Peter Hull March 27, 2017

Examiner Designs and First-Stage F Statistics: A Caution

High-dimensional instrumental variable (IV) regressions can be cumbersome to implement. To ease their computational burden, applied researchers often reduce the dimensionality of a many-IV first stage in a manual first step. For example, in the quasi-experimental "examiner" design, a researcher observes draws of an outcome Y, an endogenous variable X, and a vector of K mutually-exclusive and exhaustive binary variables, Z, which indicate as-good-as-random assignment to examiner groups.[1] Rather than directly estimating a two-stage least squares (2SLS) regression of Y on X with the K instruments, a researcher may compute the equivalent IV coefficient by first constructing the examiner-level average of the endogenous variable, $\hat{X}$, and then instrumenting X by $\hat{X}$. Often researchers use leave-one-out averages to form $\hat{X}$, in which case the two-step constructed IV coefficient matches that of the Angrist, Imbens, and Krueger (1999) jackknife IV estimator.[2]

Computing group-level averages is typically much simpler than inverting a high-dimensional instrument design matrix in 2SLS. Since the two approaches produce numerically identical coefficients, it seems natural to prefer the use of "constructed instruments" $\hat{X}$ in these cases. Nevertheless, a researcher should never forget in doing so that the dimensionality of her true identifying variation is K, not one. If she does forget, she may, for example, fall into the common trap of using the F statistic from a regression of X on $\hat{X}$ to gauge the first-stage strength of her identification. This is

$$\hat{F}_1 = \frac{(N-2)\hat{R}^2}{1-\hat{R}^2},$$

where $\hat{R}^2$ denotes the sample R-squared from this regression and N is the sample size. The "true" first-stage F statistic from the regression of X on Z is, by contrast,

$$\begin{aligned}\hat{F}_K &= \frac{(N-K-1)\hat{R}^2}{K(1-\hat{R}^2)} \\ &= \frac{N-K-1}{K(N-2)}\hat{F}_1,\end{aligned}$$

which is approximately K times smaller than $\hat{F}_1$. In practice, therefore, researchers run the risk of greatly overstating their first-stage F-statistics when using constructed instruments in examiner designs – estimators suffering from severe many-weak IV bias may go undetected. Researchers should always be sure to apply the above degree-of-freedom correction when using $\hat{F}_1$ to diagnose the strength of examiner instruments. Alternatively, they may prefer more formal dimension-reduction techniques for IV, such as the LASSO approach in Belloni et al. (2012), which may prove useful in many-weak designs.[3]

[1] See Kling (2006), Maestas, Mullen, and Strand (2013), and Doyle et al. (2015) for three recent examples of this approach.

[2] The result demonstrated here is easily extended to cases with multiple endogenous regressors and auxiliary controls.

[3] The Angrist, Imbens, and Krueger (1999) estimator also has favorable properties in many-weak environments, but also many drawbacks; see, e.g., Davidson and MacKinnon (2006).

References

Angrist, J., G. Imbens, and A. Krueger, "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14 (1999), 57-67.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen, "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain," *Econometrica*, 80 (2012), 2369-2429.

Kling, J., "Incarceration Length, Employment and Earnings," *American Economic Review*, 96 (2006), 863-876.

Davidson, R. and J. MacKinnon, "The Case Against JIVE," *Journal of Applied Econometrics*, 21 (2006), 827-833.

Doyle, J., J. Graves, J. Gruber, and S. Kleiner, "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns," *Journal of Political Economy*, 123 (2015): 170-214.

Maestas, N., K. Mullen, and A. Strand, "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt," *American Economic Review*, 103 (2013), 1797-1829.