BOSTON COLLEGE
Department of Economics
EC 228 01 Econometric Methods
Fall 2008, Prof. Baum, Ms. Phillips (tutor), Mr. Dmitriev (grader)
Problem Set 4
Due at classtime, 23 October 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. (6.3)

(i) (5 marks) The turnaround point is given by $\hat{\beta}_1/(2|\hat{\beta}_2|)$, or $.0003/.000000014 \approx 21,428.57$; remember, this is sales in millions of dollars.

(ii) (5 marks) Probably. Its t-statistic is about -1.89, which is signigicant against the one-sided alternative $H_0 : \beta_1 < 0$ at the 5 % level ($cv \approx -1.70$ with $df = 29$). In fact, the p-value is about .036.

(iii) (10 marks) Because *sales* gets divided by 1,000 to obtain *salesbil*, the corresponding coefficient gets multiplied by 1,000: $1,000 \cdot .00030 = .30$. The standard error gets multiplied by the same factor. As stated in the hint, $salesbil^2 = sales/1,000,000$, and so the coefficient on the quadratic gets multiplied by one million: $1,000,000 \cdot .0000000070 = .0070$; its standard error also gets multiplied by one million. Nothing happens to the intercept (because *rdintens* has not been rescaled) or to the R^2:

$$\widehat{rdintens} = \underset{(0.429)}{2.613} + \underset{(.14)}{.30}\ salesbil - \underset{(.0037)}{.0070}\ salesbil^2$$

$$n = 32, R^2 = .1484.$$

(iv) (5 marks) The equation in part (iii) is easier to read because it contains fewer zeros to the right of the decimal. Of course the interpretation of the two equations is identical once the different scales are accounted for.

2. (**6.4**)

(i) (5 marks) Holding all other factors fixed we have

$$\Delta \log(wage) = \beta_1 \Delta educ + \beta_2 \Delta educ \cdot pareduc = (\beta_1 + \beta_2 pareduc)\Delta educ.$$

Dividing both sides by $\Delta educ$ gives the result. The sign of β_2 is not obvious, although $\beta_2 > 0$ if we think a child gets more out of another year of education the more highly educated are the childs parents.

(ii) (5 marks)We use the values $pareduc = 32$ and $pareduc = 24$ to interpret the coefficient on $educ \cdot pareduc$. The difference in the estimated return to education is .00078(32 - 24) = .0062, or about .62 percentage points.

(iii) (5 marks) When we add *pareduc* by itself, the coefficient on the interaction term is negative. The t- statistic on $educ \cdot pareduc$ is about $-0.0016/0.0012 = 1.33$, which is not significant at the 10% level against a two-sided alternative. Note that the coefficient on *pareduc* is significant at the 5% level against a two-sided alternative. This provides a good example of how omitting a level effect (*pareduc* in this case) can lead to biased estimation of the interaction effect.

3. (**C6.2**)

(i) (5 marks) The estimated equation is

$$\begin{array}{lccccccc} \widehat{\log(wage)} = & .128 & + & .0904 & educ+ & .0410 & exper- & .000714 & exper^2 \\ & (.106) & & (.0075) & & (.0052) & & (.000116) & \end{array}$$

$$n = 526, R^2 = .300, \bar{R}^2 = .296$$

(ii) (5 marks) The t-statistic on $exper^2$ is about 6.16, which has a p-value of essentially zero. So *exper* is significant at the 1% level(and much smaller significance levels).

(iii) (5 marks) To estimate the return to the fifth year of experience, we start at $exper = 4$ and increase $exper$ by one, so $\Delta exper = 1$:

$$\%\Delta\widehat{wage} \approx 100[.0410 - 2(.000714)4] \approx 3.53\%$$

Similarly, for the 20th year of experience,

$$\%\Delta\widehat{wage} \approx 100[.0410 - 2(.000714)19] \approx 1.39\%$$

(iv) (5 marks) The turnaround point is about $.041/[2(.000714)] \approx 28.7$ years of experience. In the sample, there are 121 people with at least 29 years of experience. This is a fairly sizeable fraction of the sample.

4. (**C6.3**)

(i) (5 marks) Holding $exper$ (and the elements in u) fixed, we have

$$\Delta\log(wage) = \beta_1\Delta educ + \beta_3\Delta educ \cdot exper = (\beta_1 + \beta_3 exper)\Delta educ,$$

or

$$\frac{\Delta\log(wage)}{\Delta educ} = (\beta_1 + \beta_3 exper)$$

This is the approximate proportionate change in wage given one more year of education.

(ii) (5 marks) $H_0 : \beta_3 = 0$. If we think that education and experience interact positively so that people with more experience are more productive when given another year of education then $\beta_3 > 0$ is the appropriate alternative.

(iii) (10 marks) The estimated equation is

$$\widehat{\log(wage)} = \underset{(.24)}{5.95} + \underset{(.0174)}{.0440}\ educ - \underset{(.0200)}{.0215}\ exper + \underset{(.00153)}{.00320}\ educ \cdot exper$$

$$n = 935, R^2 = .135, \bar{R}^2 = .132$$

The t-statistic on the interaction term is about 2.13,which gives a p-value below .02 against $H_1 : \beta_3 > 0$. Therefore, we reject $H_0 : \beta_3 = 0$ at the 2 % level.

(iv) (5 marks) We rewrite the equation as

$$\log(wage) = \beta_0 + \theta_1 educ + \beta_2 exper + \beta_3 educ(exper - 10) + u,$$

and run the regression log($wage$) on $educ$, $exper$, and $educ(exper10)$. We want the coefficient on $educ$. We obtain $\hat{\theta}_1 \approx .0761$ and $se\left(\hat{\theta}_1\right) \approx .0066$. The 95 % CI for θ_1 is about .063 to .089.

5. (**C6.8**)

(i) (5 marks) The estimated equation (where $price$ is in dollars) is

$$\begin{array}{rllllll} \widehat{price} = & -21,770.3 & + & 2.068 & lotsize+ & 122.78 & sqrft+ & 13,852.5 & bdrms \\ & (29,475.0) & & (0.642) & & (13.24) & & (9,010.1) & \end{array}$$

$$n = 88, R^2 = .672, \bar{R}^2 = .661, \hat{\sigma} = 59,833$$

The predicted price at $lotsize = 10,000$, $sqrft = 2,300$, and $bdrms = 4$ is about \$336,714.

(ii) (5 marks) The regression is $price_i$ on $(lotsize_i 10,000)$, $(sqrft_i 2,300)$, and $(bdrms_i 4)$. We want the intercept estimate and the associated 95% CI from this regression. The CI is approximately $336,706.7 \pm 14,665$, or about \$322,042 to \$351,372 when rounded to the nearest dollar.

(iii) (10 marks) We must use equation (6.36) to obtain the standard error of $\hat{e}^0$ and then use equation (6.37) (assuming that $price$ is normally distributed). But from the regression in part (ii), $se(\hat{y}^0) \approx 7,374.5$ and $\hat{\sigma} \approx 59,833$. Therefore, $se\left(\hat{e}^0\right) \approx \left[(7,374.5)^2 + (59,833)^2\right]^{1/2} \approx 60,285.8$. Using 1.99 as the approximate 97.5^{th} percentile in the $t_8 4$ distribution gives the 95% CI for $price^0$, at the given values of the explanatory variables, as $336,706.7 \pm 1.99(60,285.8)$ or, rounded to the nearest dollar, \$216,738 to \$456,675. This is a fairly wide prediction interval. But we have not used many factors to explain housing price. If we had more, we could, presumably, reduce the error standard deviation, and therefore $\hat{\sigma}$, to obtain a tighter prediction interval.

6. (**7.2**)

(i) (5 marks) If $\Delta cigs = 10$ then $\Delta \log(bwght) = -.0044 \cdot (10) = -.044$, which means about a 4.4 % lower birth weight.

(ii) (5 marks) A white child is estimated to weigh about 5.5 % more, other factors in the first equation fixed. Further, $t_{white} \approx 4.23$, which is well above any commonly used critical value. Thus, the difference between white and nonwhite babies is also statistically significant.

(iii) (5 marks) If the mother has one more year of education, the childs birth weight is estimated to be .3 % lower. This is not a huge effect, and the t-statistic is only one, so it is not statistically significant.

(iv) (10 marks) The two regressions use different sets of observations. The second regression uses fewer observations because *motheduc* or $fatheduc$ are missing for some observations. We would have to reestimate the first equation (and obtain the R-squared) using the same observations used to estimate the second equation.

7. (**7.4**)

(i) (5 marks) The approximate difference is just the coefficient on *utility* times 100, or 28.3 %. The t-statistic is $-.283/.099 \approx -2.86$, which is very statistically significant.

(ii) (5 marks) $100[exp(-.283)1) \approx -24.7\%$, and so the estimate is somewhat smaller in magnitude.

(iii) (10 marks) The proportionate difference is .181 - .158 = .023, or about 2.3 %. One equation that can be estimated to obtain the standard error of this difference is

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \delta_1 consprod + \delta_2 utility + \delta_3 trans + u,$$

where *trans* is a dummy variable for the transportation industry. Now, the base group is $finance$, and so the coefficient δ_1 directly measures the difference between the consumer products and finance industries, and we can use the t-statistic on *consprod*.

8. (**C7.2**)

(i) (10 marks) The estimated equation is

$$\begin{array}{lcccccccc}
\log(\widehat{wage}) = & 5.40 & + & .0654 & educ+ & .0140 & exper+ & .0117 & tenure \\
& (.11) & & (.0063) & & (.0032) & & (.0025) & \\
+ & .199 & married- & .188 & black- & .091 & south+ & .184 & urban \\
& (.039) & & (.038) & & (.026) & & (.027) &
\end{array}$$

$$n = 935, R^2 = .253.$$

The coefficient on *black* implies that, at given levels of the other explanatory variables, black men earn about 18.8 % less than nonblack men. The t-statistic is about 4.95, and so it is very statistically significant.

(ii) (5 marks) The F-statistic for joint significance of $exper^2$ and $tenure^2$, with 2 and 925 df, is about 1.49 with p-value $\approx$ .226. Because the p-value is above .20, these quadratics are jointly insignificant at the 20 % level.

(iii) (10 marks) We add the interaction $black \cdot educ$ to the equation in part (i). The coefficient on the interaction is about -.0226 (se $\approx$ .0202). Therefore, the point estimate is that the return to another year of education is about 2.3 percentage points lower for black men than nonblack men. (The estimated return for nonblack men is about 6.7 %.) This is nontrivial if it really reflects differences in the population. But the t statistic is only about 1.12 in absolute value, which is not enough to reject the null hypothesis that the return to education does not depend on race.

(iv) (10 marks) We choose the base group to be single, nonblack. Then we add dummy variables *marrnonblck*, *singblck*, and *marrblck* for the other three groups. The result is

$$\begin{array}{lccccccc}
\log(\widehat{wage}) = & 5.40 & +.0655 & educ+ & .0141 & exper+ & .0117 & tenure \\
& (.11) & (.0063) & & (.0032) & & (.0025) & \\
& & -.092 & south+ & .184 & urban+ & .189 & marrnonblck \\
& & (.026) & & (.027) & & (.043) & \\
& & -.241 & singblck+ & .0094 & marrblck & & \\
& & (.096) & & (.0560) & & &
\end{array}$$

$$n = 935, R^2 = .253.$$

We obtain the ceteris paribus differential between married blacks and married nonblacks by taking the difference of their coefficients: .0094 - .189 = -.1796, or about -.18. That is, a married black man earns about 18 % less than a comparable, married nonblack man.

9. (**C7.6**)

(i) (10 marks) The estimated equation for men is

$$\begin{array}{cccccccc}
\widehat{sleep} = & 3{,}648.2 & - & .182 & totwrk- & 13.05 & educ \\
 & (310.0) & & (.024) & & (7.41) & \\
+ & 7.16 & age- & .0448 & age^2+ & 60.38 & yngkid \\
 & (14.32) & & (.1684) & & (59.02) &
\end{array}$$

$$n = 400, R^2 = .156$$

The estimated equation for women is

$$\begin{array}{cccccccc}
\widehat{sleep} = & 4{,}238.7 & - & .140 & totwrk- & 10.21 & educ \\
 & (384.9) & & (.028) & & (9.59) & \\
- & 30.36 & age- & .368 & age^2- & 118.28 & yngkid \\
 & (18.53) & & (.223) & & (93.19) &
\end{array}$$

$$n = 306, R^2 = .098$$

There are certainly notable differences in the point estimates. For example, having a young child in the household leads to less sleep for women (about two hours a week) while men are estimated to sleep about an hour more. The quadratic in *age* is a hump-shape for men but a U-shape for women. The intercepts for men and women are also notably different.

(ii) (5 marks) The F statistic (with 6 and 694 df) is about 2.12 with *p-value* $\approx .05$, and so we reject the null that the sleep equations are the same at the 5 % level.

(iii) (5 marks) If we leave the coefficient on *male* unspecified under H_0, and test only the five interaction terms, $male \cdot totwork$, $male \cdot educ$, $male \cdot age$, $male \cdot age^2$, and $male \cdot yngkid$, the F statistic (with 5 and 694 df) is about 1.26 and *p-value* $\approx .28$.

(iv) (10 marks) The outcome of the test in part (iii) shows that, once an intercept difference is allowed, there is not strong evidence of slope differences between men and women. This is one of those cases where the practically important differences in estimates for women and men in part (i) do not translate into statistically significant differences. We need a larger sample size to confidently determine whether there are differences in slopes. For the purposes of studying the sleep-work tradeoff, the original model with *male* added as an explanatory variable seems sufficient.