Problem set 2

Jason Parker

In this problem set, the data sets are available as tables in the `wooldridge2.db` file.

Question 1

The following model can be used to study whether campaign expenditures affect election outcomes:

$$voteA = \beta_0 + \beta_1 \ln[expendA] + \beta_2 \ln[expendB] + \beta_3 prystrA + u,$$

where `voteA` is the percentage of the vote received by Candidate A, `expendA` and `expendB` are campaign expenditures by Candidates A and B, and `prtystrA` is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

1. What is the interpretation of β_1?
2. In terms of the parameters, state the null hypothesis that a 1% increase in A's expenditures is offset by a 1% increase in B's expenditures.
3. Estimate the given model using the data in the `vote1` table and report the results in usual form. Do A's expenditures affect the outcome? What about B's expenditures? Can you use these results to test the hypothesis in part 2?
4. Estimate a model that directly gives the t-statistic for testing the hypothesis in part
5. What do you conclude? (Use a two-sided alternative.)

Question 2

Use the data in the `lawsch85` table for this exercise.

1. Using the model:

$$\ln[salary] = \beta_0 + \beta_1 LSAT + \beta_2 GPA + beta_3 \ln[libvol] + \beta_4 \ln[cost] + \beta_5 rank + u,$$

   state and test the null hypothesis that the rank of law schools has no ceteris paribus effect on median starting salary.
2. Are features of the incoming class of students—namely, LSAT and GPA—individually or jointly significant for explaining salary? (Be sure to account for missing data on LSAT and GPA.)
3. Test whether the size of the entering class (`clsize`) and the size of the faculty (`faculty`) need to be added to this equation jointly.
4. What factors might influence the rank of the law school that are not included in the salary regression?

Question 3

Use the data in `hprice1` for this excercise. Now, use the log of the housing price as the dependent variable:

$$\ln[price] = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u.$$

1. You are interested in estimating and obtaining a confidence interval for the percentage change in price when a 150-square-foot bedroom is added to a house. In decimal form, this is $\theta_1 = 150\beta_1 + \beta_2$. Use the data in the `hprice1` table to estimate θ_1.

2. Write β_2 in terms of θ_1 and β_1 and plug this into the $\ln[price]$ equation.
3. Use part 2 to obtain a standard error for $\hat{\theta}_1$ and use this standard error to construct a 95% confidence interval.

Question 4

Use the data in the `wage2` table for this exercise.

1. Consider the standard wage equation

$$\ln[wage] = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u.$$

   State the null hypothesis that another year of general workforce experience has the same effect on $\ln[wage]$ as another year of tenure with the current employer.
2. Test the null hypothesis in part 1 against a two-sided alternative, at the 5% significance level.

Question 5

The table `401Ksubs` contains information on net financial wealth (`nettfa`) age of the survey respondent (`age`), annual family income (`inc`), family size (`fsize`), and participation in certain pension plans for people in the United States. The wealth and income variables are both recorded in thousands of dollars. For this question, use only the data for single-person households (so $fsize = 1$).

1. How many single-person households are there in the data set?
2. Use OLS to estimate the model

$$nettfa = \beta_0 + \beta_1 inc + \beta_2 age + u,$$

   and report the results using the usual format. be sure to use only the single-person households in the sample. Interpret the slope coefficients. Are there any surprises in the slope estimates?
3. Does the intercept from the regression in part 2 have an interesting meaning? Explain.
4. Find the p-value for the test $\mathcal{H}_0 : \beta_2 = 1$ against $\mathcal{H}_1 : \beta_2 < 1$. Do you reject $\mathcal{H}_0$ at the 1% significance level?
5. If you do a simple regression of `nettfa` on `inc`, is the estimated coefficient on `inc` much different from the estimate in part 2? Why or why not?

Question 6

Use the data in the `kielmc` table, only for the year 1981, to answer the following questions. The data are for houses that sold during 1981 in North Andover, Massachusetts; 1981 was the year construction began on a local garbage incinerator.

1. To study the effects of the incinerator location on housing price, consider the simple regression model

$$\ln[price] = \beta_0 + \beta_1 \ln[dist] + u,$$

   where price is housing price in dollars and dist is distance from the house to the incinerator measured in feet. Interpreting this equation causally, what sign do you expect for β_1 if the presence of the incinerator depresses housing prices? Estimate this equation and interpret the results.
2. To the simple regression model in part 1, add the variables $\ln[intst]$, $\ln[area]$, $\ln[land]$, $rooms$, $bath$, and age, where `intst` is distance from the home to the interstate, `area` is square footage of the house, `land` is the lot size in square feet, `rooms` is total number of rooms, baths is number of bathrooms, and `age` is age of the house in years. Now, what do you conclude about the effects of the incinerator? Explain why 1 and 2 give conflicting results.

3. Add $\ln[intst]^2$ to the model from part 2. Now what happens? What do you conclude about the importance of functional form?
4. Is the square of $\ln[dist]$ significant when you add it to the model from part 3?

Question 7

Use the data in the `wage1` table for this exercise.

1. Use OLS to estimate the equation

$$\ln[wage] = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u$$

   and report the results using the usual format.
2. Is $exper^2$ statistically significant at the 1% level?
3. Using the approximation

$$\frac{\partial wage}{\partial exper} \approx \hat{\beta}_2 + 2\hat{\beta}_3 exper,$$

   find the approximate return to the fifth year of experience. What is the approximate return to the twentieth year of experience?
4. At what value of exper does additional experience actually lower predicted $\ln[wage]$? How many people have more experience in this sample?

Question 8

Consider a model where the return to education depends upon the amount of work experience (and vice versa):

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

1. Show that the return to another year of education (in decimal form), holding experience fixed, is $\beta_1 + \beta_3 exper$.
2. State the null hypothesis that the return to education does not depend on the level of experience. What do you think is the appropriate alternative?
3. Use the data in the `wage2` table to test the null hypothesis in 2 against your stated alternative.
4. Let θ_1 denote the return to education (in decimal form), when $exper = 10$: $\theta_1 = \beta_1 + 10\beta_3$. Estimate θ_1 and a 95% confidence interval for θ_1. (Hint: Write $\beta_1 = \theta_1 - 10\beta_3$and plug this into the equation; then rearrange. This gives the regression for obtaining the confidence interval for θ_1.)

Question 9

Use the data in the `gpa2` table for this exercise.

1. Estimate the model

$$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + u.$$

   where `hsize` is the size of the graduating class (in hundreds), and write the results in the usual form. Is the quadratic term statistically significant?
2. Using the estimated equation from part 1, what is the "optimal" high school size? Justify your answer.
3. Is this analysis representative of the academic performance of all high school seniors? Explain.
4. Find the estimated optimal high school size, using $\ln(sat)$ as the dependent variable. Is it much different from what you obtained in part 2?

Question 10

Use the housing price data in the `hprice1` table for this exercise.

1. Estimate the model

$$\ln[price] = \beta_0 + \beta_1 \ln[lotsize] + \beta_2 \ln[sqrft] + \beta_3 bdrms + u$$

   and report the results in the usual OLS format.
2. Find the predicted value of $price$, when $lotsize = 20\,000$, $sqrft = 2\,500$, and $bdrms = 4$.
3. For explaining variation in price, decide whether you prefer the model from part 1 or the model

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$