# Inequality Pset 4

Julia Du

2021-05-26

### Load necessary libraries

```
library(tidyverse)
library(lubridate)
library(tinytex)
library(stargazer)
library(lfe)

theme_set(theme_minimal())
```

### Question 4a

The causal relationship of interest is the effect of income (where your income is either increased by the changed EITC) on infant health outcomes, specifically infant birth weight.

The structural equation is:

$$Y_{ipjst} = \beta_0 + \beta_1(Income_{ipjst}) + \varepsilon_{ipjst}$$

where $i$ is the individual. The other subscripts will be discussed later.

### Question 4b

If we estimate the structural equation in cross-sectional data, then our estimate of the causal parameter of interest $\beta_1$ will suffer from Omitted Variable Bias (OVB), as income is endogeneous because it is correlated with unobserved health care inputs. In general, those with better health tend to have higher income.

Recall:

$$OVB = \beta_1^{OLS} - \beta_1 = \frac{cov(Income_{ipjst}, \varepsilon_{ipjst})}{var(Income_{ipjst})}$$

So, the direction of bias depends on the sign of the covariance. In this case, it's likely our estimate is upward-biased since income tends to be positively correlated with unobserved health.

## Question 4c

**4c(i).** I have included a graph of this a few pages earlier.

The 1st diff-in-diff is:

$$\bar{Y}_{2plus,after} - \bar{Y}_{1,after}$$

This is the difference in infant health between families with 2 or more kids and with families with just one kid after the EITC expansion occurred.

The 2nd diff-in-diff is:

$$\bar{Y}_{2plus,before} - \bar{Y}_{1,before}$$

This is the difference in infant health between families with 2 or more kids and with families with just one kid before the EITC expansion occurred.

**4c(ii).** The key identifying assumption is the parallel assumption, i.e. that absent the EITC expansion policy change (OBRA93), both families with one child and families with two or more children follow parallel trends over time in their infant health outcomes.

**4c(iii).** One possible violation could be that there is a macro shock occurs sometime before the EITC expansion, and the shock affects the two family types differently. For example, maybe there's another policy that seeks to discourage families from having more children and thus gives HHs with only 1 kid bigger subsidies. In that case, families with 1 kid and those with 2 or more would be on different trends.

## Question 4d

The **"reduced-form" equation** is:

$$Y_{pjst} = \alpha + \delta After_t \times Parity2plus_p + \beta X_{st} + \gamma_p + \eta_s + \delta_t + \phi_j + \varepsilon_{pjst}$$

where $Y_{pjst}$ is a measure of infant health (specifically, the fraction of low birth weight infants multiplied by 100) for the cell defined by parity $p$, demographic group $j$, in state $s$ for effective tax year $t$. $\gamma_p$ is a set of dummy variables for birth order, $\eta_s$ is a set of dummy variables for state of residence, $\delta_t$ is a set of dummy variables for effective tax year. We also include fixed effects for demographic group $\phi_j$. $X_{st}$ includes controls for unemployment rate, welfare reform and Medicaid or SCHIP eligibility. $a$ is the intercept, representing the baseline infant health (i.e. for families with a first-order birth before the policy expansion). $\varepsilon_{pjst}$ represents the unobserved variation.

$After$ is a dummy variable equaling one for effective tax years 1994 through 1998, $Parity2plus_p$ is a dummy variable indicating if a birth is second or higher order. Their interaction lets us make use of DD strategy in trying to suss out the difference in infant health outcomes before and after the EITC policy change, while also factoring in the difference between the treated and control groups (families with 2nd-order or higher births and families with 1st-order birth, respectively)

$\delta$ is our coefficient of interest, i.e. the DD estimate. It shows the effect of the treatment (i.e. policy expansion) on the treated's (in this case, families whose birth is 2nd-order or higher) infant birth weight.

## Question 4e

Alternative specification to equation:

$$Y_{pjst} = \alpha + \delta After_t \times Parity2plus_p + \varphi_1 After_t + \varphi_2 Parity2plus_p + \beta \tilde{X}_{pjst} + \varepsilon_{pjst}$$

In this case, $\varphi_1$ represents the difference in the outcome before & after the EITC policy change, while $\varphi_2$ represents the difference between the treated and control groups.

## Question 5: Data exercise

Just so you know, the dummy syntax for the `felm` command is as follows:

`felm(causal relation of interest | fixed effects | IVs | clusters, data = your_data)`

```
q4 <- read_csv("./dataexercise_pset4/pset_experiment_data.csv")

# 5b
q4 <- q4 %>%
  drop_na(treatment) %>%
  mutate(
    treatment = if_else(treatment == "Selected", 1, 0),
    returned_12m = if_else(returned_12m == "Yes", 1, 0))

q4 %>%
  drop_na(returned_12m) %>% # CAN I DROP NA
  group_by(treatment) %>%
  summarize(response_rate = mean(returned_12m))
```

```
## # A tibble: 2 x 2
##   treatment response_rate
##       <dbl>         <dbl>
## 1         0         0.415
## 2         1         0.399
```

```
q4 %>%
  count(returned_12m)
```

```
## # A tibble: 3 x 2
##   returned_12m     n
##          <dbl> <int>
## 1            0 34628
## 2            1 23777
## 3           NA 16517
```

(5b) The response rates are pretty similar across the control and treatment groups (41.5% and 39.9%, respectively). Given this closeness, I'd say there isn't evidence of differential survey response rates.

## Question 5c

```
q4 <- q4 %>%
  mutate(
    ohp_all_mo_survey = parse_number(ohp_all_mo_survey),
    ohp_all_ever_survey = if_else(ohp_all_ever_survey == "Enrolled", 1, 0))

dummies_q4 <- q4 %>%
  select(starts_with ("ddd")) %>%
  colnames()

# drop NAs for weights to
```

```
q4_test <- q4 %>%
  drop_na(weight_12m) %>%
  filter(sample_12m_resp == "12m mail survey responder")

## does including the fixed effs ("ddd" var) include the regular HH size dummy var & survey wave dummy v

# fyi: if put in fixed effs, you might be creating the dummies & interactions - but we already have cre
# column 6, table iii

q4_1ststage_ols_ever <- felm(as.formula(
  paste("ohp_all_ever_survey ~  treatment", "+",
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test, weights = q4_test$weight_12m)

q4_1ststage_ols_mo <- felm(as.formula(
  paste("ohp_all_mo_survey ~  treatment", "+",
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test, weights = q4_test$weight_12m)

q4_1ststage_ols_end <- felm(as.formula(
  paste("ohp_all_end_survey ~  treatment", "+",
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test, weights = q4_test$weight_12m)

#q4_1ststage_ols_end %>%
 # summary("robust")
```

### Question 5c(ii): 1st stage OLS results

```
stargazer(q4_1ststage_ols_ever, q4_1ststage_ols_mo, q4_1ststage_ols_end, type = "latex",
          keep = "treatment",
          title = "Question 5c(ii)",
          dep.var.labels =c("Ever on Medicaid", "Number of months on Medicaid", "On Medicaid at end"),
          omit.stat =c("f","rsq","adj.rsq","ser"),
          dep.var.caption = "OLS 1st stage",
          digits = 4)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Wed, May 26, 2021 - 9:56:37 PM

### Question 5d(i)

```
#healthnotpoor has 1, 0, & NA. notbaddays is dbl w/ NAs - don't need if_else for them, but might need d

q4 <- q4 %>%
  mutate(
```

Table 1: Question 5c(ii)

| | OLS 1st stage | | |
|---|---|---|---|
| | Ever on Medicaid | Number of months on Medicaid | On Medicaid at end |
| | (1) | (2) | (3) |
| treatment | 0.2902*** | 3.9427*** | 0.1890*** |
| | (0.0066) | (0.0896) | (0.0061) |
| Observations | 23,741 | 23,741 | 23,741 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

```
    rx_any_12m = if_else(rx_any_12m == "Yes", 1, 0),
    doc_any_12m = if_else(doc_any_12m == "Yes", 1, 0),
    er_any_12m = if_else(er_any_12m == "Yes", 1, 0),
    hosp_any_12m = if_else(hosp_any_12m == "Yes", 1, 0),
    cost_any_oop_12m = if_else(cost_any_oop_12m == "Yes", 1, 0),
    cost_any_owe_12m = if_else(cost_any_owe_12m == "Yes", 1, 0),
    )

# drop NAs for weights too & select only survey data
# (have to rerun same command as before since we changed q4 above)
q4_test_1 <- q4 %>%
  #drop_na(weight_12m) %>%
  filter(sample_12m_resp == "12m mail survey responder")

q4_struct_ols_a <- felm(as.formula(
  paste("rx_any_12m ~ ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | (ohp_all_ever_survey ~ treatment) | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_struct_ols_a <- felm(as.formula(
  paste("rx_any_12m ~ ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | (ohp_all_ever_survey ~ treatment) | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_struct_ols_b <- felm(as.formula(
  paste("doc_any_12m ~ ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | (ohp_all_ever_survey ~ treatment) | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_struct_ols_c <- felm(as.formula(
  paste("er_any_12m ~ ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | (ohp_all_ever_survey ~ treatment) | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_struct_ols_d <- felm(as.formula(
```

```r
  paste("hosp_any_12m ~ ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | (ohp_all_ever_survey ~ treatment) | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_struct_ols_e <- felm(as.formula(
  paste("cost_any_oop_12m ~ ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | (ohp_all_ever_survey ~ treatment) | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_struct_ols_f <- felm(as.formula(
  paste("cost_any_owe_12m ~ ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | (ohp_all_ever_survey ~ treatment) | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_struct_ols_g <- felm(as.formula(
  paste("health_notpoor_12m ~ ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | (ohp_all_ever_survey ~ treatment) | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_struct_ols_h <- felm(as.formula(
  paste("notbaddays_tot_12m ~ ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | (ohp_all_ever_survey ~ treatment) | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_struct_ols_f %>%
  summary("robust")
```

```
##
## Call:
##    felm(formula = as.formula(paste("cost_any_owe_12m ~ ", paste(dummies_q4,      collapse = " + "),
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1798 -0.5734  0.3473  0.4370  1.2224
##
## Coefficients:
##                              Estimate Cluster s.e. t value Pr(>|t|)
## (Intercept)                  0.645080     0.016833  38.322  < 2e-16 ***
## ddddraw_sur_2                0.002546     0.019126   0.133   0.8941
## ddddraw_sur_3                0.001773     0.019251   0.092   0.9266
## ddddraw_sur_4               -0.002577     0.018148  -0.142   0.8871
## ddddraw_sur_5               -0.009894     0.018176  -0.544   0.5862
## ddddraw_sur_6               -0.018467     0.016950  -1.089   0.2759
## ddddraw_sur_7                0.007669     0.016530   0.464   0.6427
## dddnumhh_li_2               -0.119395     0.023298  -5.125 3.00e-07 ***
## dddnumhh_li_3               -0.251549     0.170246  -1.478   0.1395
## ddddraXnum_2_2               0.045128     0.033603   1.343   0.1793
## ddddraXnum_2_3               0.229080     0.216548   1.058   0.2901
```

```
## ddddraXnum_3_2              0.050895     0.033852   1.503   0.1327
## ddddraXnum_3_3              0.322450     0.236543   1.363   0.1728
## ddddraXnum_4_2              0.020385     0.031788   0.641   0.5214
## ddddraXnum_5_2              0.047253     0.031888   1.482   0.1384
## ddddraXnum_6_2              0.058068     0.029421   1.974   0.0484 *
## ddddraXnum_7_2              0.025103     0.056812   0.442   0.6586
## `ohp_all_ever_survey(fit)` -0.179834     0.026455  -6.798 1.09e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5513 on 23433 degrees of freedom
##   (290 observations deleted due to missingness)
## Multiple R-squared(full model): -0.00636   Adjusted R-squared: -0.00709
## Multiple R-squared(proj model): -0.00636   Adjusted R-squared: -0.00709
## F-statistic(full model, *iid*):12.22 on 17 and 23433 DF, p-value: < 2.2e-16
## F-statistic(proj model): 9.108 on 17 and 20761 DF, p-value: < 2.2e-16
## F-statistic(endog. vars):46.21 on 1 and 20761 DF, p-value: 1.091e-11
```

### Question 5d(ii)

```r
q4_itt_ols_a <- felm(as.formula(
  paste("rx_any_12m ~ treatment + ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_itt_ols_b <- felm(as.formula(
  paste("doc_any_12m ~ treatment + ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_itt_ols_c <- felm(as.formula(
  paste("er_any_12m ~ treatment + ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_itt_ols_d <- felm(as.formula(
  paste("hosp_any_12m ~ treatment + ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_itt_ols_e <- felm(as.formula(
  paste("cost_any_oop_12m ~ treatment + ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_itt_ols_f <- felm(as.formula(
  paste("cost_any_owe_12m ~ treatment + ",
```

```r
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_itt_ols_g <- felm(as.formula(
  paste("health_notpoor_12m ~ treatment + ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

q4_itt_ols_h <- felm(as.formula(
  paste("notbaddays_tot_12m ~ treatment + ",
        paste(dummies_q4, collapse = " + "),
        "| 0 | 0 | household_id", sep = "")),
  data = q4_test_1, weights = q4_test_1$weight_12m)

#q4_itt_ols_a %>%
 # summary("robust")
```

, q4__struct__ols__b, q4__struct__ols__c, q4__struct__ols__d, q4__struct__ols__d, q4__struct__ols__b, q4__struct__ols__c, q4__struct__ols__d, q4__struct__ols__d

### Question 5d(iii): $\hat{\pi}$ Coefficients

```r
stargazer(q4_struct_ols_a, q4_struct_ols_b, q4_struct_ols_c, q4_struct_ols_d, q4_struct_ols_e, q4_struc
          type = "latex",
          omit = c(dummies_q4, "Constant"),
          title = "Question 5d(iii): Pi Estimates (Structural Eqn)",
          covariate.labels = "Ever on Medicaid",
          dep.var.labels = c("rx", "doc", "er", "hosp", "cost oop", "cost owe", "health", "notbad"),
          omit.stat =c("f","rsq","adj.rsq","ser"),
          dep.var.caption = "",
          digits = 4,
         font.size = "small",
         column.sep.width = "0pt")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Wed, May 26, 2021 - 9:56:43 PM

Table 2: Question 5d(iii): Pi Estimates (Structural Eqn)

| | rx | doc | er | hosp | cost oop | cost owe | health | notbad |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Ever on Medicaid | $0.0878^{***}$ | $0.2124^{***}$ | 0.0223 | 0.0077 | $-0.1995^{***}$ | $-0.1798^{***}$ | $0.0990^{***}$ | $1.3171^{**}$ |
| | (0.0288) | (0.0252) | (0.0231) | (0.0136) | (0.0262) | (0.0265) | (0.0176) | (0.5629) |
| Observations | 18,308 | 23,492 | 23,514 | 23,573 | 23,426 | 23,451 | 23,361 | 21,881 |
| *Note:* | | | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | | |

## Question 5d(iii): $\hat{\beta}$ coefficients

```r
stargazer(q4_itt_ols_a, q4_itt_ols_b, q4_itt_ols_c, q4_itt_ols_d, q4_itt_ols_e, q4_itt_ols_f, q4_itt_ols
          type = "latex",
          omit = c(dummies_q4, "Constant"),
          title = "Question 5d(iii): Beta Estimates (ITT Equation)",
          covariate.labels = "Ever on Medicaid",
          dep.var.labels = c("rx", "doc", "er", "hosp", "cost oop", "cost owe", "health", "notbad"),
          omit.stat =c("f","rsq","adj.rsq","ser"),
          dep.var.caption = "",
          digits = 4,
         font.size = "small",
         column.sep.width = "0pt")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Wed, May 26, 2021 - 9:56:43 PM

Table 3: Question 5d(iii): Beta Estimates (ITT Equation)

| | rx | doc | er | hosp | cost oop | cost owe | health | notbad |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Ever on Medicaid | $0.0252^{***}$ | $0.0617^{***}$ | 0.0065 | 0.0022 | $-0.0580^{***}$ | $-0.0523^{***}$ | $0.0288^{***}$ | $0.3810^{**}$ |
| | (0.0083) | (0.0074) | (0.0067) | (0.0040) | (0.0077) | (0.0076) | (0.0051) | (0.1618) |
| Observations | 18,308 | 23,492 | 23,514 | 23,573 | 23,426 | 23,451 | 23,361 | 21,881 |
| *Note:* | | | | | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |