Problem set 3: Empirical Methods for Applied Microeconomics

Due 11/26.

General instructions. Please work in a group no larger than 3. When you write up your results, please let me know who is in your group. (Only turn in 1 completed homework.). Present your answers in a concise way (typed is highly preferred). Please include relevant Stata output and well-commented do files and ado files for all the exercises (or equivalent in the package of your choice.) Please do NOT include lots of undigested log files.

Put the do files in an appendix and make clear reference to the regression output and/or figures.

Problem 1

IV simulation.

Start with 1000 observations of a standard normal u.

Create e, $v1$, and $v2$. Let e and $v1$ both be $u + 0.5$ times another standard normal.

Let $v2$ be u plus a standard normal.

Create 41 instruments (all standard normals), $z0, z1, ..., z40$.

Create $x1 = 0.2 \cdot z0 + v1$ and $x2 = 0.2 \cdot z0 + v2$.

Create $y1 = x1 + e$ and $y2 = x2 + e$.

(i) Which x is more highly correlated with the error term, $x1$ or $x2$?

(ii) Run OLS with $y1$ on $x1$, and then separately with $y2$ on $x2$.

(iii) Run 2SLS with the 1 good instrument for both $y1$ and $y2$ (separately). Do you get substantively different answers than in (ii)?

(iv) Run 2SLS with the 1 good and 40 bad instruments for both $y1$ and $y2$ (separately). What are the first stage Fs? Which estimate is further from the true value? Is this what you expected?

(v) Run LIML for both. Does LIML help?

(vii) Pretend you did not test the first stage in both models. Use either the Anderson-Rubin, Kleibergen

K or Moreira (et al.) conditional approaches for obtaining CIs.

For Anderson-Rubin or Moreira, see

http://www.stata.com/statalist/archive/2013-10/msg00136.html

and for the Moreira et al. approach see

http://www.fgv.br/professor/mjmoreira/

Do these do better?

Problem 2

Quantile regression.

Here we will use some data from a paper by Angrist, Kevin Lang, and Phil Oreopoulous on another experiment. Here is the link to the paper:

http://homes.chass.utoronto.ca/ oreo/research/compositions/Incentives_and_Services_For_College_Achievement.p

The data are at

http://econ-www.mit.edu/faculty/angrist/data1/data,

click on Angrist, Lang, and Oreopoulos (2009).

Load in the data. We will use the following sample.

```
(control==1 | sfsp==1) & noshow==0 & (GPA_year1!=. &
grade_20059_fall!=.)
```

We will in some specifications use some of the following controls.

```
List of controls:
female english hsgroup1 hsgroup2  numcourses6 numcourses5 lastminusof
lastminocc momedlehs momedsomcol dadedlehs dadedsomcol


Definitions:
hsgroup1 = hsgroup==1
hsgroup2 = hsgroup==2
hsgroup3 = hsgroup==3
gen numcourses6 = numcourses_==6
gen numcourses5 = numcourses_==5
gen numcoursesle4 = numcourses_<=4
gen lastminusof = lastmin==1 | lastmin==2
```

```
gen lastminocc = lastmin==3
gen lastminrar = lastmin==4 | lastmin==5
gen momedlehs = mom_edn<=2
gen momedsomcol = mom_edn>=3 & mom_edn<=5
gen momedgecol = mom_edn>=6 & mom_edn<=9
gen dadedlehs = dad_edn<=2
gen dadedsomcol = dad_edn>=3 & dad_edn<=5
gen dadedgecol = dad_edn>=6 & dad_edn<=9
```

(i) Start by regressing GPA in year 1 on the SFSP dummy and the controls above in our preferred sample (control group or SFSP group, no noshows).
(ii) Now calculate quantile regression estimates of the effects of the program on GPA in year 1, at the first decile, the median, the 9th decile, using the same sample and controlling for the Xs.
(iii) Run the same quantile regression without any other Xs but the treatment dummy in the specification. Why did the coefficients change from (ii)?
(iv) Now we will calculate the same results by hand. Figure out the 10th, 50th, and 90th percentile of the GPA year 1 distribution for the SFSP group and for the control group. Use these to get estimates of the QTE at these percentiles. Do you get the same estimates for the QTE as in (iii)?
(v) Now we will bootstrap the calculation with replacement. So, for 999 replicates, you will draw the data with replacement, calculate the relevant percentiles of the treatment and control groups, and then for the each percentile, get the QTE within the bootstrap replicates, sort it, and get the 90% CI for each percentile (pointwise) as the 50th smallest bootstrap estimate and 950th largest.
(vi) Now we will do the Abadie, Angrist, and Imbens IV QTE. Download ivqte from Blaise Melly's webpage http://www.econ.brown.edu/fac/Blaise_Melly/code_ivqte.html. You may also need to install "moremata" and "kdens".

First estimate 2SLS (instrument for sfsp_p with sfsp), with all the controls in.

Then estimate the AAI IV QTE using the ivqte command. (See right below, for command with ";" as the delimiter and for the 0.1 decile.

```
ivqte GPA_year1 (sfsp_p=sfsp) if (control==1 | sfsp==1) &
noshow==0 ,  q(.1) variance dummy(female hsgroup1 hsgroup2
numcourses6 numcourses5  lastminusof lastminocc momedlehs momedsomcol
dadedlehs dadedsomcol)  aai;
```

Do you think effects of the program are constant across the distribution? (1/2 page max only)

(vii) Now, do unconditional IVQTE (you can also estimate this using the same package ivqte as above but a different set of options). Does this change your view of whether effects are constant? (1/2 page max only)