Econ 4223 - Fall 2018 **Problem Set 3 - Solutions** **Due: Feb. 26 beginning of class**

Warm Up

1. Wooldridge (2015), exercise 4.1

Solution: (i) and (iii) generally cause the t statistics not to have a t distribution under H_0. Homoskedasticity is one of the CLM assumptions. An important omitted variable violates Gauss-Markov Assumption 4. The CLM assumptions contain no mention of the sample correlations among independent variables, except to rule out the case where the correlation is exactly 1.

Exercises

2. *Wooldridge*, 4.5

Solution: Part (i): $.412 \pm 1.96(.094)$, or about .228 to .596.

Part (ii): No, because the value .4 is well inside the 95% CI

Part (iii): Yes, because 1 is well outside the 95% CI.

3. *Wooldridge*, 8.4

Solution: Part (i): These coefficients have the anticipated signs. If a student takes courses where grades are, on average, higher—as reflected by higher *crsgpa*—then his/her grades will be higher. The better the student has been in the past—as measured by *cumgpa*—the better the student does (on average) in the current semester. Finally, *tothrs* is a measure of experience, and its coefficient indicates an increasing return to experience.

The t statistic for *crsgpa* is very large, over five using the usual standard error (which is the largest of the two). Using the robust standard error for *cumgpa*, its t statistic is about 2.61, which is also significant at the 5% level. The t statistic for *tothrs* is only about 1.17 using either standard error, so it is not significant at the 5% level.

Part (ii): This is easiest to see without other explanatory variables in the model. If *crsgpa* were the only explanatory variable, $H_0 : \beta_{crsgpa} = 1$ means that, without any information about the student, the best predictor of term GPA is the average

GPA in the students' courses; this holds essentially by definition. (The intercept would be zero in this case.) With additional explanatory variables, it is not necessarily true that $\beta_{crsgpa} = 1$ because *crsgpa* could be correlated with characteristics of the student. (For example, perhaps the courses students take are influenced by ability—as measured by test scores—and past college performance.) But it is still interesting to test this hypothesis.

The t statistic using the usual standard error is $t = (.900 - 1)/.175 \approx -.57$; using the heteroskedasticity-robust standard error gives $t \approx -.60$. In either case we fail to reject $H_0 : \beta_{crsgpa} = 1$ at any reasonable significance level, certainly including 5%.

Part (iii): The in-season effect is given by the coefficient on *season*, which implies that, other things equal, an athlete's GPA is about .16 points lower when his/her sport is competing. The t statistic using the usual standard error is about -1.60, while that using the robust standard error is about -1.96. Against a two-sided alternative, the t statistic using the robust standard error is just significant at the 5% level (the standard normal critical value is 1.96), while using the usual standard error, the t statistic is not quite significant at the 10% level ($cv \approx 1.65$). So the standard error used makes a difference in this case. This example is somewhat unusual, as the robust standard error is more often the larger of the two.

Computer Exercises You should use `R` to complete these exercises. Any data set referred to in the question should be available in the `wooldridge` package in `R`. You do not need to turn in an R-script for these questions, but you are welcome to do so if you would like to.

4. *Wooldridge*, 4.C12 (Chapter 4, Computer Exercise 12)

Solution: Part (i): The estimated equation, with standard errors in parentheses below coefficient estimates, is

$$\widehat{colgpa} = \underset{(.17)}{.028} + \underset{(.053)}{.659}\,hsgpa + \underset{(.005)}{.013}\,actmth + \underset{(.005)}{.012}\,acteng$$
$$N = 856, R^2 = .256.$$

The explanatory variables are individually significant at 5% level, with $t_{hsgpa} \approx 12.4$, $t_{actmth} \approx 2.6$, and $t_{acteng} \approx 2.4$.

Part (ii): One standard deviation increase in *hsgpa* is associated with an increase in *colgpa* of about $0.659(.343) \approx .226$, or a three percent growth. The *actmth* would have to increase by about five standard deviations to change $\widehat{colgpa}$ by the same amount as a one standard deviation in *hsgpa*. The variable *actmth* has a small effect on $\widehat{colgpa}$ compared to *hsgpa*.

Part (iii): You can easily do this with the R code `linearHypothesis(est,'actmth=acteng')`. The F-statistic is about 0.0088 and the p-value is 0.9253. There is essentially no evidence to overturn $H_0 : \beta_{actmth} = \beta_{acteng}$.

Part (iv): The variables *hsgpa*, *actmth*, and *acteng* can explain only about 26% variation in *colgpa* ($R^2 = .256$). To create an equation that explains at least 50% variation in *colgpa*, we should include other explanatory variables that are correlated with *colgpa* along with the above variables.

5. *Wooldridge*, 8.C4 and 8.C7

Solution: 8.C4, Part (i): The estimated equation is

$$\widehat{voteA} = \underset{(4.74)}{37.66} + \underset{(.071)}{.252}\, prtystrA + \underset{(1.407)}{3.793} democA + \underset{(.392)}{5.779} \log(expendA) - \underset{(.397)}{6.238} \log(expendB) + \hat{u}$$

$$N = 173, R^2 = .801, \overline{R}^2 = .796.$$

You can convince yourself that regressing the $\hat{u}_i$ on all of the explanatory variables yields an R-squared of zero, although it might not be exactly zero in your computer output due to rounding error. Remember, OLS works by choosing the estimates, $\hat{\beta}_j$, such that the residuals are uncorrelated in the sample with each independent variable (and the residuals have a zero sample average, too).

8.C4, Part (ii): The B-P test entails regressing the $\hat{u}_i^2$ on the independent variables in part (i). The F statistic for joint significance (with 4 and 168 df) is about 2.33 with p-value $\approx .058$. Therefore, there is some evidence of heteroskedasticity, but not quite at the 5% level.

8.C4, Part (iii): Now we regress $\hat{u}_i^2$ on $\widehat{voteA}_i$ and $(\widehat{voteA}_i)^2$, where the $\widehat{voteA}_i$ are the OLS fitted values from part (i). The F test, with 2 and 170 df, is about 2.79 with p-value $\approx .065$. This is slightly less evidence of heteroskedasticity than provided by the B-P test, but the conclusion is very similar.

8.C7, Part (i): The heteroskedasticity-robust standard error for $\hat{\beta}_{white} \approx .129$ is about .026, which is notably higher than the nonrobust standard error (about .020). The heteroskedasticity-robust 95% confidence interval is about .078 to .179, while the nonrobust CI is, of course, narrower, about .090 to .168. The robust CI still excludes the value zero by some margin.

8.C7, Part (ii): There are no fitted values less than zero, but there are 213 greater than one. Unless we do something to those fitted values, we cannot directly apply WLS, as $\hat{h}_i$ will be negative in 213 cases.

Cool Down

6. *Wooldridge*, 8.1.

Solution: (ii) and (iii). The homoskedasticity assumption played no role in Chapter 5 in showing that OLS is consistent. But we know that heteroskedasticity causes statistical inference based on the usual t and F statistics to be invalid, even in large samples. As heteroskedasticity is a violation of the Gauss-Markov assumptions, OLS is no longer BLUE.

References

Wooldridge, Jeffrey M. 2015. *Introductory Econometrics: A Modern Approach*. Cengage Learning, 6 ed.