Appendix to "How to Use Spatial Instruments"[1]

To confirm and illustrate the points made in the article, I implemented a spatial instruments strategy on a large number of randomly generated datasets. 500 cross-sectional datasets were generated for each of 46 scenarios, ranging from $k = 5$ to $k = 50$ regions, always with five units per region. The data generating process mirrored that described by equations (1) and (2) in the article. I then estimated β_2 using OLS, a leave-one-out spatial instrument γ_i, and a modified spatial instrument Γ_i. I also completed the same steps for panel data, which included fifty units per time period. Replication code is available on the Harvard Dataverse.

The OLS estimates were biased, with the median estimate in the cross-sectional case missing the mark by 0.5, regardless of the number of groups. The spatial instrument γ_i performed much better, with a median error of approximately zero. However, Figure 1 reveals the limits of this strategy by plotting the interval from the first to the ninety-ninth percentile of estimates. For $k < 15$ regions, using γ_i as an instrument returned hugely varying estimates of β_2. The modified spatial instrument Γ_i exhibited less variance, but was biased by about 0.16, which is more easily seen in the middle row of results. For the simulated panel data, the bias and variance of $\widehat{\beta_2}$ were much lower when estimated using Γ_{it} (note the changing scales across rows of Figure 1). This was a result of the much larger number of units in each group: fifty per year rather than five per region.

These results confirm the main points of the article: spatial instruments can overcome unobserved confounding to identify a parameter of interest. Doing so requires that such confounding be limited to within groups, and doing so efficiently requires an adequate number of groups and units per group.

[1] James Sundquist, Yale University (james.sundquist@yale.edu)

Figure 1: Simulation Study Results