FINM3123 Introduction to Econometrics

Chapter 8 Exercises

Solutions

1. With Var(u|$inc, price, educ, female$) = $\sigma^2 inc^2$, $h(\mathbf{x}) = inc^2$, where $h(\mathbf{x})$ is the heteroskedasticity function defined in equation (8.21). Therefore, $\sqrt{h(\mathbf{x})} = inc$, and so the transformed equation is obtained by dividing the original equation by inc:

$$\frac{beer}{inc} = \beta_0(1/inc) + \beta_1 + \beta_2(price/inc) + \beta_3(educ/inc) + \beta_4(female/inc) + (u/inc).$$

Notice that β_1, which is the slope on inc in the original model, is now a constant in the transformed equation. This is simply a consequence of the form of the heteroskedasticity and the functional forms of the explanatory variables in the original equation.

2. (i) No. For each coefficient, the usual standard errors and the heteroskedasticity-robust ones are practically very similar.

(ii) The effect is $-.029(4) = -.116$, so the probability of smoking falls by about .116.

(iii) As usual, we compute the turning point in the quadratic: $.020/[2(.00026)] \approx 38.46$, so about 38 and one-half years.

(iv) Holding other factors in the equation fixed, a person in a state with restaurant smoking restrictions has a .101 lower chance of smoking. This is similar to the effect of having four more years of education.

(v) We just plug the values of the independent variables into the OLS regression line:

$$\widehat{smokes} = .656 - .069 \cdot \log(67.44) + .012 \cdot \log(6,500) - .029(16) + .020(77) - .00026(77^2) \approx .0052.$$

Thus, the estimated probability of smoking for this person is close to zero. (In fact, this person is not a smoker, so the equation predicts well for this particular observation.)

3.

(i) The estimated equation is

$$\widehat{voteA} = \underset{(4.74)}{37.66} + \underset{(.071)}{.252}\, prtystrA + \underset{(1.407)}{3.793}\, democA + \underset{(0.392)}{5.779}\log(expendA)$$

$$- \underset{(0.397)}{6.238}\log(expendB) + \hat{u}$$

$$n = 173, \quad R^2 = .801, \quad \bar{R}^2 = .796.$$

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).

(ii) The B-P test entails regressing the $\hat{u}_i^2$ on the independent variables in part (i). The F statistic for joint significant (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.

(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.

4.

(i) The estimated equation is

$$\widehat{math4} = \underset{\substack{(19.96)\\ [23.09]}}{91.93} - \underset{\substack{(.015)\\ [.017]}}{.449}\, lunch - \underset{\substack{(0.94)\\ [1.13]}}{5.40}\, lenroll + \underset{\substack{(2.10)\\ [2.35]}}{3.52}\, lexppp$$

$n = 1{,}692,\ R^2 = .373$

The heteroskedasticity-robust standard errors are somewhat larger, in all cases, than the usual OLS standard errors. The robust *t* statistic on *lexppp* is about 1.50, which raises further doubt about whether performance is linked to spending.

(ii) The value of the *F* statistic is 132.7, which gives a *p*-value of zero to at least four decimal places. Therefore, there is strong evidence of heteroskedasticity.

(iii) The equation estimated by WLS is

$$\begin{aligned}\widehat{math4} = \ &50.48 \quad - .449\ lunch \quad - 2.65\ lenroll \quad + 6.47\ lexppp\\ &(16.51) \quad\ (.015) \qquad\qquad (0.84) \qquad\qquad (1.69)\end{aligned}$$

$n = 1{,}692,\ R^2 = .360$

where the usual WLS standard errors are in (·). The OLS and WLS coefficients on *lunch* are the same to three decimal places, but the other coefficients differ in practically important ways. The most important is that the WLS coefficient on *lexppp* is much larger than the OLS coefficient. Now, a 10 percent increase in spending (so *lexppp* increases by .1) is associated with roughly a .65 percentage point increase in the math pass rate. The WLS *t* statistic is much larger, too: $t \approx 3.83$.

(iv) Because our model of heteroskedasticity might be wrong, it is a good idea to compute the robust standard errors for WLS. On the key variable *lexppp*, the robust standard error is about 1.82, which is somewhat higher than the usual WLS standard error. The robust standard error on *lenroll* is also somewhat higher, 1.05. That on *lunch* is slightly lower: .014 to three decimal places. For *lexppp*, the robust *t* statistic is about 3.55, which is still very statistically significant.

(v) WLS is more precise: its robust standard error is 1.82, compared with the robust standard error for OLS of 2.35. Of course, the *t* for WLS is much larger partly because the coefficient estimate is much larger. The lower standard error has an effect, too.