Econ 312: Problem Set 3

Professor Magne Mogstad

Due Tuesday, April 23 in the lecture

Problem 1

Consider the simple regression model:

$$Y = Y_0 + D(Y_1 - Y_0) = \beta_0 + \beta_1 D + U$$

where D is a binary treatment variable.

a) In general, the average treatment effect for the treated (ATT) and the LATE will not be the same. This is because ATT is a weighted average of two effects: one on always- takers and one on compliers. Show that this is the case.

b) When is LATE = ATE?

Problem 2

Consider the adjusted set of assumptions conditional on covariates:

- Exogeneity: $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z | X$
- Relevance: $\mathbb{P}[D = 1|X, Z = 1] \neq \mathbb{P}[D = 1|X, Z = 0]$ a.s.

- Monotonicity: $\mathbb{P}[D_1 \geq D_0|X] = 1$ a.s.

- Overlap: $\mathbb{P}[Z = 1|X] \in (0, 1)$ a.s.

a) Show that we can identify:

$$LATE(x) \equiv \mathbb{E}[Y_1 - Y_0|\underbrace{T = c}_{\text{compliers}}, X = x]$$

b) Show:

$$\mathbb{E}[Y_1 - Y_0|T = c] = \mathbb{E}\left[\frac{LATE(X)\mathbb{P}[T = c|X]}{\mathbb{P}[T = c]}\right]$$

c) Suppose that X is descrete (a set of binary indicators); Excluded instruments are Z and XZ where $Z \in \{0, 1\}$; the model is fully saturated. Show that:

$$\beta_{TSLS} = \mathbb{E}[LATE(X)\frac{\text{Var}(p(X, Z)|X)}{\mathbb{E}[\text{Var}(p(X, Z)|X)]}]$$

d) Show that for binary D and Z and any function $G = g(Y, X, D)$:

$$\mathbb{E}[G|T = c] = \frac{1}{\mathbb{P}[T = c]}\mathbb{E}[\kappa G]$$

where $\kappa \equiv 1 - \frac{D(1-Z)}{\mathbb{P}[Z=0|X]} - \frac{Z(1-D)}{\mathbb{P}[Z=1|X]}$

Problem 3

1. Seats in Dutch medical schools are assigned through a lottery. Applicants to medical studies in the Netherlands are assigned to lottery categories based on their high school grades. The categories diffr by the probability to be awarded a place (to win the lottery). If people loose a lottery they can try again the following year.

Below you find a link to a dataset that has results from peoples' first lottery outcome for participants in 1988 and 1989, and whether they attended medical school, as well as earnings from a survey that was sent out in 2007.

`https://www.dropbox.com/s/dy6cmnmmvvkfv4l/lottery.dta?dl=1`

You plan to estimate the return to attending medical school (D) on earnings in 2007 ($\ln w$) using the lottery outcome (Z) as your instrument.

a) Discuss instrument exogeneity, exclusion and monotonicity.

b) Assess instrument relevance.

c) Estimate the return to attending medical school on earnings in 2007 using IV, and interpret the results.

d) Count the number of compliers, and compare them to the population of appli- cants in terms of gender.

e) Is the IV estimate an estimate of the ATT? Explain why or why not.

f) Estimate the mean and distribution of Y_0 and Y_1 for compliers.

g) What can you say about Y_0 and Y_1 for always- and never-takers?

h) The lottery is within lottery category and year, so your instrument is only exogenous within these groups. Estimate lottery category×year specific LATEs and combine these in one estimate. Compare this to the specification where you control for lottery category×year dummies and also interact the instrument with these dummies.

Problem 4

We are interested in how health insurance affects out-of-pocket expenditure on drugs, and have access to an extract from the Medical Expenditure Panel Survey of individuals over the age of 65 years. We want to estimate the following equation:

$$ldrugexp = \alpha + \gamma hi\_empunion + X\beta + U$$

where *ldrugexp* is log expenditure on prescribed medical drugs, *hi_empunion* is equal to one if the individual has supplemental health insurance and zero otherwise, and we control in X for *age*, *gender*, *linc* (log of household income), *totchr* (the no. of children), and *blhisp* (a dummy for being black or hispanic). Below is output from Stata with a number of results that may be useful in this exercise.

a) Explain why we may worry that having supplemental health insurance is endogenous in the equation above.

b) A suggested instrument is *multlc*, a dummy for whether the firm at which the individual is employed is a large operator with multiple locations. Why or why not may this be a good instrument (think about the conditions that need to hold to identify the LATE)? Using the output below, do you think that *multlc* is a weak instrument?

c) Derive the indirect least squares representation of the IV-estimator using *multlc* as an instrument, and calculate it using the output below. Interpret the estimate.

d) Assuming $\beta = 0$, derive the IV-estimator using the moments (covariances), and calculate it using the output below.

e) What is the share of females in the complier group? How does this compare to the overall population? How does this affect your interpretation of the estimates?

f) Assuming $\beta = 0$, what is the share of females in the three groups of compliers, always-takers and never-takers?

g) Using the means and counts of *ldrugexp* from the output below, estimate $\mathbb{E}[Y_0|\text{never taker}]$, $\mathbb{E}[Y_1|\text{always taker}]$, $\mathbb{E}[Y_0|\text{complier}]$, and $\mathbb{E}[Y_1|\text{complier}]$, where Y's are the potential outcomes for *ldrugexp* with and without supplemental health insurance. How do the compliers compare to the other groups, and what do you conclude about external validity?

```
. su ldrugexp hi_empunion multlc totchr age female blhisp linc

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    ldrugexp |     10089    6.481361    1.362052          0   10.18017
 hi_empunion |     10089    .3821984    .4859488          0          1
      multlc |     10089    .0620478    .2412543          0          1
      totchr |     10089    1.860938    1.292858          0          9
         age |     10089    75.05174    6.682109         65         91
-------------+--------------------------------------------------------
      female |     10089    .5770641    .4940499          0          1
      blhisp |     10089    .1635445      .36988          0          1
        linc |     10089    2.743275    .9131433  -6.907755   5.744476

.
. reg ldrugexp multlc totchr age female blhisp linc , robust

Linear regression                                      Number of obs =   10089
                                                       F(  6, 10082) =  376.72
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1775
                                                       Root MSE      =  1.2356

------------------------------------------------------------------------------
             |               Robust
    ldrugexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      multlc |  -.2002194   .0540601    -3.70   0.000    -.3061878   -.0942509
      totchr |   .4401428   .0093589    47.03   0.000     .4217975    .4584882
         age |  -.0053332   .0019369    -2.75   0.006    -.0091299   -.0015366
      female |   .0501264   .0252882     1.98   0.047     .0005566    .0996962
      blhisp |  -.1481236   .0341141    -4.34   0.000    -.2149941    -.081253
        linc |   .0252773   .0137866     1.83   0.067    -.0017472    .0523018
       _cons |   6.000931   .1559161    38.49   0.000     5.695304    6.306557
------------------------------------------------------------------------------

. reg hi_empunion multlc totchr age female blhisp linc , robust

Linear regression                                      Number of obs =   10089
                                                       F(  6, 10082) =  120.25
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0643
                                                       Root MSE      =   .4702

------------------------------------------------------------------------------
             |               Robust
 hi_empunion |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      multlc |   .1487593    .020504     7.26   0.000     .1085674    .1889513
      totchr |   .0109104   .0036859     2.96   0.003     .0036853    .0181354
         age |  -.0091799   .0007101   -12.93   0.000    -.0105717    -.007788
      female |  -.0792221   .0096843    -8.18   0.000    -.0982052    -.060239
      blhisp |  -.0741602   .0123788    -5.99   0.000    -.0984251   -.0498953
        linc |   .0720981   .0062189    11.59   0.000     .0599079    .0842883
       _cons |     .90169   .0589985    15.28   0.000     .7860412    1.017339
------------------------------------------------------------------------------
```

```
. correlate ldrugexp hi_empunion multlc, cov
(obs=10089)

             | ldrugexp hi_emp~n   multlc
-------------+---------------------------
    ldrugexp |  1.85519
 hi_empunion |  .021107  .236146
      multlc | -.016529  .014051  .058204


.
. reg hi_empunion multlc totchr age blhisp linc if female == 1 , robust

Linear regression                                      Number of obs =    5822
                                                       F(  5,  5816) =   75.57
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0618
                                                       Root MSE      =  .45887

------------------------------------------------------------------------------
             |               Robust
 hi_empunion |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      multlc |   .1667599   .0290762     5.74   0.000     .1097599      .22376
      totchr |   .0031734   .0047225     0.67   0.502    -.0060845    .0124314
         age |  -.0101485   .0008897   -11.41   0.000    -.0118926   -.0084043
      blhisp |  -.0628657   .0156485    -4.02   0.000    -.0935426   -.0321888
        linc |   .0749594   .0084465     8.87   0.000     .0584012    .0915176
       _cons |   .8998824   .0745252    12.07   0.000     .7537852     1.04598
------------------------------------------------------------------------------

. reg hi_empunion multlc totchr age blhisp linc if female == 0 , robust

Linear regression                                      Number of obs =    4267
                                                       F(  5,  4261) =   43.34
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0478
                                                       Root MSE      =  .48478

------------------------------------------------------------------------------
             |               Robust
 hi_empunion |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      multlc |   .1352384   .0288941     4.68   0.000     .0785909    .1918859
      totchr |   .0221272   .0059031     3.75   0.000     .0105541    .0337004
         age |  -.0075946   .0011705    -6.49   0.000    -.0098894   -.0052998
      blhisp |  -.0885161   .0200636    -4.41   0.000    -.1278512   -.0491811
        linc |   .0689149   .0092015     7.49   0.000     .0508751    .0869546
       _cons |   .7758469   .0960955     8.07   0.000     .5874497    .9642441
------------------------------------------------------------------------------

.
```

```
. table hi_empunion multlc if female == 1

-----------------------
Insured   |   Multiple
thro      |  locations
emp/union |     0      1
----------+------------
        0 | 3,721    125
        1 | 1,792    184
-----------------------

. table hi_empunion multlc if female == 0

-----------------------
Insured   |   Multiple
thro      |  locations
emp/union |     0      1
----------+------------
        0 | 2,267    120
        1 | 1,683    197
-----------------------
.
. table hi_empunion multlc if female == 0

-----------------------
Insured   |   Multiple
thro      |  locations
emp/union |     0      1
----------+------------
        0 | 2,267    120
        1 | 1,683    197
-----------------------


.
. table hi_empunion multlc, c(mean ldrugexp N ldrugexp)

------------------------------
Insured   |
thro      | Multiple locations
emp/union |        0         1
----------+-------------------
        0 | 6.464303  6.029153
          |    5,988       245
          |
        1 | 6.558737    6.3345
          |    3,475       381
------------------------------
```