EC508 - Fall 2019 - Problem Set 4

Chapter 7 Question 2

The following equations were estimated using the data in BWGHT.DTA:

$$\begin{aligned}\widehat{ln(bwght)} = \underset{(0.22)}{4.66} &- \underset{(0.0009)}{0.0044}\, cigs + \underset{(0.0059)}{0.0093}\, ln(faminc) + \underset{(0.006)}{0.016}\, parity \\ &+ \underset{(0.010)}{0.027}\, male + \underset{(0.013)}{0.055}\, white\end{aligned}$$

$n = 1,388, R^2 = 0.0472$

and

$$\begin{aligned}\widehat{ln(bwght)} = \underset{(0.38)}{4.65} &- \underset{(0.0010)}{0.0052}\, cigs + \underset{(0.0085)}{0.0110}\, ln(faminc) + \underset{(0.006)}{0.017}\, parity \\ &+ \underset{(0.011)}{0.034}\, male + \underset{(0.015)}{0.045}\, white - \underset{(0.0030)}{0.0030}\, motheduc \\ &+ \underset{(0.0026)}{0.0032}\, fatheduc\end{aligned}$$

$n = 1,191, R^2 = 0.0493$

The variables are defined as in Example 4.9:

- $bwght$ = birth weight, in pounds
- $cigs$ = average number of cigarettes the mother smoked per day during pregnancy
- $parity$ = the birth order of this child
- $faminc$ = annual family income
- $motheduc$ = years of schooling for the mother
- $fatheduc$ = years of schooling for the father

but we have added a dummy variable for whether the child is male and a dummy variable indicating whether the child is classified as white.

(i)

In the first equation, interpret the coefficient on the variable *cigs*. In particular, what is the effect on birth weight from smoking 10 more cigarettes per day?

(ii)

How much more is a white child predicted to weigh than a nonwhite child, holding the other factors in the first equation fixed? Is the difference statistically significant?

(iii)

Comment on the estimated effect and statistical significance of *motheduc*.

(iv)

From the given information, why are you unable to compute the F statistic for joint significance of *motheduc* and *fatheduc*? What would you have to do to compute the F statistic?

Chapter 7 Question 6

To test the effectiveness of a job training program on the subsequent wages of workers, we specify the model

$$\ln(wage) = \beta_0 + \beta_1 train + \beta_2 educ + \beta_3 exper + u$$

where *train* is a binary variable equal to unity if a worker participated in the program. Think of the error term u as containing unobserved worker ability. If less able workers have a greater chance of being selected for the program, and you use an OLS analysis, what can you say about the likely bias in the OLS estimator of β_1? (Hint: Refer back to chapter 3.)

Chapter 7 Question C6

Use the data in SLEEP75.DTA for this exercise. The equation of interest is

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 age^2 + \beta_5 yngkid + u$$

(i)

Estimate this equation separately for men and women and report the results in the usual form. Are there notable differences in the two estimated equations?

(ii)

Compute the Chow test for equality of the parameters in the sleep equation for men and women.

Use the form of the test that adds *male* and the interaction terms $male \times totwrk, \ldots, male \times yngkid$ and uses the full set of observations. What are the relevant df for the test? Should you reject the null at the 5% level?

(iii)

Now, allow for a different intercept for males and females and determine whether the interaction terms involving *male* are jointly significant.

(iv)

Given the results from parts (ii) and (iii), what would be your final model?

Chapter 7 Question C14

Use the data in CHARITY.DTA to answer this question. The variable *respond* is a dummy variable equal to one if a person responded with a contribution on the most recent mailing sent by a charitable organization. The variable *resplast* is a dummy variable equal to one if a person responded to the previous mailing, $avggift$ is the average of past gifts (in Dutch guilders), and *propresp* is the proportion of times the person has responded to past mailings.

(i)

Estimate a linear probability model relating *respond* to *resplast* and $avggift$. Report the results in the usual form, and interpret the coefficient on *resplast*.

(ii)

Does the average value of past gifts seem to affect the probability of responding?

(iii)

Add the variable *propresp* to the model, and interpret its coefficient. (Be careful here: an increase of one in the *propresp* is the largest possible change.)

(iv)

What happened to the coefficient on *resplast* when *propresp* was added to the regression? Does this make sense?

(v)

Add *mailsyear*, the number of mailings per year, to the model. How big is its estimated effect? Why might this not be a good estimate of the causal effect of mailings on responding?

Chapter 8 Question 2

Consider a linear model to explain monthly beer conumption:

$$
\begin{aligned}
beer &= \beta_0 + \beta_1 inc + \beta_2 price + \beta_4 educ + \beta_5 female + u \\
E(u|inc, price, educ, female) &= 0 \\
Var(u|inc, price, educ, female) &= \sigma^2 inc^2
\end{aligned}
$$

Write the transformed equation that has a homoskedastic error term.

Chapter 8 Question 6

There are different ways to combine features of the Breusch-Pagan and White tests for heteroskedasticity. One possibility not covered in the text is to run the regression $\hat{u}_i^2$ on $x_{i1}, x_{i2}, \ldots, x_{ik}, \hat{y}_i^2$, $i = 1, \ldots, n$, where the $\hat{u}^i$ are the OLS residuals and the $\hat{y}^i$ are the OLS fitted values. Then, we would test joint significance of $x_{i1}, x_{i2}, \ldots, x_{ik}, \hat{y}_i^2$. (Of course, we always include an intercept in this regression.)

(i)

What are the degrees of freedom associated with the proposed 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 Breusch-Pagan 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 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?

Chapter 8 Question C4

Use VOTE1.DTA for this exercise.
(i)

Estimate a model with $voteA$ as the dependent variable and $prtystrA$, $democA$, $ln(expendA)$ and $ln(expendB)$ as independent variables. Obtain the OLS residuals $\hat{u}_i$ and regress these on all of the independent variables. Explain why you obtain $R^2 = 0$.

(ii)

Now, compute the Breusch-Pagan test for heteroskedasticity. Use the F statistic version and report the p-value.

(iii)

Compute the special case of the White test for heteroskedasticity, again using the F statistic form. How strong is the evidence for heteroskedasticity now?

Chapter 8 Question C8

Use the data set $GPA1.DTA$ for this exercise.

(i)

Use OLS to estimate a model relating $colGPA$ to $hsGPA$, ACT, $skipped$ and PC. Obtain the OLS residuals.

(ii)

Compute the special case of the White test for heteroskedasticity. In the regression of $\hat{u}_i^2$ on $\widehat{colGPA}_i$, $\widehat{colGPA}_i^2$, obtain the fitted values, say $\hat{h}_i$.

(iii)

Verify that the fitted values from part (ii) are all strictly positive. Then, obtain the weighted least squares estimates using weights $\frac{1}{\hat{h}_i}$. Compare the weighted lest 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)

In the WLS estimation from part (iii), obtain heteroskedasticity-robust standard errors. In other words, allow for the fact that the variance function estimated in part (ii) might be misspecified. (See Question 8.4) Do the standard errors change much from part (iii)?