Problem Set 5: Instrumental Variables

MGMT 737

1. PART I (Complier share): We explore Angrist and Evan (1998), who study the impact of children on parents' supply. This paper has two instruments: 1) whether the first two children of a mother are the same sex *Same sex*; 2) whether the second (successful) pregnancy is twins *Twins-2*. The 1980 Data for this paper is available in the repository in a csv `ang_ev_1980¸csv`.

   I was not able to perfectly match AE's sample but it is very close. As a result, the point estimates are not identical, but are close enough that you should be able to feel confident. You may use pre-built estimation packages unless otherwise stated.

   More than 2 children is `morekids`, *Number of children* is `kidcount`, *worked for pay* is `mom_worked`, *Weeks worked* is `mom_weeks_worked` and the two dummy instruments are `samesex` and `twins_2`.

   (a) Replicate the coefficients from Table 5, Column 1, rows 1-4, using a linear regression.

   (b) Replicate the coefficients from Table 5, Column 2, rows 3-4, using 2SLS. Convince yourself you could construct this estimate by hand using the result in the previous answer.

   (c) Replicate the coefficients from Table 5, Column 7, rows 1-4, using a linear regression.

   (d) For the endogeneous vairable "More than 2 children", what is the complier share for each of the two instruments?

   (e) For the endogeneous vairable "More than 2 children" and each of the two instruments, what is the average share of the complier population with an education greater than high school (`moreths`)? What about different mother race shares?

   (f) Using the Same sex instrument, construct the Weak IV robust Anderson-Rubin confidence intervals using the algorithm outlined in Chernozhukov and Hansen (2007) (see the slides)

2. PART II: **Kitagawa test** Now we implement the Kitagawa (2015) test for instrument validity. We will focus on the *worked for pay* outcome as our endogeneous variable Y, more than two children as our endogeneous variable, and our two instruments will be as before (same sex and twins). Implement the procedure below for both instruments.

   (a) Let $P(y,d) = n_1^{-1}\sum_{i:Z_i=1} 1(Y_i = y, D_i = d)$, where n_1 is the number of observations where $Z_i = 1$ in the sample, and $Q(y,d) = n_0^{-1}\sum_{i:Z_i=0} 1(Y_i = y, D_i = d)$, where n_1 is the number of observations where $Z_i = 0$ in the sample. $n = n_0 + n_1$ Estimate $P(y,d)$ and $Q(y,d)$ and report them for the 4 possible pairs of (y,d).

   (b) Next, we want to calculate the following statistic (Note that this is simplified relative to the general Kitagawa case because Y is binary):

   $$T = \left(\frac{n_0 n_1}{n}\right)^{1/2} \max\Big\{ \sup_{y\in\{0,1\}} \left\{ \frac{Q(y,1) - P(y,1)}{\xi \vee \sigma_{P,Q}(y,1)} \right\}, \sup_{y\in\{0,1\}} \left\{ \frac{Q(y,0) - P(y,0)}{\xi \vee \sigma_{P,Q}(y,0)} \right\}\Big\}$$

   where ξ is a trimming constant that you pick, $\vee$ denotes the "max" between the two objects, and

   $$\sigma^2_{P,Q}(y,d) = (1-\hat{\lambda})P(y,d)(1-P(y,d)) + \hat{\lambda}Q(y,d)(1-Q(y,d)), \qquad \hat{\lambda} = n_1/n.$$

   Estimate σ for each of the four possible pairs of (y,d), and then calculate the T based using a $\xi = 0.1$. Report this T statistic.

   (c) Now, we need to construct critical values for the statistic T. First, take the subsample of n_1 observations where $Z_i = 1$. Sample a new set of n_1 observations from this subsample, with replacement using weights $H(y,d) = \hat{\lambda}P + (1-\hat{\lambda})Q$. (E.g. the weight of drawing each observation is a function of this combined probability). Use this new data to construct a new empirical distribution P^*. Similarly construct a new sample of n_0 observations based on the subsample of observations where $Z_0 = 0$ using $H(y, d)$. Use this to construct Q^*. Then, calculate the T statistic above, but using P^* and Q^* in place of P and Q.

(d) Repeat this last process 500 times and construct a distribution of the T^*. We reject the null of instrument validity (exclusion and monotonicity) at the 5 percent level if T is greater than the 95th percentile of the distribution of T^*. Identify what quantile T sits in the distribution of T^*.