HOMEWORK 4

SOLUTIONS TO PROBLEMS

4.1 (i) and (iii) generally cause the *t* statistics not to have a *t* distribution under H_0. Homoskedasticity is one of the CLM assumptions. An important omitted variable violates Assumption MLR.3. The CLM assumptions contain no mention of the sample correlations among independent variables, except to rule out the case where the correlation is one.

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

(ii) The proportionate effect on $\widehat{salary}$ 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 from zero only because of sampling variation. On 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).

SOLUTIONS TO COMPUTER EXERCISES

C4.1 (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) The estimated equation (with standard errors in parentheses below estimates) is

$$\widehat{voteA} = \underset{(3.93)}{45.08} + \underset{(0.382)}{6.083}\log(expendA) - \underset{(0.379)}{6.615}\log(expendB) + \underset{(0.062)}{.152}\, prtystrA$$

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

The coefficient on log(*expendA*) is very significant (t statistic $\approx$ 15.92), as is the coefficient on log(*expendB*) (t statistic $\approx$ –17.45). The estimates imply that a 10% ceteris paribus increase in spending by candidate A increases 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

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

When we estimate this equation, we obtain $\hat{\theta}_1 \approx -.532$ and 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: $\beta_2 = -\beta_1$.

C4.6 (i) In the model

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

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

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

to obtain the 95% CI for θ_2. This turns out to be about .0020 ± 1.96(.0047), or about -.0072 to .0112. Because zero is in this CI, θ_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.

C4.8(i) There are 2,017 single people in the sample of 9,275.

(ii) The estimated equation is

$$\widehat{nettfa} = \underset{(4.08)}{-43.04} + \underset{(.060)}{.799}\, inc + \underset{(.092)}{.843}\, age$$

$$n = 2{,}017,\ R^2 = .119.$$

The coefficient on *inc* indicates that one dollar more in income (holding *age* fixed) is reflected in about 80 more cents in predicted *nettfa*; no surprise there. The coefficient on *age* means that, holding income fixed, if a person gets another year older, his/her *nettfa* is predicted to increase by about $843. (Remember, *nettfa* is in thousands of dollars.) Again, this is not surprising.

(iii) The intercept is not very interesting as it gives the predicted *nettfa* for *inc* = 0 and *age* = 0. Clearly, there is no one with even close to these values in the relevant population.

(iv) The *t* statistic is $(.843 - 1)/.092 \approx -1.71$. Against the one-sided alternative H_1: $\beta_2 < 1$, the *p*-value is about .044. Therefore, we can reject H_0: $\beta_2 = 1$ at the 1% significance level (against the one-sided alternative).

(v) The slope coefficient on *inc* in the simple regression is about .821, which is not very different from the .799 obtained in part (ii). As it turns out, the correlation between *inc* and *age* in the sample of single people is only about .039, which helps explain why the simple and multiple regression estimates are not very different; refer back to page 84 of the text.