**Mixtape SSIV Workshop: Coding Lab**

This lab will walk through some basic SSIV analyses using data from Autor, Dorn, and Hanson (ADH, 2013). As discussed in lecture, ADH use a shift-share instrument aggregating Chinese import shocks across 397 manufacturing instruments with exposure weights calculated as the economy-wide share of (baseline) industry employment. We will use three cleaned datasets from their setup:

* *adh\_shocks.dta*: an industry-by-year dataset of the shocks
* *adh\_shares.dta*: a location-by-industry-by-year dataset of the shares
* *adh\_noIV.dta*: a location-by-year dataset of the main outcome (manufacturing employment growth, *y*), treatment (local growth of China import exposure, *x*), and other useful variables – excluding the ADH instrument

Exercises:

1. Construct the ADH (location-by-year) instrument by appropriately combining the data on shocks and shares. Merge this into the *adh\_noIV* dataset, and estimate an IV regression of the outcome onto the treatment which controls for year (i.e. the *post* variable) and weights by baseline total employment (the *weight* variable), clustering by *state*. Then estimate the exact same IV regression replacing the outcome *y* with the lagged outcome *y\_lag*, capturing growth in manufacturing employment that took place before the ADH “China Shock” quasi-experiment. How does the latter IV regression help build support for the former IV regression?

|  |  |
| --- | --- |
| *Main IV Estimate:* | *Standard Error:* |
| *Lag Outcome IV Estimate:* | *Standard Error:* |

*Comments:*

1. Construct the “sum-of-shares” control from the *adh\_shares* dataset and add this control to both of the previous IV regressions. How does the main IV estimate change? Why, intuitively, is this control important to include?

|  |  |
| --- | --- |
| *Main IV Estimate:* | *Standard Error:* |
| *Lag Outcome IV Estimate:* | *Standard Error:* |

*Comments:*

1. Interact the “sum-of-shares” control with year and add this control to both of the previous IV regressions. How do both IV estimates change? Can you see why, intuitively, the interaction control shifts the main IV estimate so much?

|  |  |
| --- | --- |
| *Main IV Estimate:* | *Standard Error:* |
| *Lag Outcome IV Estimate:* | *Standard Error:* |

*Comments:*

1. Use the *ssaggregate* command to run both of the previous IV regressions at the shock level. You should control for year fixed effects in the shock-level IV regressions and cluster by industry. The coefficients should be identical to the previous estimates, but the standard errors will be different.

|  |  |
| --- | --- |
| *Main IV Estimate:* | *Standard Error:* |
| *Lag Outcome IV Estimate:* | *Standard Error:* |