BOSTON COLLEGE
Department of Economics
EC 228 Econometrics, Prof. Baum, Mr. Barbato, Spring 2003

**Problem Set 3**

**Problem 3.4**

(i) A larger rank for a law school means that the school has less prestige; this lowers starting salaries. For example, a rank of 100 means there are 99 schools thought to be better.

(ii) $\beta_1 > 0, \beta_2 > 0$. Both $LSAT$ and $GPA$ are measures of the quality of the entering class. No matter where better students attend law school, we expect them to earn more, on average. $\beta_3, \beta_4 > 0$. The number of volumes in the law library and the tuition cost are both measures of the school quality. (Cost is less obvious than library volumes, but should reflect quality of the faculty, physical plant, and so on.)

(iii) This is just the coefficient on GPA, multiplied by 100: 24.8%

(iv) This is an elasticity: a one percent increase in library volumes implies a .095% increase in predicted median starting salary, other things equal.

(v) It is definitely better to attend a law school with a lower rank. If law school A has a ranking 20 less than law school B, the predicted difference in starting salary is 100(.0033)(20) = 6.6% higher for law school A.

**Problem 3.14**

(i) In the dataset, *price* is measured in dollars. The estimated equation is:

$$\widehat{price} = -19315 + 128.43 sqrft + 15198 bdrms$$

$$n = 88, R^2 = .632$$

(ii) Holding square footage constant, $\Delta\widehat{price} = 15198 bdrms$, and so $\widehat{price}$ increases by \$15,198.

(iii) Now $\Delta\widehat{price} = 128.43\Delta sqrft + 15198\Delta bdrms = 128.43(149) + 15198 =$ \$33,120. Because the size of the house is increasing, this is a much larger effect than in (ii).

(iv) About 63.2%.

(v) The predicted value is $-19315 + 128.43(2,438) + 15198(4) =$ \$353,544

(iv) From part (v), the estimated value of the home based only on square footage and number of bedrooms is \$353,544. The actual selling price was \$300,000, which suggest the buyer underpaid by some margin. But, of course, there are many other features of a house (some that we cannot even measure) that affect price, and we have not controlled for these.

**Problem 3.18**

(i) The slope coefficient from the regression IQ on *educ* is approximately $\delta_1 =$ 3.53383.

(ii) The slope coefficent from log(wage) on *educ is* $\beta_1 = .05984$

(iii) The slope coefficient from log(wage) on *educ*, IQ are $\beta_1 = .03912$ and $\beta_2 = .00586$, respectively.

**Problem 4.8**

(i) We use property VAR.3 from Appendix B: Var($\beta_1$- $\beta_2$) = Var($\beta_1$) + 9Var($\beta_2$) - 6Cov($\beta_1, \beta_2$).

(ii)$t = (\beta_1 - 3\beta_2 - 1)/se(\beta_1 - 3\beta_2)$,where *se* stands for the standard error of the expression in parentheses.

(iii) Because $\theta_1 = \beta_1 - 3\beta_2$,we can write $\beta_1 = \theta_1 + 3\beta_2$. Plugging this into the population model gives

$$\begin{aligned} y &= \beta_0 + (\theta_1 + 3\beta_2)x_1 + \beta_2 x_2 + \beta_3 x_3 + u \\ &= \beta_0 + \theta_1 x_1 + \beta_2(3x_1 + x_2) + \beta_3 x_3 + u \end{aligned}$$

This last equation is what we would estimate by regressing $y$ on $x_1$,$3x_1+x_2$,and $x_3$.The coefficient and standard error on $x_1$are what we want.

**Problem 4.9**

(i) With degrees of freedom $= 706 - 4 = 702$, we use the standard normal critical value ($df = \infty$ in Table G.2), which is 1.96 for a two-tailed test at the 5% level. Now $t_{educ} = -11.13/5.88 \approx -1.89$,so $|t_{educ}| = 1.89 < 1.96$, and we fail to reject $H_0 : \beta_{educ} = 0$ at the 5% level. Also, $t_{age} \approx 1.52$,so age is also statistically insignificant at the 5% level.

(ii) We need to compute the R-squared from the F statistic for joint significance. But $F = [(.113 - .103)/(1 - .113)](702/2) \approx 3.96$.The 5% critical value in the $F_{2,702}$can be obtained from Table G.3b with denominator $df = \infty : cv = 3.00$. Therefore, *educ* and *age* are jointly significant at the 5% level (3.96 > 3.00). In fact, the *p*-value is about .019, and so *educ* and *age* are jointly significant at the 2% level.

(iii) Not really. These variables are jointly significant, but including them only changes the coefficient on *totwrk* from $-.151$ to $-.148$.

(iv) The standard t and F statistics that we used assume homoskedasticity, in addition to the other CLM assumptions. If there is heteroskedasticity in the equation, the tests are no longer valid.

**Problem 4.13**

(i) In the model

$$\log(salary) = \beta_0 + \beta_1 LSAT + \beta_2 GPA + \beta_3 \log(libvol) + \beta_4 \log(\cos t) + \beta_5 rank + u,$$

the hypothesis that *rank* has no effect on log(*salary)* is $H_0 : \beta_5 = 0$. The estimated equation (now with standard errors) is

$$\begin{aligned} \log(salary) &= \underset{(0.53)}{8.34} + \underset{(.0040}{.0047}\, LSAT + \underset{(.090)}{.248}\, GPA + \underset{(.033)}{.095} \log(libvol) + \underset{(.032)}{.038} \log(\cos t) - \underset{(.0003)}{.0033}\, rank \\ n &= 136,\ and\ R^2 = .842 \end{aligned}$$

The t statistic on *rank* is $-11$, which is very significant. If *rank* decreases by 10 (which is a move up for a law school), median starting salary is predicted to increase by about 3.3%.

(ii)LSAT is not statistically significant ( *t statistic* $\approx$ 1.18) but GPA is very significant ( *t statistic* $\approx$ 2.76). The test for joint significance is moot given that GPA is so significant, but for completeness the F statistic is about 9.95(with 2 and 130 *df)* and *p*-value $\approx$ .0001.

(iii)When we add *clsize* and *faculty* to the regression we lose 5 observations. The test of their joint significance (with 2 and 131 - 8 = 123 *df*) gives $F \approx .95$ and $p - value \approx .39$.So these two variables are not jointly significant unless we use a very large significance level.

(iv) If we want to just determinethe effect of numerical rankings on starting law school salaries, we should control for other factors that affect salaries and rankings. The idea is that there is some randomness in rankings, or the rankings might depend partly on frivolous factors that do not affect quality of the students. LSAT and GPA are perhaps good controls for student quality. However, if there are differences in genderand racial composition across schools, and systematic gender and race differences in salaries, we could also control for these. However, it is unclear why these would be correlated with *rank*. Faculty quality, as perhaps measured by publication records, could be included. Such things do enter rankings of law schools.

**Problem 4.14.**

(i) The estimated model is

$$\begin{aligned}\log(price) &= \underset{(0.10)}{11.67} + \underset{(.000043)}{.000379}\, sqrft + \underset{(.0296)}{.0289}\, bdrms \\ n &= 88.\ R^2 = .588.\end{aligned}$$

Therefore, $\theta_1 = 150(.000379) + .0289 = .0858$,which means that an additional 150 square foot bedroom increases the predicted price by about 8.6%

(ii) $\beta_2 = \theta_1 - 150\beta_1$,and so

$\log(price) = \beta_0 + \beta_1 sqrft + (\theta_1 - 150\beta_1)bdrms + u$

$\log(price) = \beta_0 + \beta_1(sqrft - 150bdrms) + \theta_1 bdrms + u$

(iii) From part (ii), we run the regression

$\log(price)$ on $(sqrft - 150bdrms)$ and $bdrms$,

and obtain the standard error on $bdrms$. We already know that $\widehat{\theta_1} = .0858$; now we also get $se(\widehat{\theta_1}) = .268$. The 95% confidence interval is .0326 to .1390 (or about 3.3% to 13.9%).

Alternatively, use Stata's `lincom` command to evaluate the expression (and compute its confidence interval), via `lincom 150*_b[sqrft] + _b[bdrms]`.

**Problem 4.17**

(i) In the model

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 \exp er \ + \beta_3 tenure + u$$

the null hypothesis of interest is $H_0 : \beta_2 = \beta_{3.}$

(ii)Let $\theta_2 = \beta_2 - \beta_3$.Then we can estimate the equation

$$\log(wage) = \beta_0 + \beta_1 educ + \theta_2 \exp er + \beta_3(\exp er + tenure) + u$$

to obtain the 95% CI for $\theta_2$. This turns out to be about $.0020 \pm 1.96(,0047)$,or about $-.0072$ to .0112. Because zero is in this CI, $\theta_2$is not statistically different from zero at the 5% level, and we fail to reject $H_0 : \beta_2 = \beta_3$ at the 5% level.