Homework Set #5

John Vandivier, 00859508

1

.60456 is the fraction of women who work in the data set. It is the mean of the worked variable.

23.16 is the mean number of weeks worked by women in the data set.

The median labor income is 1005.

2

.85536 is the fraction of women who have had a second child.

The coefficient on the variable called second in the simple regression is -6.813862.

The first stage estimate of the effect of twin1st on second is .2746051.

The reason the first stage estimate is not 1 is because many people have additional children without having twins. So the observation of twins does not fully explain even a single additional child, since the woman may have had additional children in the absence of twins.

The Wald estimate of the coefficient of second is -3.605, which is obviously smaller than the previous OLS estimate.

The R squared from the first stage is .1519.

For the regression on worked, the ratio requested is (.012376)2 / (.0317631)2 = .151815.

3

Each of the 6 regressions except 1 presented a statistically significant coefficient. The economic importance of these variables is questionable. Altogether these factors present an R squared of less than .02. The effects are not large when compared to other factors such as the kids variable. One possible exception is the black ethnicity flag which has a coefficient of about .1. The judgement is arbitrary, but it could be considered important when the presence of a factor is associated with a 10% increase in the probability of the output. These findings suggest the frequency of twins on first birth is not random.

4

The observation of a second child is associated with a statistically significant and economically strong negative effect on both worked and weeks. The coefficients are -.1797937 and -9.801238, respectively. After instrumenting, the estimates are reduced but the direction and statistical significance are preserved.

5

The correlation coefficient is .3898. Yes, this is concerning.

6

The first stage regression of kids on agefsta indicates a strong correlation between the age at which a mother first had twins and their overall fertility. With an F > 500, the p(F) ~= 0 and we may reject the null hypothesis that kids is invariant to agefsta. Finite sample bias occurs due to the presence of significant error-induced variation in the first-stage1. With a finite sample, random assignment can produce unbiased and consistent estimates of the effect. However, we have established that assignment using twin1st is non-random. Therefore, finite sample bias is a concern.

Citation:

1 - <http://www.mdrc.org/sites/default/files/Finite%20Sample%20Bias%20from%20Instumental%20Variables%20full%20report.pdf>

DO file content:

sum worked

sum weeks

sum linc, detail

cumul lincome, gen(cum) //graph cumulative bc i don't believe the median I'm seeing

line cum lincome

gen second = 0 //question 2 begin

replace second = 1 if kids > 1

reg weeks second

ivregress 2sls weeks (second = twin1st), first //iv son watup

ivregress 2sls worked (second = twin1st), first

gen racewhite = 0 //question 3 begin

replace racewhite = 1 if race == 1

gen raceblack = 0

replace raceblack = 1 if race == 2

gen raceother = 0

replace raceother = 1 if race == 3

reg educ twin1st

reg agefst twin1st

reg agem twin1st

reg racewhite twin1st

reg raceblack twin1st

reg raceother twin1st

reg worked agem agefst educ raceblack raceother second //question 4 begin

reg weeks agem agefst educ raceblack raceother second

ivregress 2sls worked agem agefst educ raceblack raceother (second = twin1st), first

ivregress 2sls weeks agem agefst educ raceblack raceother (second = twin1st), first

correlate second twin1st //question 5

gen agebefore = 0 //question 6

replace agebefore = 1 if agefst < 20

gen agemid = 0

replace agemid = 1 if agefst >= 20

replace agemid = 0 if agefst > 24

gen ageafter = 0

replace ageafter = 1 if agefst > 24

gen agefsta = 0

replace agefsta = 1 if twin1st == 1 & agebefore == 1

replace agefsta = 2 if twin1st == 1 & agemid == 1

replace agefsta = 3 if twin1st == 1 & ageafter == 1

reg kids agefsta