Jialin Li
Professor Phillips (Renyu) Zhang
Final Project: Causal Inference in Experiment Practice
Due: 2021/5/20

Making Causal Inference with Instrumental Variable

Introduction

Discovering what effects do increasing Medicaid coverage has on emergency department use provides an important decision basis for policymakers. However, making a direct causal inference is difficult due to the potential unobservable observable confounders even with a larger sample and better observational design. Therefore, the purest way of the causal investigation is to introduce an extra well-randomized mechanism. For example, the lottery system in the *Oregan Health Plan.* The lottery is drawn purely at random but also correlates with the enrollment of the health care program (needed to be verified). Therefore, we may shift our attention from "Medicaid coverage → Emergency Department Use" to "Lottery Result → Emergency Department Use" with solid evidence.

Method and Procedure

Method

Though the lottery system is perceived as purely random, a balance check on the lottery drawing process is still required to ensure the validity of this instrumental variable. Therefore, I conduct a balance check on the variables of the first data set described below. Then, two-stage least squares is used in the model fitting process. In the first stage, we are to verify the transferability between the independent variable and instrumental variable, an OLS is fitted for the weak instrument test. Second, another OLS between the instrumental variable (lottery treatment) and dependent variables (whether visit and number of visits the emergence department).

Dataset

1. *oregonhie_descriptive_vars.dta (Z)*
   It contains demographic information about the lottery mechanism.
2. *oregonhie_stateprograms_vars.dta (W)*
   The feature *ohp_all_ever_firstn_30sep2009* in it contains the measures of insurance coverage.
3. *oregonhie_ed_vars.dta (Y)*
   It contains information about the number of emergency department visits (Mainly Look at the total number).

Procedure

Read the Data and Drop the Unnecessary Variables

Referencing the user guide, I keep these variables for my analysis:

1. *oregonhie_descriptive_vars.dta (Z)*

   numhh_list: Since the Medicaid opportunity is given in the unit of household. The number of registered members of each household will influence each member's probability of getting selected. Therefore, this must be included in the balance check. (The selection is still balanced under CIA)

birthyear_list: Older people may devote more to insurance while younger may not

first_day_list: Timing may result in a selection difference

Other selections: female_list/english_list/have_phone_list/pobox_list

2. *oregonhie_stateprograms_vars.dta (W)*

   ohp_all_ever_firstn_30sep2009: This variable was used as the definition of insurance coverage in estimating the effect of Medicaid.

3. *oregonhie_ed_vars.dta (Y)*

   any_visit_ed: This variable is equal to 1 if an individual had any ED visits between the treatment effective period.

   num_visit_ed: This variable is equal to the number of ED visits an individual had between the treatment effective period.

Dummy Textual Information

Change gender, English, phone, postbox information from text to 0-1.

Balance Check

The OLA results yield that most of the features are well-balanced except for the English level (I tried to put it in the dependent variable but then the OLS provides a worse result). Overall, no selection bias is observed in the lottery mechanism with most p-value bigger than 0.05.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         birthyear_list   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     2.734
Date:                Thu, 20 May 2021   Prob (F-statistic):             0.0650
Time:                        20:23:19   Log-Likelihood:                -96304.
No. Observations:               24646   AIC:                         1.926e+05
Df Residuals:                   24643   BIC:                         1.926e+05
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1967.8769      0.239   8218.458      0.000    1967.408    1968.346
treatment      0.0975      0.160      0.608      0.543      -0.217       0.412
numhh_list     0.4013      0.191      2.099      0.036       0.027       0.776
------------------------------------------------------------------------------
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            female_list   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                  0.002
Method:                 Least Squares   F-statistic:                     21.44
Date:                Thu, 20 May 2021   Prob (F-statistic):           4.98e-10
Time:                        20:24:01   Log-Likelihood:                -17758.
No. Observations:               24646   AIC:                         3.552e+04
Df Residuals:                   24643   BIC:                         3.555e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.6074      0.010     61.420      0.000       0.588       0.627
treatment     -0.0099      0.007     -1.497      0.134      -0.023       0.003
numhh_list    -0.0471      0.008     -5.969      0.000      -0.063      -0.032
------------------------------------------------------------------------------
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              self_list   R-squared:                       0.443
Model:                            OLS   Adj. R-squared:                  0.443
Method:                 Least Squares   F-statistic:                     9811.
Date:                Thu, 20 May 2021   Prob (F-statistic):               0.00
Time:                        20:25:39   Log-Likelihood:                 1715.4
No. Observations:               24646   AIC:                            -3425.
Df Residuals:                   24643   BIC:                            -3400.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.4915      0.004    332.379      0.000       1.483       1.500
treatment      0.0009      0.003      0.291      0.771      -0.005       0.007
numhh_list    -0.4927      0.004   -137.535      0.000      -0.500      -0.486
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         first_day_list   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     1.370
Date:                Thu, 20 May 2021   Prob (F-statistic):              0.254
Time:                        20:31:04   Log-Likelihood:                -4478.2
No. Observations:               24646   AIC:                             8962.
Df Residuals:                   24643   BIC:                             8987.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0937      0.006     16.233      0.000       0.082       0.105
treatment      0.0063      0.004      1.631      0.103      -0.001       0.014
numhh_list    -0.0027      0.005     -0.591      0.555      -0.012       0.006
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           english_list   R-squared:                       0.072
Model:                            OLS   Adj. R-squared:                  0.071
Method:                 Least Squares   F-statistic:                     949.1
Date:                Thu, 20 May 2021   Prob (F-statistic):               0.00
Time:                        20:28:22   Log-Likelihood:                -7617.3
No. Observations:               24646   AIC:                         1.524e+04
Df Residuals:                   24643   BIC:                         1.526e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.1329      0.007    172.879      0.000       1.120       1.146
treatment      0.0092      0.004      2.105      0.035       0.001       0.018
numhh_list    -0.2255      0.005    -43.114      0.000      -0.236      -0.215
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        have_phone_list   R-squared:                       0.003
Model:                            OLS   Adj. R-squared:                  0.002
Method:                 Least Squares   F-statistic:                     31.28
Date:                Thu, 20 May 2021   Prob (F-statistic):           2.71e-14
Time:                        20:33:44   Log-Likelihood:                -8098.6
No. Observations:               24646   AIC:                         1.620e+04
Df Residuals:                   24643   BIC:                         1.623e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.8197      0.007    122.675      0.000       0.807       0.833
treatment      0.0027      0.004      0.610      0.542      -0.006       0.011
numhh_list     0.0407      0.005      7.623      0.000       0.030       0.051
```

2SLS

Fit the model accordingly with ED together with numhh_list as dependent variable. The table shows that there is a positive causal relationship between the increase in visits and Medicaid coverage. And the second 2SLS gives even concrete support that a one unit increase in Medicaid coverage results in a 34.5% increase in overall emergency department use. Taubman et.al s' conclusion has been confirmed.

```
                          IV-2SLS Estimation Summary
==============================================================================
Dep. Variable:              num_visit   R-squared:                      0.0220
Estimator:                    IV-2SLS   Adj. R-squared:                 0.0219
No. Observations:               24622   F-statistic:                    468.38
Date:                Thu, May 20 2021   P-value (F-stat)                0.0000
Time:                        20:56:12   Distribution:                  chi2(2)
Cov. Estimator:                robust

                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
const          1.6197     0.0506     32.036     0.0000      1.5206      1.7188
numhh_list    -0.5866     0.0272    -21.596     0.0000     -0.6398     -0.5333
Enrollment     0.3454     0.1293     2.6705     0.0076      0.0919      0.5989
==============================================================================
```

Conclusion

Medicaid coverage significantly increases overall emergency use by 0.34 visits per person, close to the given number 0.41.