Economics 8379
Prof. Williams
Homework 4
due Friday, April 10

1. **weak IV.** Use the file `AKMHEdata.dta` for this exercise. This file contains 5 variables: log wage (`Lwage`), highest grade of education completed (`hgc`), place of birth (`pob`), year of birth (`yob`), and quarter of birth (`qob`). Watch the ivreg2 video I posted on Blackboard before doing this exercise.

   a. Recreate columns (1), (3), and (5) in Table 4.6.2 in Mostly Harmless Econometrics. Summarize Angrist and Pischke's discussion of whether the weak instruments problem is a concern here.

   b. Use the Montiel-Olea and Pflueger test to test for the presence of weak instruments for each of the 3 specifications. State explicitly what the null hypothesis is and what your conclusion is.

   c. For each of the three specifications, calculate the AR and CLR confidence intervals for the coefficient on `hgc`. Compare each to the standard confidence interval from the IV regression.

   d. For each of the three specifications, implement the Fuller estimator, the 2-step optimal GMM estimator, and the continuously-updated GMM estimator (CUE). (Just use the `Fuller(1)`, `gmm2s`, and `cue` options to `ivreg2`. Compare to the 2SLS and LIML estimators in Table 4.6.2.

2. **IV estimators Monte Carlo.** For this exercise, use the file IVMonteCarlo.m.

   a. Inspect the model being simulated in the .m file. For this model, which of the 4 estimators is a consistent estimator?

   b. Run the Monte Carlo simulation to determine the bias of the 4 estimators for $n = 100$, $n = 1000$ and $n = 10000$. Make sure that the number of simulations is large enough! Use the theory discussed in class (and in your answer to a) to explain your simulation results.

   c. Compare the standard deviation of the 4 estimators in the Monte Carlo simulations.

   d. Make at least one modification to the Monte Carlo exercise that leads to a substantial difference in the bias comparison in b. Explain your results carefully.

3. **Shift-share design.** This problem is about the Bartik instrument, which is a popular example of what a shift-share IV design. Consider the regression model,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

   where y_i is wage growth in location i and x_i is employment growth in location i. Employment growth, which is endogenous, can be written as

$$x_i = \sum_k s_{ik} g_{ik}$$

where g_{ik} is employment growth in location i in industry k and s_{ik} is the employment share in location i in industry k. This is an accounting identity, not an assumption. Next, suppose we write $g_{ik} = g_k + \tilde{g}_{ik}$. The Bartik instrument is $z_i = \sum_k s_{ik} g_k$. To simplify, suppose there are only two industries.

a. Explain when the two identifying assumptions for using z_i as an instrument in the regression model are valid. Write out a formula for the 2SLS estimator. Is the model just-identified or over-identified?

b. Suppose we instead use s_{i1} and s_{i2} as two separate instruments. Write down the 2SLS estimator and the 2 IV conditions under which this estimator is consistent. Is this estimator equivalent to the estimator in a?

c. Now suppose β_1 is heterogeneous across locations. That is, $y_i = \beta_0 + \beta_{1i} x_i + \varepsilon_i$. Are there conditions under which the estimators in a and b are consistent for the same estimand? Explain.