on wearing seatbelts). Comment on the results for the two policy variables. Why is it important to include a time trend here?

   (ii) Test the errors for AR(1) serial correlation using the Breusch-Godfrey test and the Durbin-Watson test. Do the tests agree? Does it make sense to use tests that assume strict exogeneity of the regressors in this case?

   (iii) Obtain serial correlation- and heteroskedasticity-consistent (HAC) standard errors for the coefficients. Do the coefficients change? Explain. How do these HAC standard errors affect the statistical significance of the two policy variables?