Exercise instructions week 2:

Exercise 1:

We know that for unbiasedness the most critical assumption is E(Xe)=0 and that the OLS estimator in a bivariate model can be obtained from dividing the sample covariance in X and Y by the variance in X.

a. Write down the formula for the OLS bias if E(Xe) does not equal 0 for the univariate case.

b. How do you have to modify this formula if you are interested in the same beta but include further control variables in your regression (hint: regression anatomy theorem).

Exercise 2:

Read the Altonji, Elter and Taber paper. Some instructions for doing this: when reading, focus in the things that we are interested in here, which roughly start in section III on page 20.

Note we do not want to reproduce everything they have done. We are only interested in the alternative key assumption that they are making to replace the normal CIA. For example, do not get distracted by the fact that they are estimating a probit models as well (we are just discussing the OLS case), and that they have some asymptotics to prove that their alternative assumption makes sense. Of course, they have to argue that what they are doing is not at hoc.

A simplified OLS version of their regression of interest is Y=a+b_1CH+b_2X+e where W is a matrix that captures both sets of independent variables, the dummy CH and the X.

Note that throughout they continue assuming E(Xe)=0 for the set of X variables within W. What we are replacing/being worried about is E(CHe)=0.

a. In simple words, what do CH and X and Y refer to in their paper?

b. What is the assumption that they use as alterative to E(CH,e)=0. Without going into the details on how they derive this, how do they justify this assumption? Do you think this makes sense?

c. Can you modify your results from question 1b for allow for this alternative assumption? Hint (substitute for cov(e,CH~) where CH~ comes from the auxiliary regression, and note that cov(e,CH)= cov(e,CH~))

d. Can we estimate the various elements of the bias-corrected beta estimate under their alternative assumption? Write down the individual elements and the specifications to estimate these. How many regressions do we need to run?

Exercise 3: Stata implementation

This paper by Altonji, Elter and Taber initiated further research on coefficient stability. Emily Oster, an econometrician at Brown, recently augmented the AET-argument to also capture movements in the R-squared. This is readily usable in Stata in the “psacalc” package. Lets take a look:

a. Open Stata and install psacalc by typing `ssc install psacalc`

b. Load the automobile data that we already looked at. `sysuse auto`

c. We already examined the sensitivity of the effect of weight on price – remember the regression anatomy? Clearly things are not well specified. Now lets use the Oster-method (AET plus looking at changes in R-squared) to get bounds for the estimate of the effect of weight on price. For this, run:

   ```
   regress price foreign mpg weight headroom trunk
   ```

   ```
   psacalc beta weight
   ```

   -What is the estimated beta under the assumption that the violation of the CIA is proportional to selection on observables? Does this increase or decrease our confidence in our estimated effect of weight on price?

d. From the regression anatomy example in the lecture we saw that the homoscedasticity assumption seems unlikely to hold. In our first lecture we learnt how we can estimate heteroscedasticity-robust errors. In Stata, this is implemented for us by adding “`, robust`” to the multiple regression at the end. How do the coefficient on weight and the estimated standard errors change? When thinking about the hypothesis test with H0 weight has no effects, what is more problematic in this application: violating homoscedasticity or violating E(CHe)=0?