BOSTON COLLEGE
Department of Economics
EC 228 01 Econometric Methods
Fall 2008, Prof. Baum, Ms. Phillips (tutor), Mr. Dmitriev (grader)
Problem Set 5
Due at classtime, Thursday 6 Nov 2008

Problem sets should be your own work. You may work together with classmates, but if you're not figuring this out on your own, you will eventually regret it.

**1. (C7.10)**

(i) (5 marks) The estimated equation is

$$\begin{array}{lccccccccc} points = & 4.76 & + & 1.28 & exper- & .072 & exper^2+ & 2.31 & guard+ & 1.54 & forward \\ & (1.18) & & (.33) & & (.024) & & (1.00) & & (1.00) & \end{array}$$

$$n = 269, R^2 = .091, \bar{R}^2 = .077$$

(ii) (5 marks) Including all three position dummy variables would be redundant, and result in the dummy trap. Each player falls into one of the three categories, and the overall intercept is the intercept for centers.

(iii) (5 marks) A guard is estimated to score about 2.3 points more per game, holding experience fixed. The t statistic is 2.31, so the difference is statistically different from zero at the 5% level, against a two-sided alternative.

(iv) (5 marks) When *marr* is added to the regression, its coefficient is about .584 (se=.740). Therefore, a married player is estimated to score just over half a point more per game (experience and position held fixed), but the estimate is not statistically different from zero (p-value=.43). So, based on points per game, we cannot conclude married players are more productive.

(v) (5 marks) Adding the terms $marr \cdot exper$ and $marr \cdot exper^2$ leads to complicated signs on the three terms involving *marr*. The F test for the joint significance, with 3 and 261 df, gives F= 1.44 and p-value=.23. Therefore, there is not very strong evidence that marital status has any partial effect on points scored.

(vi) (5 marks) If in the regression from part (iv) we use *assists* as the dependent variable, the coefficient on marr becomes .322 (se=.222). Therefore, holding experience and position fixed, a married man has almost one-third more assist per game. The p-value against a two-sided alternative is about .15, which is stronger, but not overwhelming, evidence that married men are more productive when it comes to assists.

**2. (8.2)**

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

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

Notice that $\beta_1$, 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.

**3. (8.4)**

(i) (5 marks) 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.

(ii) (5 marks) 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 hetero-skedasticity-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 %.

(iii) (5 marks) The in-season effect is given by the coefficient on *season*, which implies that, other things equal, an athletes 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.

**4.** (**C8.2**)

(i) (10 marks) The estimated equation with both sets of standard errors (heteroskedasticity-robust standard errors in brackets) is

$$\begin{array}{lllllll} price = & -21.77 & + \; .00207 & lotsize+ & .123 & sqrft+ & 13.85 \quad bdrms \\ & (29.48) & (.00064) & & (.013) & & (9.01) \\ & [37.13] & [.00125] & & [.017] & & [8.48] \end{array}$$

$$n = 88, R^2 = .672$$

The robust standard error on *lotsize* is almost twice as large as the usual standard error, making *lotsize* much less significant (the t statistic falls from about 3.23 to 1.70). The t statistic on *sqrft* also falls, but it is still very significant. The variable *bdrms* actually becomes somewhat more significant, but it is still barely significant. The most important change is in the significance of *lotsize*.

(ii) (10 marks) For the log-log model,

$$\begin{array}{rcccccccc} \log(\widehat{price}) = & -1.30 & + & .168 & \log(lotsize)+ & .700 & \log(sqrft)+ & .037 & bdrms \\ & (0.65) & & (.038) & & (.093) & & (.028) & \\ & [.78] & & [.041] & & [.103] & & [.030] & \end{array}$$

$$n = 54, R^2 = .643$$

Here, the heteroskedasticity-robust error is always slightly greater than the corresponding usual standard error, but the differences are relatively small. In particular, log(*lotsize*) and log(*sqrft*) still have very large t statistics, and the t statistic on bdrms is not significant at the 5% level against a one-sided alternative using either standard error.

(iii) (5 marks) As we discussed in Section 6.2, using the logarithmic transformation of the dependent variable often mitigates, if not entirely eliminates, heteroskedasticity. This is certainly the case here, as no important conclusions in the model for log(*price*) depend on the choice of the standard error. (We have also transformed two of the independent variables to make the model of the constant elasticity variety in *lotsize* and *sqrft*.

**5.** (**C8.4**)

(i) (10 marks) The estimated equation is

$$\begin{array}{rcccccccc} voteA = & 37.66 & + & .252 & prtystrA+ & 3.793 & democA+ & 5.779 & \log(expendA) \\ & (4.74) & & (.071) & & (1.407) & & (.392) & \\ & & - & 6.238 & \log(expendB) & +\hat{u} & & & \\ & & & (.397) & & & & & \end{array}$$

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

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) (5 marks) The B-P test entails regressing on the $\hat{u}_i^2$ 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) (5 marks) 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.