BOSTON COLLEGE
Department of Economics
EC 228 Econometrics, Prof. Baum, Mr. Barbato, Spring 2003
Problem Set 5.
Answer sheet

**Problem 7.13**

The estimated equation is:

$$\widehat{\log(salary)} = \underset{(0.29)}{4.30} + \underset{(.034)}{.288} \log(sales) + \underset{(.0040)}{.0167}\ roe - \underset{(.109)}{.226}\ rosneg$$
$$n = 209,\ R^2 = .297,\ \bar{R}^2 = .286$$

The coefficient on *rosneg* implies that if the CEO's firm had a negative return on its stock over the 1988 to 1990 period, the CEO salary was predicted to be about 22.6% lower, for given levels of *sales* and *roe*. The t statistic is about -2.07, which is significant at the 5% level against a two sided alternative.

**Problem 7.15**

(i) When *educ* = 12.5, the approximate proportionate difference in estimated *wage* between women and men is $-.227 - .0056(12.5) = -.297$. When *educ* = 0, the difference is -.227. So the differential at 12.5 years of education is about 7 percentage points greater.

(ii) We can write the model underlying (7.18) as

$$\log(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta female \bullet educ + other factors$$
$$= \beta_0 + (\delta_0 + 12.5\delta_1) female + \beta_1 educ + \delta_1 female \bullet (educ - 12.5) + other factors$$
$$= \beta_0 + \theta_0 female + \beta_1 educ + \delta_1 female \bullet (educ - 12.5) + other factors,$$

where $\theta_0 \equiv \delta_0 + 12.5\delta_1$ is the gender differential at 12.5 years of education. When we run this regression we obtain about -.294 as the coefficient on *female* (which differs from -.297 due to rounding error). Its standard error is about .036.

(iii) The *t* statistic on *female* from part (ii) is about -8.17, which is very significant. This is because we are estimating the gender differential at a reasonable number of years of education, 12.5, which is close to the average. In equation (7.18), the coefficient on *female* is the gender differential when *educ* = o. There are no people of either gender with close to zero years of education, and so we cannot hope - nor do we want to - to estimate the gender differential at *educ* = 0.

**Problem 7.18**

(i) The estimated equation is

$$\widehat{points} = \underset{(1.18)}{4.76} + \underset{(.33)}{1.28}\ exper - \underset{(.024)}{.072}\ exper^2 + \underset{(1.00)}{2.31}\ guard + \underset{(1.00)}{1.54}\ forward$$
$$n = 269,\ R^2 = .091,\ \bar{R}^2 = .077.$$

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

(iii) A gueard 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 zeroat the 5% level, against a two sided alternative.

(iv) 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 = .23. So, based on points per game, we cannot conclude married players are more productive.

(v) Adding the terms *marr•exper* and *marr•exper*$^2$ leads to complicated signs on the three terms involving *marr*. The F test for their 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) 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 assists 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.

**Problem 8.7**

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

$$\widehat{price} = \underset{\substack{(29.48)\\ [36.28]}}{-21.77} + \underset{\substack{(.00064)\\ [.00122]}}{.00207}\ lotsize + \underset{\substack{(.013)\\ [.017]}}{.123}\ sqrft + \underset{\substack{(9.01)\\ [8.28]}}{13.85}\ bdrms$$

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

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 about 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) For the log-log model,

$$\widehat{\log(price)} = \underset{\substack{(0.65)\\ [0.76]}}{5.61} + \underset{\substack{(.038)\\ [.041]}}{1.68}\ \log(lotsize) + \underset{\substack{(.093)\\ [.101]}}{.700}\ \log(sqrft) + \underset{\substack{(.028)\\ [.030]}}{.037}\ bdrms$$

$$n = 88,\ R^2 = .643.$$

Here, the heteroskedasticity-robust standard 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) 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 standard error. (We have also transformed two of the independent variables to make the model of the constant elasticity variety in *lotsize* and *sqrft*.)

**Problem 8.9**

(i) The estimated equation is

$$\widehat{voteA} = \underset{(4.74)}{37.66} + \underset{(.071)}{.252}\ prtystrA + \underset{(1.407)}{3.793}\ democA + \underset{(0.392)}{5.779}\ \log(expendA) - \underset{(0.397)}{6.238}\ \log(expendB) + \hat{\mu}$$

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

You can convince yourslef that regressing the $\widehat{\mu_i}$ 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, this is how OLS works: the estimates $\widehat{\beta_j}$ are chosen to make the residuals be uncorrelated in the sample with each independent variable (as well as have zero sample average).

(ii) The B-P test entail regressing the $\widehat{\mu_i}$ on the independent variables in part (i). The F statistic for joint significance (with 4 and 168 *df*) is about 2.33 with *p*-value ≈ .058. Therefore, there is some evidence of the heteroskedasticity but not quite at the 5% level.

(iii) Now we regress $\widehat{\mu_i^2}$ 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 ≈ .065. This is slightly less evidence of heteroskedasticity than provided by the B-P test, but the conclusion is very similar.

**Problem 9.7**

(i) We estimate the model from column (2) but with KWW in place of IQ. The coefficient on *educ* becomes about .058 (se ≈ .006), so this is similar to the estimate obtained with IQ, although slightly largerand more precisely estimated.

(ii) When KWW and IQ are both used as proxies, the coefficient on *educ* becomes about .049 (se ≈ .007). Compared with the estimate when only KWW is used as a proxy, the return to education has fallen by almost a full percentage point.

(iii) The $t$ statistic on IQ is about 3.08 while that on KWW is about 2.07, so each is significant at the 5% level against a two-sided alternative. They are jointly significant, with $F_{2,925} \approx 8.59$ and $p$-value ≈ .0002.