ECON 626: Empirical Microeconomics

Instrumental Variables: replicating a non-RCT use of IV

Department of Economics
University of Maryland
Fall 2019

The data set `dinkelman_aer2011_matched_censusdata.dta` contains observations from Taryn Dinkelman's 2011 AER paper, "The Effects of Rural Electrification on Employment: New Evidence from South Africa." The .do file `dinkelman_aer2011_start.do` has some material to get you started working with this dataset.

1. First, regress the outcome, the proportion of women employed (`d_prop_emp_f`) , on the endogenous regressor, electrification (`T`), controlling for a list of exogenous variables, with the correct standard error option. This is the naive OLS specification. Each of these four elements has been placed in a "global" macro for your convenience, so you should be able to type

   ```
   regress $macro1 $macro2 $macro3, $macro4
   ```

   and obtain the result, with the correct use of these existing macros (they are actually called `outcome`, `controls`, etc.) Your result should match Dinkelman (2011), Table 4, Column 4. The key coefficient is on the variable `T`, which is an indicator for the electrification program.

2. Estimate the first stage: regress the endogenous treatment variable on the slope of land instrument (`mean_grad_new`), controlling for the same list of exogenous variables. Again, this can be done via the global macros that have already been set up, or can be done manually. Your result should match Dinkelman (2011), Table 3, Column 4. After estimation, do three additional things:

   - Store the coefficient on the instrument as a local macro;
   - Calculate the F statistic for the excluded instrument (one way is to use the `test` command; it should match the value at the bottom of Table 3, Column 4); and
   - use the `predict` command to generate a new variable that contains the fitted values of $\hat{X}^{end}$, that is, the values of `T` predicted by the instrument and the exogenous regressors.

3. Calculate the reduced form and Wald estimate, and estimate the second stage. The reduced form is the regression of the outcome on the instrument with controls. The Wald estimate is the ratio of this value to the coefficient you stored from the first stage. The second stage is the regression of the outcome on the fitted values of $\hat{X}^{end}$ that is, the values of `T` predicted by the instrument and the exogenous regressors that you generated in the previous part. (Note, of course, that these standard errors are not correct.) The Wald estimate and the second stage estimate should have the same value.

4. Use the `ivregress 2sls` command to estimate 2SLS with better standard errors. Your coefficients should agree with those in Dinkelman (2011), Table 4, Column 8, and the coefficient of interest should be identical to that in the previous part. At what level does it now appear to be significant?

5. To correct the confidence interval (given the borderline weakness of the instrument), use the `condivreg` command, the `weakiv` command, or both, to compute the Anderson-Rubin 95-percent confidence interval. Suggested syntax for each command is provided in the comments. As with Dinkelman's reported results, this interval should now not include zero, but it may not be exactly the interval Dinkelman gave in her paper, as she did not have access to these Stata routines at the time she started her work.