Collinear sectors in Shift-Share Designs

Michal Kolesár

September 4, 2019

If columns of W are collinear, so it that it has rank $S_0 < S$, then we cannot recover $\mathcal{X}$ from X, or $\tilde{\mathcal{X}}$ from $\ddot{X}$. There doesn't appear to be a way to fix the way we estimate the asymptotic variance, but there are three partial solutions:

1. Drop the collinear sectors, and adjust X_i accordingly, so that (assuming we drop the last $S - S_0$ sectors) $X_i = \sum_{s=1}^{S_0} w_{is}\mathcal{X}_s$. This effectively puts shocks to the collinear sectors into the residual (which is analogous to letting say the shock to non-manufacturing sectors be part of the residual), so that the estimand is now different, $\sum_i \sum_{s=1}^{S_0} \pi_{is}\beta_{is} / \sum_i \sum_{s=1}^{S_0} \pi_{is}$. With only a few collinear sectors, this seems like the cleanest solution.

2. Aggregate the sectors. For example, if originally w_{ist} were weights for 3-digit sectors s with the 4th digit equal to t, we can aggregate up to a 3-digit level and instead use the weights $\sum_t w_{ist}$ and shocks $\overline{\mathcal{X}}_s$, which is a (weighted) average of the 4-digit shocks $\mathcal{X}_{st}$. Note that this changes the model, since we're now saying that the shocks only affect the outcome through the 3-digit sector aggregate $\overline{\mathcal{X}}_s$. One could also only aggregate the collinear sets of sectors.

3. If the only controls are those with shift-share structure, and we have data on $\mathcal{Z}_s$, we can estimate $\tilde{\mathcal{X}}$ by running a sector-level regression of $\mathcal{X}$ onto $\mathcal{Z}$, and taking the residual.

Note that solutions 1 and 2 involve changing the shock vector X_i, so that the aggregation or dropping the collinear sector must be done before using the `reg_ss` and `ivreg_ss` commands.