8.6 Problem Set 3

1. Chapter 8, Q4

4 Using the data in GPA3, the following equation was estimated for the fall and second semester students:

$$\begin{aligned}
\widehat{trmgpa} = &-2.12 + .900\, crsgpa + .193\, cumgpa + .0014\, tothrs \\
&(.55)\ \ (.175) \qquad (.064) \qquad (.0012) \\
&[.55]\ \ [.166] \qquad [.074] \qquad [.0012] \\
&+ .0018\, sat - .0039\, hsperc + .351\, female - .157\, season \\
&(.0002) \qquad (.0018) \qquad (.085) \qquad (.098) \\
&[.0002] \qquad [.0019] \qquad [.079] \qquad [.080] \\
n = 269,\ &R^2 = .465.
\end{aligned}$$

Here, *trmgpa* is term GPA, *crsgpa* is a weighted average of overall GPA in courses taken, *cumgpa* is GPA prior to the current semester, *tothrs* is total credit hours prior to the semester, *sat* is SAT score, *hsperc* is graduating percentile in high school class, *female* is a gender dummy, and *season* is a dummy variable equal to unity if the student's sport is in season during the fall. The usual and heteroskedasticity-robust standard errors are reported in parentheses and brackets, respectively.

(i) Do the variables *crsgpa*, *cumgpa*, and *tothrs* have the expected estimated effects? Which of these variables are statistically significant at the 5% level? Does it matter which standard errors are used?

(ii) Why does the hypothesis H_0: $\beta_{crsgpa} = 1$ make sense? Test this hypothesis against the two-sided alternative at the 5% level, using both standard errors. Describe your conclusions.

(iii) Test whether there is an in-season effect on term GPA, using both standard errors. Does the significance level at which the null can be rejected depend on the standard error used?

2. Chapter 8, Q5

5 The variable *smokes* is a binary variable equal to one if a person smokes, and zero otherwise. Using the data in SMOKE, we estimate a linear probability model for *smokes*:

$$
\begin{aligned}
\widehat{smokes} = {} & \underset{\substack{(.855)\\ [.856]}}{.656} - \underset{\substack{(.204)\\ [.207]}}{.069}\log(cigpric) + \underset{\substack{(.026)\\ [.026]}}{.012}\log(income) - \underset{\substack{(.006)\\ [.006]}}{.029}\,educ \\
& + \underset{\substack{(.006)\\ [.005]}}{.020}\,age - \underset{\substack{(.00006)\\ [.00006]}}{.00026}\,age^2 - \underset{\substack{(.039)\\ [.038]}}{.101}\,restaurn - \underset{\substack{(.052)\\ [.050]}}{.026}\,white \\
& n = 807,\ R^2 = .062.
\end{aligned}
$$

The variable *white* equals one if the respondent is white, and zero otherwise; the other independent variables are defined in Example 8.7. Both the usual and heteroskedasticity-robust standard errors are reported.

(i) Are there any important differences between the two sets of standard errors?

(ii) Holding other factors fixed, if education increases by four years, what happens to the estimated probability of smoking?

(iii) At what point does another year of age reduce the probability of smoking?

(iv) Interpret the coefficient on the binary variable *restaurn* (a dummy variable equal to one if the person lives in a state with restaurant smoking restrictions).

(v) Person number 206 in the data set has the following characteristics: $cigpric = 67.44$, $income = 6{,}500$, $educ = 16$, $age = 77$, $restaurn = 0$, $white = 0$, and $smokes = 0$. Compute the predicted probability of smoking for this person and comment on the result.

3. Chapter 8, Q6

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

$$\hat{u}_i^2 \text{ on } x_{i1}, x_{i2}, \ldots, x_{ik}, \hat{y}_i^2, \; i = 1, \ldots, n,$$

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 $x_{i1}, x_{i2}, \ldots, x_{ik}$ and $\hat{y}_i^2$. (Of course, we always include an intercept in this regression.)

(i) What are the *df* 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 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 BP 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?

4. Chapter 8, Q7

7 Consider a model at the employee level,

$$y_{i,e} = \beta_0 + \beta_1 x_{i,e,1} + \beta_2 x_{i,e,2} + \cdots + \beta_k x_{i,e,k} + f_i + v_{i,e},$$

where the unobserved variable f_i is a "firm effect" to each employee at a given firm i. The error term $v_{i,e}$ is specific to employee e at firm i. The *composite error* is $u_{i,e} = f_i + v_{i,e}$, such as in equation (8.28).

(i) Assume that $\text{Var}(f_i) = \sigma_f^2$, $\text{Var}(v_{i,e}) = \sigma_v^2$, and f_i and $v_{i,e}$ are uncorrelated. Show that $\text{Var}(u_{i,e}) = \sigma_f^2 + \sigma_v^2$; call this σ^2.

(ii) Now suppose that for $e \neq g$, $v_{i,e}$ and $v_{i,g}$ are uncorrelated. Show that $\text{Cov}(u_{i,e}, u_{i,g}) = \sigma_f^2$

(iii) Let $\bar{u}_i = m_i^{-1}\sum_{e=1}^{mi} u_{i,e}$ be the average of the composite errors within a firm. Show that $\text{Var}(\bar{u}_i) = \sigma_f^2 + \sigma_v^2/m_i$.

(iv) Discuss the relevance of part (iii) for WLS estimation using data averaged at the firm level, where the weight used for observation i is the usual firm size.

5. Chapter 8, Q8

8 The following equations were estimated using the data in ECONMATH. The first equation is for men and the second is for women. The third and fourth equations combine men and women.

$$\widehat{score} = \underset{(3.72)}{20.52} + \underset{(0.94)}{13.60}\, colgpa + \underset{(0.150)}{0.670}\, act$$

$$n = 406.\ R^2 = .4025,\ \text{SSR} = 38{,}781.38.$$

$$\widehat{score} = \underset{(4.11)}{13.79} + \underset{(1.09)}{11.89}\, colgpa + \underset{(0.18)}{1.03}\, act$$

$$n = 408,\ R^2 = .3666,\ \text{SSR} = 48{,}029.82.$$

$$\widehat{score} = \underset{(2.80)}{15.60} + \underset{(0.73)}{3.17}\, male + \underset{(0.72)}{12.82}\, colgpa + \underset{(0.116)}{0.838}\, act$$

$$n = 814,\ R^2 = .3946,\ \text{SSR} = 87{,}128.96.$$

$$\widehat{score} = \underset{(3.91)}{13.79} + \underset{(5.55)}{6.73}\, male + \underset{(1.04)}{11.89}\, colgpa + \underset{(0.17)}{1.03}\, act + \underset{(1.44)}{1.72}\, male \cdot colgpa - \underset{(0.232)}{0.364}\, male \cdot act$$

$$n = 814,\ R^2 = .3968,\ \text{SSR} = 86{,}811.20.$$

(i) Compute the usual Chow statistic for testing the null hypothesis that the regression equations are the same for men and women. Find the p-value of the test.

(ii) Compute the usual Chow statistic for testing the null hypothesis that the slope coefficients are the same for men and women, and report the p-value.

(iii) Do you have enough information to compute heteroskedasticity-robust versions of the tests in (ii) and (iii)? Explain.