Warm Up

1. Wooldridge (2015), exercise 3.7

Solution: Only (ii), omitting an important variable, can cause bias, and this is true *only* when the omitted variable is correlated with the included explanatory variables. The homoskedasticity assumption plays no role in showing that the OLS estimators are unbiased. (Homoskedasticity was used to obtain the usual variance formulas for the $\hat{\beta}_j$.) Further, the degree of collinearity between the explanatory variables in the sample, even if it is reflected in a correlation as high as .95, does not affect the Gauss-Markov assumptions. Only if there is a *perfect* linear relationship among two or more explanatory variables is MLR.3 violated.

Exercises

2. *Wooldridge*, 3.9

Solution: Part (i): $\beta_1 < 0$ because more pollution can be expected to lower housing values; note that β_1 is the elasticity of price with respect to nox. β_2 is probably positive because rooms roughly measures the size of a house. (However, it does not allow us to distinguish homes where each room is large from homes where each room is small.)

Part (ii): If we assume that rooms increases with quality of the home, then $\log(nox)$ and *rooms* are negatively correlated when poorer neighborhoods have more pollution, something that is often true. We can use Table 3.2 to determine the direction of the bias. If $\beta_2 > 0$ and $Corr(x_1, x_2) < 0$, the simple regression estimator $\tilde{\beta}_1$ has a downward bias. But because $\beta_1 < 0$, this means that the simple regression, on average, overstates the importance of pollution. [$E(\tilde{\beta_1})$ is more negative than β_1.]

Part (iii): This is what we expect from the typical sample based on our analysis in part (ii). The simple regression estimate, -1.043, is more negative (larger in magnitude) than the multiple regression estimate, -.718. As those estimates are only for one sample, we can never know which is closer to β_1. But if this is a "typical" sample, β_1 is closer to -.718.

3. *Wooldridge*, 7.4 (ignore any questions about statistical significance)

Solution: Part (i): The approximate difference is just the coefficient on utility times 100, or -28.3%.

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

Part (iii): The proportionate difference is $.181 - .158 = .023$, or about 2.3%.

Computer Exercises You should use `R` to complete these exercises. Any data set referred to in the question should be available in the `wooldridge` package in `R`. You do not need to turn in an R-script for these questions, but you are welcome to do so if you would like to.

4. *Wooldridge*, 3.C8 (Chapter 3, Computer Exercise 8)

Solution: Part (i): The average of *prpblck* is .113 with standard deviation .182; the average of *income* is 47,053.78 with standard deviation 13,179.29. It is evident that *prpblck* is a proportion and that *income* is measured in dollars.

Part (ii): The results from the OLS regression are

$$\widehat{psoda} = .956 + .115 prpblck - .0000016 income$$
$$N = 401, R^2 = .064$$

If, say, *prpblck* increases by .10 (ten percentage points), the price of soda is estimated to increase by .0115 dollars, or about 1.2 cents. While this does not seem large, there are communities with no black populations and others that are almost all black, in which case the difference in *psoda* is estimated to be almost 11.5 cents.

Part (iii): The simple regression estimate on *prpblck* is .065, so the simple regression estimate is actually lower. This is because *prpblck* and *income* are negatively correlated (-.43) and *income* has a positive coefficient in the multiple regression.

Part (iv): To get a constant elasticity, income should be in logarithmic form. I estimate the constant elasticity model:

$$\widehat{psoda} = -.794 + .122 prpblck + .077 \log(income)$$
$$N = 401, R^2 = .068$$

If *prpblck* increases by .20, $\log(psoda)$ is estimated to increase by .20(.122) = .0244, or about 2.44 percent.

Part (v): $\hat{\beta}_{prpblck}$ falls to about .073 when $prppov$ is added to the regression

5. *Wooldridge*, 7.C8 (Ignore anything about statistical significance since we haven't covered this yet)

Solution: Part (i): If the appropriate factors have been controlled for, $\beta_1 > 0$ signals discrimination against minorities: a white person has a greater chance of having a loan approved, other relevant factors fixed.

Part (ii): The simple regression results are

$$\widehat{approve} = .708 + .201 white$$
$$N = 1989, R^2 = .049$$

The coefficient on *white* means that, in the sample of 1,989 loan applications, an application submitted by a white application was 20.1% more likely to be approved than that of a nonwhite applicant.

Part (iii): When we add the other explanatory variables as controls, we obtain $\hat{\beta}_1 \approx .129$. The coefficient has fallen by some margin because we are now controlling for factors that should affect loan approval rates, and some of these clearly differ by race. (On average, white people have financial characteristics—such as higher incomes and stronger credit histories—that make them better loan risks.) But the race effect is still fairly strong.

Part (iv): When we add the interaction $white \times obrat$ to the regression, its coefficient is about .0081. Therefore, there is an interactive effect: a white pplicant is penalized less than a nonwhite applicant for having other obligations as a larger percent of income.

Part (v): The trick is to replace $obrat$ with $obrat - 32$ in the interaction (see Example 7.10 in the textbook). Replace $white \times obrat$ with $white \times (obrat - 32)$; the coefficient on $white$ is now the race differential when $obrat = 32$. We obtain a coefficient of about .113. There is evidence of discrimination (or, at least loan approval rates that differ by race for some other reason that is not captured by the control variables).

Cool Down

6. *Wooldridge*, 7.5.

Solution: Part (i): Following the hint,

$$\begin{aligned}\widehat{colGPA} &= \hat{\beta}_0 + \hat{\delta}_0 (1 - noPC) + \hat{\beta}_1 hsGPA + \hat{\beta}_2 ACT \\ &= \left(\hat{\beta}_0 + \hat{\delta}_0\right) - \hat{\delta}_0 noPC + \hat{\beta}_1 hsGPA + \hat{\beta}_2 ACT\end{aligned}$$

For the specific estimates in equation (7.6), 0 $\hat{\beta}_0 = 1.26$ and $\hat{\delta}_0 = .157$, so the new intercept is 1.26 + .157 = 1.417. The coefficient on *noPC* is -.157.

Part (ii): Nothing happens to the R^2. Using *noPC* in place of *PC* is simply a different way of including the same information on PC ownership.

Part (iii): It makes no sense to include both dummy variables in the regression: we cannot hold *noPC* fixed while changing *PC*. We have only two groups based on PC ownership so, in addition to the overall intercept, we need only to include one dummy variable. If we try to include both along with an intercept, we have perfect multicollinearity (the dummy variable trap).

References

Wooldridge, Jeffrey M. 2015. *Introductory Econometrics: A Modern Approach*. Cengage Learning, 6 ed.