EC 771B Spring 2000 Problem Set 2

Christopher F. Baum

March 12, 2000

Due at classtime, Thursday 23 March 2000

Set up a Stata program to provide the empirical results requested. Hand in a copy of the program, annotated with your comments as warranted. The comments may be handwritten on the printout if they are clearly legible.

Set up a Stata program (do-file) to provide the empirical results requested. Hand in a copy of the program, annotated with your comments as warranted. The comments may be handwritten on the printout if they are clearly legible.

Use the Wooldridge CARD dataset, available from within Stata via the command

```
use http://fmwww.bc.edu/ec-p/data/wooldridge/CARD
```

This dataset contains 3,010 observations on the following variables:

1.	id	person identifier
2.	nearc2	=1 if near 2 yr college, 1966
3.	nearc4	=1 if near 4 yr college, 1966
4.	educ	years of schooling, 1976
5.	age	in years
6.	fatheduc	father's schooling
7.	motheduc	mother's schooling
8.	weight	NLS sampling weight, 1976
9.	momdad14	=1 if live with mom, dad at 14
10.	sinmom14	=1 if with single mom at 14
11.	step14	=1 if with step parent at 14
12.	reg661	=1 for region 1, 1966
13.	reg662	=1 for region 2, 1966
14.	reg663	=1 for region 3, 1966
15.	reg664	=1 for region 4, 1966
16.	reg665	=1 for region 5, 1966
17.	reg666	=1 for region 6, 1966
18.	reg667	=1 for region 7, 1966
19.	reg668	=1 for region 8, 1966
20.	reg669	=1 for region 9, 1966
21.	south66	=1 if in south in 1966
22.	black	=1 if black
23.	smsa	=1 in in SMSA, 1976
24.	south	=1 if in south, 1976
25.	smsa66	=1 if in SMSA, 1966
26.	wage	hourly wage in cents, 1976

```
27. enroll                  =1 if enrolled in school, 1976
28. KWW                     knowledge world of work score
29. IQ                      IQ score
30. married                 =1 if married, 1976
31. libcrd14                =1 if lib. card in home at 14
32. exper                   age - educ - 6
33. lwage                   log(wage)
34. expersq                 exper^2
```

Test the following hypotheses:

1. Log wages differ significantly across regions of the country (note that the region variable identifies where the respondent resided in 1966).

2. Log wages can be explained by age, age^2, years of education, race and SMSA (whether the respondent lived in an urban area in 1976). Comment on the expected signs for each of these explanatory variables and their estimated coefficients. Use White's general test for heteroskedasticity (-whitetst- from SSC-IDEAS) to evaluate the residuals from this equation.

3. Test whether the estimated error variances for SMSA and non-SMSA observations are equal (hint: see -sdtest-). If they are unequal, reestimate the equaion, correcting for groupwise heteroskedasticity.

4. Test whether the equation in #2 can be improved significantly by taking account of region. How do you interpret the coefficients on the region dummies?

5. Estimate the equation in #2 separately for each region (hint: -for- is handy) and comment on how its fit and estimated coefficients differ by region.

6. An alternative form of model #2 would express age and years of education in logs, and regress log wages on log(age), log(educ), race and SMSA. (a) Why is age^2 excluded? (b) Since these models (linear and log RHS) have the same dependent variable, can we compare R^2 and standard error of regression?

7. Non-nested models (the linear and log RHS of #2 and #6) can be compared by a Davidson-MacKinnon "J" test, in which each model's predicted values are added to the other. Under the hypothesis that model I is adequate, the predicted values from model II will not have a significant coefficient in model I, and vice versa. (The test need not be conclusive; neither or both predicted vectors may be significant in each other's equations). Carry out this test, and indicate whether it identifies one model as the better model.