Homework 4

ECON 24450 Spring, 2021

Jake Underland

2021-05-27

Contents

Read the journal article posted on Canvas, "The Oregon Health Insurance Experiment: Evidence from the First Year," by Finkelstein et al. (QJE 2013). Provide brief answers to the following questions. The question is interesting because, although there have been many attemtps to elucidate the effects of medicaid on health and other individual outcomes, research has been confounded by difficulty of controlling for unobserved differences between insured and uninsured originating from selection. It is rare to have a random assignment setting where such confounding factors can be controlled.

Question 1.

What is the research question being addressed and why is it an interesting empirical question?

Solution.
The research question asks what the effect of Medicaid and expanding access to public health insurance is on the use of health care, financial strain, and health of low-income adults.

Question 2.

The following questions relate to the internal validity of the Oregon Health Insurance experiment (OHIE).

(a) What is the identifying assumption of the randomized controlled trial? What are possible violations of the identifying assumption?

Solution.
The identifying assumption is that the treated and untreated groups are identical other than the treatment they receive, i.e., assignment of the ability to apply for OHP Standard was random, and that the treatment and control individuals in the subsamples used to analyze outcomes are not differentially selected from the full sample. This can be violated by nonrandom assignment, or when certain types of people are more likely to select into either the treatment or control group. Even when we enforce random assignment in our experimental design, there may be correlations in treated

vs. untreated with other unobservable coavariates. Additionally, differential participation, where the participation rate across populations is nonrandom and certain subgroups are more likely to drop out when assigned to one of the treatment or control groups, can also bias estimates.

(b) What do Finkelstein et al. do to test for random assignment? Refer to specific page numbers, figures, and tables as appropriate.

Solution.

They verified the randomization procedure followed by Oregon's DHS through computer simulations and demonstrated balance of treatment and control characteristics in the online appendix and Table A12. Additionally, in Table II, they conduct a covariance balance test for differences in attrition for both credit reports and survey responses, nine "lotterly list variables" they constructed from prerandomization characteristics as well as for pre-randomization outcomes to examine treatment and control balance, found on page 1075.

(c) Why is it a problem if the treatment and control groups differentially report outcomes? Is differential reporting expected to be a problem in this context? Why or why not?

Solution.

Differential reporting biases estimates as it increases the weight of the outcomes more likely to be reported in the analysis. If a ceratin subset of people are more likely to report their outcomes in one group than the other, we can no longer say that the two groups we are examining (via the reports) are identical on average. Differential reporting in this paper is not a problem for the outcomes measured by administrative records, but is a problem for the survey data with an effective response rate of 50%. However, from the covariance balance test, the authors find a very small effect of treatment on the survey response rate.

(d) s imperfect take-up of Medicaid in the treatment group a threat to validity? Why or why not? Compare the consequences of imperfect take-up of Medicaid in the treatment group on the OHIE to the situation we saw in class with differential attrition across treatment groups in the RAND health insurance experiment.

Solution.

Imperfect take up of Medicaid in the treatment group is not a threat in this context. In terms of the ITT, the authors are interested in the effects of expansion of access to public health insurance, which does not require take-up of Medicaid after treatment. Second, in the IV model they employ to estimate the effect of actual coverage on Medicaid, the authors are estimating the local average treatment effect of the compliers in the dataset, to which the population that did not take-up medicaid do not belong and therefore pose no threat to the validity of the estimation method. In the RAND experiment, the causal relationship studied was not expansion of access to health care but actual enrollment. Examining this model required randomness and identicalness upon average across the groups, but since the distribution of refusal and attrition rates were nonrandom and increased significantly in the non-generous group, being unable to access administrative records, this led to a nonrandom loss of data for that group, potentially biasing estimates.

(e) What other possible violations of the exclusion restriction do Finkelstein et al. consider? How do they test for these violations?

Solution.

They consider the possibility of direct effects that the event of winning the lottery may have on the individual's outcomes, and whether application for Medicaid encourages individuals to apply for other public programs. These do not change the validity of the effects of expanded access to public health care, but do violate the exclusion principle necessary to pin these effects on actual coverage.

The authors conduct a covariate balance test reported in Table III. They conclude that selection by the lottery is not associated with TANF receipt and has a trivial association with food stamp receipt.

Question 3.

The following questions relate to the paper's estimation strategy.

(a) Write down the structural equation of interest. Write down the reduced form equation and first stage equation. Why do Finkelstein et al. say that there is ambiguity about what the correct endogeneous variable is?

Solution.
The structural equation of interest is:

$$y_{ihj} = \pi_o + \pi_1 INSURANCE_{ih} + X_{ih}\pi_2 + V_{ih}\pi_3 + v_{ihj}$$

And the first stage is:

$$INSURANCE_{ih} = \delta_0 + \delta_1 LOTTERY_h + X_{ih}\delta_2 + V_{ih}\delta_3 + \mu_{ihj}$$

And the reduced form and intent to treat effect is:

$$y_{ihj} = \beta_0 + \beta_1 LOTTERY_h + X_{ih}\beta_2 + V_{ih}\beta_3 + \epsilon_{ihj}$$

The authors use "Ever on Medicaid" as the endogenous variable, but other measures of insurance such as "the number of months on medicaid", and "on medicaid at the end of our study" are also useful in that the former may be more appropriate for modeling outcomes where the effect of insurance on outcomes is linear in the duration insured, and the latter provides an insightful lower bound on the first stage.

(b) Why do Finkelstein et al. include controls in equation (1)?

Solution.
The controls in X are correlated with treatment and must be controlled for in order to achieve randomness and get an unbiased estimate, as the identifying assumption of RCT requires. The controls in V are included for increased precision.

(c) What is the first stage estimate for ever on Medicaid? Interpret the estimate. Why is it less than one?

Solution.
The results show a first stage of 0.26 for the full sample and credit-report subsample and a first stage of 0.29 for the survey respondents. This implies that winning the lottery increases the chance of a household enrolling in medicaid by 30%. The coefficient is less than 1 because of the 30

(d) Describe who the complier population is for this local average treatment effect. How generalizable is this particular LATE is to the effect of Medicaid on other populations (e.g., children and pregnant women)? How generalizable is this LATE to the effect of the Medicaid expansions under the Affordable Care Act?

Solution.
The compiler population is somewhat older white people who are worse in health than average. Thus, the results of this paper are hard to generalize, as children are much younger than the complier population observed here, while pregnant women would possibly be in better health given their pregnancy, again deviating from the complier population described here. Medicaid expansions under the ACA mainly increased coverage of non-white minorities, and the age range that saw the largest increase in coverage was 19-34 as seen here. Thus, the effects of this study also seem difficult to generalize to expansions under the ACA.

Question 4.

This question asks you to consider the results of this paper, plus the results of the Oregon Health Insurance Experiment more generally.

(a) Summarize this paper's results on the effect of Medicaid on

i. Health care utilization

Solution.
Medicaid increased emergency care by 40%, outpatient office visits by 55%, prescription drug use by 15%, non-emergency department hospitalizations, and diabetes and depression diagnoses. In sum, Medicaid increased health care utilization for almost all variables measured save ER use.

ii. Financial strain

Solution.
The average standardized treatment effect suggests no evidence of a decline in bankruptcy, lien, judgment, collection, or delinquency. However, self reported surveys show that Medicaid eliminated out of pocket catastrophic expenditures by 35% and reduced probability of having to borrow money or skip paying other bills because of medical expenses by more than 40%. Credit reports tell a similar story, as Medicaid decreased probability of having unpaid medical bill sent to collection agency by 25%.

iii. Health

Solution.
Medicaid improved self-reported health and reduced depression. There is a 32% increase in self-reported overall happiness, and rates of depression dropped by 30%. The effects on phsyical health, on the other hand, were not statistically significant, even among higher risk groups.

(b) The Oregon Health Insurance Experiment has received extensive media coverage. In 2011, an NPR headline read "Medicaid Makes 'Big Difference' in Lives, Study Finds." In 2013, a Washington Post headline read "Spending on Medicaid doesn't actually help the poor." These headlines seem contradictory. Given the results of this paper and the other OHIE papers we summarized in class, which of these headlines do you think is correct and why?

Solution.
Although the effects of medicare can vary widely between complier populations, this paper and others seem to support the notion that Medicaid does improve the lives of lower socioeconomic individuals. From this paper alone, we find that Medicaid has benefits on mental health, hospital utilization, and consumption smoothing. Medicaid might not contribute, at least from this study, to physical health, but that does not mean that it does not significantly improve the lives of the poor. The 32% increase in self-reported overall happiness documented in this paper should be sufficient in dismissing that notion. Some argue that bankruptcy essentially acts as a standin for Medicaid among low income individuals, but even then, funding for Medicaid is being used to fund hospitals treating low income uninsured individuals who are not formal recipients of medicaid. Pooling all of these effects, not to mention the benefits of being relieved from debt and bankruptcy with the help of medicaid, it is safe to say that Medicaid does make a big difference in people's lives.

Question 5.

a.

```r
raw_df <- read.csv("pset_experiment_data.csv.crdownload")
```

b.

```r
df <- raw_df
# coding binary variables
```

```
df[df==""] <- NA
df$treatment <- as.numeric(df$treatment == "Selected", na.rm = T)
df$returned_12m <- as.numeric(df$returned_12m == "Yes", na.rm=T)
df$ohp_all_ever_survey <- as.numeric(df$ohp_all_ever_survey == "Enrolled", na.rm = T)
df$ohp_all_mo_survey <- as.integer(str_remove(df$ohp_all_mo_survey, " months|month"), na.rm = T)
X <- names(df)[grep("ddd", names(df))]
```

```
avg_treatment <- mean(df[df$treatment == 1,]$returned_12m, na.rm = T)
avg_control <- mean(df[df$treatment == 0,]$returned_12m, na.rm = T)
# Testing for difference in response rate
response_rate <- df %>%
  filter(treatment == 1 | treatment == 0) %>%
  select(treatment, returned_12m)
res <- t.test(returned_12m ~ treatment, data = df)
cat("Average 12 month survey response rate for treatment:", avg_treatment,
    "\nAverage 12 month survey response rate for control:", avg_control,
    "\nDifference:", abs(avg_treatment - avg_control))
```

```
## Average 12 month survey response rate for treatment: 0.3991686
## Average 12 month survey response rate for control: 0.4152554
## Difference: 0.0160868
```

```
res
```

```
##
##  Welch Two Sample t-test
##
## data:  returned_12m by treatment
## t = 3.9564, df = 58341, p-value = 7.618e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.008117376 0.024056231
## sample estimates:
## mean in group 0 mean in group 1
##       0.4152554       0.3991686
```

From the above results, there seems to be a statistically significant difference in the means of the treatment and control groups, but very small and trivial.

c.

i.

```
full_df <- df[!is.na(df$weight_12m), ]
sub_df <- df[df$sample_12m_resp == "12m mail survey responder", ]
sub_df <- sub_df[!is.na(sub_df$weight_12m), ]
```

```
# A
# Full sample
ever_lottery <- felm(as.formula(paste("ohp_all_ever_survey",
                            "~ treatment +", paste(X, collapse = "+"),
                            "|0|0|household_id")), data = full_df,
                            weights = full_df$weight_12m)
# survey respondents
ever_lottery_sub <- felm(as.formula(paste("ohp_all_ever_survey",
                            "~ treatment +", paste(X, collapse = "+"),
```

```r
                            "|0|0|household_id")), data = sub_df,
                            weights = sub_df$weight_12m)

# B
months_lottery <- felm(as.formula(paste("ohp_all_mo_survey",
                            "~ treatment +", paste(X, collapse = "+"),
                            "|0|0|household_id")), data = full_df,
                            weights = full_df$weight_12m)
months_lottery_sub <- felm(as.formula(paste("ohp_all_mo_survey",
                            "~ treatment +", paste(X, collapse = "+"),
                            "|0|0|household_id")), data = sub_df,
                            weights = sub_df$weight_12m)
# C
end_lottery <- felm(as.formula(paste("ohp_all_end_survey",
                            "~ treatment +", paste(X, collapse = "+"),
                            "|0|0|household_id")), data = full_df,
                            weights = full_df$weight_12m)
end_lottery_sub <- felm(as.formula(paste("ohp_all_end_survey",
                            "~ treatment +", paste(X, collapse = "+"),
                            "|0|0|household_id")), data = sub_df,
                            weights = sub_df$weight_12m)
```

ii.

Table 1: First Stage Results

	Dependent variable:					
	Ever on Medicaid		Num of Months on Medicaid		On Medicaid, end of study period	
	Full Sample	Survey Respondents	Full Sample	Survey Respondents	Full Sample	Survey Respondents
	(1)	(2)	(3)	(4)	(5)	(6)
treatment	0.255^{***}	0.290^{***}	3.293^{***}	3.943^{***}	0.148^{***}	0.189^{***}
	(0.006)	(0.007)	(0.072)	(0.090)	(0.005)	(0.006)
Constant	0.188^{***}	0.209^{***}	2.430^{***}	2.914^{***}	0.114^{***}	0.138^{***}
	(0.010)	(0.013)	(0.148)	(0.185)	(0.009)	(0.012)
fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	58,405	23,741	58,405	23,741	58,405	23,741
R^2	0.086	0.110	0.085	0.112	0.038	0.058
Adjusted R^2	0.086	0.109	0.084	0.111	0.038	0.057

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

d.

i.

```
# creating variables
full_df$rx_any_12m <- as.numeric(full_df$rx_any_12m == "Yes", na.rm = T)
full_df$doc_any_12m <- as.numeric(full_df$doc_any_12m == "Yes", na.rm = T)
full_df$er_any_12m <- as.numeric(full_df$er_any_12m == "Yes", na.rm = T)
full_df$hosp_any_12m <- as.numeric(full_df$hosp_any_12m == "Yes", na.rm = T)
full_df$cost_any_oop_12m <- as.numeric(full_df$cost_any_oop_12m == "Yes", na.rm = T)
full_df$cost_any_owe_12m <- as.numeric(full_df$cost_any_owe_12m == "Yes", na.rm = T)
#full_df$health_notpoor_12m <- as.numeric(full_df$health_notpoor_12m == "Yes", na.rm = T)
#full_df$notbaddays_tot_12m <- as.numeric(full_df$notbaddays_tot_12m == "Yes", na.rm = T)

sub_df <- full_df[full_df$sample_12m_resp == "12m mail survey responder", ]
sub_df <- sub_df[!is.na(sub_df$weight_12m), ]
```

```
# A
# survey respondents
prescription.sub <- felm(as.formula(paste("rx_any_12m ~ ",
                         paste(X, collapse = " + "),
                         "| 0 | (ohp_all_ever_survey",
                         "~ treatment)",
                         " | household_id")), data = sub_df,
                         weights = sub_df$weight_12m)
# B
# survey respondents
primary.care.sub <- felm(as.formula(paste("doc_any_12m ~ ",
                         paste(X, collapse = " + "),
                         "| 0 | (ohp_all_ever_survey",
                         "~ treatment)",
                         " | household_id")), data = sub_df,
                         weights = sub_df$weight_12m)
# C
# survey respondents
ER.visits.sub <- felm(as.formula(paste("er_any_12m ~ ",
                       paste(X, collapse = " + "),
                       "| 0 | (ohp_all_ever_survey",
                       "~ treatment)",
                       " | household_id")), data = sub_df,
                       weights = sub_df$weight_12m)
# D
# survey respondents
hosp.visits.sub <- felm(as.formula(paste("hosp_any_12m ~ ",
                     paste(X, collapse = " + "),
                     "| 0 | (ohp_all_ever_survey",
                     "~ treatment)",
                     " | household_id")), data = sub_df,
                     weights = sub_df$weight_12m)
# E
# survey respondents
oop.cost.sub <- felm(as.formula(paste("cost_any_oop_12m ~ ",
                         paste(X, collapse = " + "),
                         "| 0 | (ohp_all_ever_survey",
                         "~ treatment)",
```

```
                           " | household_id")), data = sub_df,
                           weights = sub_df$weight_12m)
# F
# survey respondents
owe.medexp.sub <- felm(as.formula(paste("cost_any_owe_12m ~ ",
                           paste(X, collapse = " + "),
                           "| 0 | (ohp_all_ever_survey",
                           "~ treatment)",
                           " | household_id")), data = sub_df,
                           weights = sub_df$weight_12m)
# G
# survey respondents
health.good.sub <- felm(as.formula(paste("health_notpoor_12m ~ ",
                           paste(X, collapse = " + "),
                           "| 0 | (ohp_all_ever_survey",
                           "~ treatment)",
                           " | household_id")), data = sub_df,
                           weights = sub_df$weight_12m)
# H
# survey respondents
not.bad.days.sub <- felm(as.formula(paste("notbaddays_tot_12m ~ ",
                           paste(X, collapse = " + "),
                           "| 0 | (ohp_all_ever_survey",
                           "~ treatment)",
                           " | household_id")), data = sub_df,
                           weights = sub_df$weight_12m)
```

ii.

```
# A
# survey respondents
prescription.ols <- felm(as.formula(paste("rx_any_12m",
                              "~ treatment +", paste(X, collapse = "+"),
                              "|0|0|household_id")), data = sub_df,
                              weights = sub_df$weight_12m)

# B
primary.care.ols <- felm(as.formula(paste("doc_any_12m ~ ",
                              "treatment +", paste(X, collapse = "+"),
                              "|0|0|household_id")), data = sub_df,
                              weights = sub_df$weight_12m)
# C
ER.visits.ols <- felm(as.formula(paste("er_any_12m ~ ",
                              "treatment +", paste(X, collapse = "+"),
                              "|0|0|household_id")), data = sub_df,
                              weights = sub_df$weight_12m)
# D
hosp.visits.ols <- felm(as.formula(paste("hosp_any_12m ~ ",
                              "treatment +", paste(X, collapse = "+"),
                              "|0|0|household_id")), data = sub_df,
                              weights = sub_df$weight_12m)
# E
oop.cost.ols <- felm(as.formula(paste("cost_any_oop_12m ~ ",
```

```r
                              "treatment +", paste(X, collapse = "+"),
                              "|0|0|household_id")), data = sub_df,
                              weights = sub_df$weight_12m)
# F
owe.medexp.ols <- felm(as.formula(paste("cost_any_owe_12m ~ ",
                              "treatment +", paste(X, collapse = "+"),
                              "|0|0|household_id")), data = sub_df,
                              weights = sub_df$weight_12m)
# G
health.good.ols <- felm(as.formula(paste("health_notpoor_12m ~ ",
                              "treatment +", paste(X, collapse = "+"),
                              "|0|0|household_id")), data = sub_df,
                              weights = sub_df$weight_12m)
# H
not.bad.days.ols <- felm(as.formula(paste("notbaddays_tot_12m ~ ",
                              "treatment +", paste(X, collapse = "+"),
                              "|0|0|household_id")), data = sub_df,
                              weights = sub_df$weight_12m)
```

iii.

2SLS rsults reported in Table 2, ITT results reported in Table 3.

iv.

The coefficient on insurance from IV estimation of equation (3) can be interpreted as a local average treatment effect of insurance, or the causal impact of insurance among the compliers who would usually not obtain insurance.

The coefficient on the lottery dummy gives the average difference in means between the treated and control groups, or the impact of expanded *access to health care* on various outcome variables.

Table 2: 2SLS Results

	Dependent variable:							
	Prescription	Primary Care	ER visits	Hospital visits	Out of Pocket Costs	Owe Expenses	Health not bad	Not Bad Days
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LATE	0.088***	0.212***	0.022	0.008	−0.200***	−0.180***	0.099***	1.317**
	(0.029)	(0.025)	(0.023)	(0.014)	(0.026)	(0.026)	(0.018)	(0.563)
Constant	0.623***	0.555***	0.267***	0.080***	0.597***	0.645***	0.837***	20.736***
	(0.019)	(0.016)	(0.015)	(0.009)	(0.017)	(0.017)	(0.011)	(0.360)
fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,308	23,492	23,514	23,573	23,426	23,451	23,361	21,881
R^2	0.028	0.036	0.010	0.006	0.014	-0.006	-0.019	0.001
Adjusted R^2	0.027	0.035	0.009	0.005	0.013	-0.007	-0.020	-0.0002

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: ITT:OLS

	Dependent variable:							
	Prescription	Primary Care	ER visits	Hospital visits	Out of Pocket Costs	Owe Expenses	Health not bad	Not Bad Days
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITT	0.025***	0.062***	0.006	0.002	−0.058***	−0.052***	0.029***	0.381**
	(0.008)	(0.007)	(0.007)	(0.004)	(0.008)	(0.008)	(0.005)	(0.162)
Constant	0.642***	0.599***	0.272***	0.081***	0.556***	0.607***	0.858***	21.003***
	(0.016)	(0.014)	(0.013)	(0.008)	(0.014)	(0.014)	(0.009)	(0.307)
fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,308	23,492	23,514	23,573	23,426	23,451	23,361	21,881
R^2	0.016	0.007	0.006	0.004	0.005	0.009	0.004	0.014
Adjusted R^2	0.015	0.006	0.006	0.003	0.004	0.008	0.003	0.013

Note: *p<0.1; **p<0.05; ***p<0.01