BOSTON COLLEGE
Department of Economics
EC 228 Econometrics, Prof. Baum, Ms. Yu, Fall 2003

Problem Set 4 Solutions

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. (**4.2**)

(i) $H_0 : \beta_3 = 0$. $H_1 : \beta_3 > 0$.

(ii) The proportionate effect on *salary* is $.00024(50) = .012$. To obtain the percentage effect, we multiply this by 100: 1.2%. Therefore, a 50 point ceteris paribus increase in *ros* is predicted to increase salary by only 1.2%. Practically speaking this is a very small effect for such a large change in *ros*.

(iii) The 10% critical value for a one-tailed test, using $df = \infty$, is obtained from Table G.2 as 1.282. The t statistic on *ros* is $.00024/.00054 \approx .44$, which is well below the critical value. Therefore, we fail to reject H_0 at the 10% significance level.

(iv) Based on this sample, the estimated *ros* coefficient appears to be different appears to be different from zero only because of sampling variation. One the other hand, including ros may not be causing any harm; it depends on how correlated it is with the other independent variables (although these are very significant even with *ros* in the equation).

2. (**4.3**)

(i) Holding *profmarg* fixed, $\Delta \widehat{rdintens} = .321 \Delta \log(sales) = (.321/100)[100 \cdot \Delta \log(sales)] \approx .00321(\%\Delta sales)$. Therefore, if $\%\Delta sales = 10$, $\Delta \widehat{rdintens} \approx .032$, or only about 3/100 of a percentage point. For such a large percentage increase in sales, this seems like a practically small effect.

(ii) $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 > 0$, where β_1 is the population slope on $\log(scales)$. The t statistic is $.321/.216 \approx 1.486$. The 5% critical value for a one-tailed test, with $df = 32 - 3 = 29$, is obtained from Table G.2 as 1.699; so we cannot reject H_0 at the 5% level. But the 10% critical value is 1.311; since the t statistic is above this value, we reject H_0 in favor of H_1 at the 10% level.

(iii) Not really. Its t statistic is only 1.087, which is well below even the 10% critical value for a one-tailed test.

3. (**4.5**)

(i) $.412 \pm 1.96(.094)$, or about .228 to .596.

(ii) No, because the value .4 is well inside the 95% CI.

(iii) Yes, because 1 is well outside the 95% CI.

4. (**4.7**)

(i) While the standard error on hrsemp has not changed, the magnitude of the coefficient has increased by half. The t statistics on *hrsemp* has gone from about -1.47 to -2.21, so now the coefficient is statistically less than zero at the 5% level. (From Table G.2 the 5% critical value with 40 *df* is -1.684. The 1% critical value is -2.423, so the p-value is between .01 and .05).

(ii) If we add and subtract $\beta_2 \log(employ)$ from the right-hand-side and collect terms, we have

$$\begin{aligned}\log(scrap) &= \beta_0 + \beta_1 hrsemp + [\beta_2 \log(sales) - \beta_2 \log(employ)] \\ &\quad +[\beta_2 \log(employ) + \beta_3 \log(employ)] + u \\ &= \beta_0 + \beta_1 hrsemp + \beta_2 \log(sales/employ) \\ &\quad +(\beta_2 + \beta_3) \log(employ) + u,\end{aligned}$$

where the second equality follows from the fact that $\log(sales/employ) = \log(sales) - \log(employ)$. Defining $\theta_3 = \beta_2 + \beta_3$ gives the result.

(iii) No, we are interested in the coefficient on $\log(employ)$, which has a t statistic of .2, which is very small. Therefore, we conclude that the size of the firm, as measured by employees, does not matter, once we control for training *and* sales per employee (in a logarithmic functional form).

(iv) The null hypothesis in the model from part (ii) is $H_0 : \beta_2 = -1$. The t statistic is $[-.951-(-1)]/.37 = (1-.951)/.37 \approx .132$; this is very small, and we fail to reject whether we specify a one- or two-sided alternative.

5. (**4.9**)

(i) With $df = 706 - 4 = 702$, 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 of the F statistic for joint significance. But $F = [(.113 - .103)/(1 - .113)](702/2) \approx 3.96$. The 5% critical value in the $F_{2,702}$ distribution 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. This 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.

6. (**4.10**)

(i) We need to compute the F statistic for the overall significance of the regression with $n = 142$ and $k = 4$: $F = [.0395/(1 - .0395)](137/4) \approx 1.41$. The 5% critical value with 4 numerator df and using 120 for the numerator df, is 2.45, which is well above the value of F. Therefore, we fail to reject $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ at the 10% level. No explanatory variable is individually significant at the 5% level. The largest absolute t statistic is on dkr, $t_{dkr} \approx 1.60$, which is not significant at the 5% level against a two-sided alternative.

(ii) The F statistic (with the same df) is now $[.330/(1 - .0330)](137/4) \approx 1.17$, which is even lower than in part (i). None of the t statistic is significant at a reasonable level.

(iii) Because observation of a firm's debt to capital ratio, i.e., dkr and the earning per share eps, can be negative, we can not use the logs of dkr and eps in part (ii). If we only take those observations with positive dkr and eps, the sampling will not be random.

(iv) It seems very weak. There are no significant t statistics at the 5% level (against a two-sided alternative), and the F statistics are insignificant in both cases. Plus, less than 4% of the variation in $return$ is explained by the independent variables.

7. (**4.12**)

(i) Holding other factors fixed,

$$\begin{aligned} \Delta voteA &= \beta_1 \Delta \log(expendA) = (\beta_1/100)[100 \cdot \Delta \log(expendA)] \\ &\approx (\beta_1/100)(\%\Delta expendA), \end{aligned}$$

where we use the fact that $100 \cdot \Delta \log(expendA) \approx \%\Delta expendA$. So $\beta_1/100$ is the (ceteris paribus) percentage point change in $voteA$ when $expendA$ increases by one percent.

(ii) The null hypothesis is $H_0 : \beta_2 = -\beta_1$, which means a $z\%$ increase in expenditure by A and a $z\%$ increase in expenditure by B leaves $voteA$ unchanged. We can equivalently write $H_0 : \beta_1 + \beta_2 = 0$.

(iii)

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/VOTE1

. regress  voteA lexpendA lexpendB prtystrA

      Source |       SS       df       MS              Number of obs =     173
-------------+------------------------------           F(  3,   169) =  215.23
       Model |  38405.1089     3   12801.703           Prob > F      =  0.0000
    Residual |  10052.1396   169  59.4801161           R-squared     =  0.7926
-------------+------------------------------           Adj R-squared =  0.7889
       Total |  48457.2486   172  281.728189           Root MSE      =  7.7123
```

```
------------------------------------------------------------------------------
       voteA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lexpendA |   6.083316     .38215    15.92   0.000     5.328914    6.837719
    lexpendB |  -6.615417   .3788203   -17.46   0.000    -7.363246   -5.867588
    prtystrA |   .1519574   .0620181     2.45   0.015     .0295274    .2743873
       _cons |   45.07893   3.926305    11.48   0.000     37.32801    52.82985
------------------------------------------------------------------------------
```

The estimated equation (with standard errors in parentheses below estimates) is

$$\widehat{voteA} = \underset{(3.93)}{45.08} + \underset{(.382)}{6.083} \log(expendA) - \underset{(.379)}{6.615} \log(expendB) + \underset{(.062)}{.152}\, prtystrA$$

$$n = 173, R^2 = .793.$$

The coefficient on $\log(expendA)$ is very significant ($t$statistic ≈ 15.92), as is the coefficient on $\log(expendB)$ ($t$statistic ≈ -17.45). The estimates imply that a 10% ceteris paribus increase in spending by candidate A increase the predicted share of the vote going to A by about .61 percentage points. [Recall that, holding other factors fixed, $\Delta\widehat{voteA} \approx (6.083/100)\%\Delta expendA$.] Similarly, a 10% ceteris paribus increase in spending by B reduces $\widehat{voteA}$ by about .66 percentage points. These effects certainly cannot be ignored.

While the coefficients on $\log(expendA)$ and $\log(expendB)$ are of similar magnitudes (and opposite in sign, as we expect), we do not have the standard error of $\hat{\beta}_1 + \hat{\beta}_2$, which is what we would need to test the hypothesis from part (ii).

(iv) Write $\theta_1 = \beta_1 + \beta_2$, or $\beta_1 = \theta_1 - \beta_2$. Plugging this into the original equation, and rearranging, gives

$$\widehat{voteA} = \beta_0 + \theta_1 \log(expendA) + \beta_2[\log(expendB) - \log(expendA)] + \beta_3 prtystrA + u,$$

```
. gen leAleB= lexpendA- lexpendB

. regress  voteA lexpendA leAleB prtystrA

      Source |       SS       df       MS              Number of obs =     173
-------------+------------------------------           F(  3,   169) =  215.23
```

```
       Model |  38405.1089     3   12801.703           Prob > F      =  0.0000
    Residual |  10052.1397   169  59.4801165           R-squared     =  0.7926
-------------+------------------------------           Adj R-squared =  0.7889
       Total |  48457.2486   172  281.728189           Root MSE      =  7.7123

------------------------------------------------------------------------------
       voteA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lexpendA |   -.532101   .5330858    -1.00   0.320    -1.584466     .520264
      leAleB |   6.615417   .3788203    17.46   0.000     5.867588    7.363246
    prtystrA |   .1519574   .0620181     2.45   0.015     .0295274    .2743873
       _cons |   45.07893   3.926305    11.48   0.000     37.32801    52.82985
------------------------------------------------------------------------------
```

When we estimate this equation we obtain $\hat{\theta}_1 \approx -.532$ and $\text{se}(\hat{\theta}_1) \approx .533$. The t statistic for the hypothesis in part (ii) is $-.532/.533 \approx -1$. Therefore, we fail to reject $H_0 : \beta_2 = -\beta_1$.

Note: We can also use the following command after the original regression to get the estimate of $\hat{\theta}_1$.

```
. lincom lexpendA+ lexpendB

 ( 1)  lexpendA + lexpendB = 0.0

------------------------------------------------------------------------------
       voteA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.5321009   .5330858    -1.00   0.320    -1.584466     .520264
------------------------------------------------------------------------------
```

Note: Without estimating $\hat{\theta}_1$, we can also use the following command to test hypothesis:

```
. test lexpendA=- lexpendB

 ( 1)  lexpendA + lexpendB = 0.0

       F(  1,   169) =    1.00
            Prob > F =    0.3196
```

8. (**4.14**)

(i)
```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/HPRICE1

. regress lprice sqrft bdrms

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =   60.73
       Model |  4.71671468     2  2.35835734           Prob > F      =  0.0000
    Residual |  3.30088884    85  .038833986           R-squared     =  0.5883
-------------+------------------------------           Adj R-squared =  0.5786
       Total |  8.01760352    87  .092156362           Root MSE      =  .19706

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       sqrft |   .0003794   .0000432     8.78   0.000     .0002935    .0004654
       bdrms |   .0288844   .0296433     0.97   0.333    -.0300543    .0878232
       _cons |   4.766027   .0970445    49.11   0.000     4.573077    4.958978
------------------------------------------------------------------------------
```

The estimated model is

$$\begin{aligned} \log(\widehat{price}) &= \underset{(.10)}{4.76} + \underset{(.000043)}{.000379} sqrft + \underset{(.0296)}{.0289} bdrms \\ n &= 88, R^2 = .588. \end{aligned}$$

Therefore, $\hat{\theta}_1 = 150(.000379) + .0289 = .0858$, which means that an additional 150 square foot bedroom increases the predicted price by about 8.6%.

Alternatively, using the following command:

```
. lincom 150*sqrft+ bdrms

 ( 1)  150.0 sqrft + bdrms = 0.0
```

```
------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .0858013   .0267675     3.21   0.002     .0325804    .1390223
------------------------------------------------------------------------------
```

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

$$\begin{aligned}\log(price) &= \beta_0 + \beta_1 sqrft + (\theta_1 - 150\beta_1)bdrms + u \\ &= \beta_0 + \beta_1(sqrft - 150bdrms) + \theta_1 bdrms + u.\end{aligned}$$

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

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

```
. gen sqrftbdrms= sqrft-150* bdrms

. regress lprice sqrftbdrms bdrms

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =   60.73
       Model |  4.71671468     2  2.35835734           Prob > F      =  0.0000
    Residual |  3.30088884    85  .038833986           R-squared     =  0.5883
-------------+------------------------------           Adj R-squared =  0.5786
       Total |  8.01760352    87  .092156362           Root MSE      =  .19706

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  sqrftbdrms |   .0003794   .0000432     8.78   0.000     .0002935    .0004654
       bdrms |   .0858013   .0267675     3.21   0.002     .0325804    .1390223
       _cons |   4.766027   .0970445    49.11   0.000     4.573077    4.958978
------------------------------------------------------------------------------
```

and obtain the standard error on *bdrms*. We already know that $\hat{\theta}_1 = .0858$; now we also get $\text{se}(\hat{\theta}_1) = .0268$. The 95% confidence interval reported by my software package is .0326 to .1390 (or about 3.3% to 13.9%).