Econ 220B, Winter 2020

Problem Set 6
Due Tuesday, March 10

This problem set asks you to perform certain calculations with the shortened version of the Griliches data set available from http://fhayashi.fc2web.com/datasets.htm and described on pp. 250-251 in Chapter 3 of Hayashi's text.

1.) Construct time dummies for the 9 possible values for the $YEAR$ variable of the form

$$yr66 \;=\; \begin{cases} 1 & \text{if } YEAR_t = 66 \\ 0 & \quad\text{otherwise} \end{cases}$$

$$\vdots$$

$$yr73 \;=\; \begin{cases} 1 & \text{if } YEAR_t = 73 \\ 0 & \quad\text{otherwise} \end{cases}.$$

What problem do you encounter if you try to estimate an OLS regression of LW on S,IQ, and $\mathbf{h}$ for

$$\mathbf{h} = (EXPR, TENURE, RNS, SMSA, yr66, ..., yr73)'?$$

Why does the problem get fixed if you drop the $yr72$ dummy?

2.) Drop the $yr72$ dummy from $\mathbf{h}$ and show that the OLS estimate of the coefficient on S is 0.06195. What is the standard error for this coefficient?

3.) Why might we worry that S could be enodgenous in this regression? Try to estimate the equation now by 2SLS using $\{MED, KWW, MRT, AGE\}$ as instruments for S. You should now get a standard error on S of 0.0144. What is the estimated value?

4.) Conduct an F test of the relevance of the instruments on the basis of the first-stage regression used in (3), and calculate Sargan's test of the specification in 3. How many degrees of freedom does it have, and what is the p-value?

5.) On the basis of your results in (4), do you have any concerns about either the relevance or the validity of the instruments? What could be the source of the problem?

6.) Suppose that as a result of these concerns, you decide that KWW, MRT, AGE might all have direct effects on the wage, independently of their implications for the level of schooling, so you're now interested in a regression of LW on $S, IQ, \mathbf{h}$,KWW, MRT, AGE using MED as the sole instrument for S. What do you calculate now for the coefficient on S and its standard error?

7.) Show that the Anderson-Rubin test of the hypothesis that the coefficient on S is zero in the above regression is 1.763. What are the degrees of freedom and p-value for this test? Use a grid search over possible values for the coefficient to invert the Anderson-Rubin acceptance region to obtain a 95% confidence interval for the coefficient.