Inequality PSet 1 - Question #6d (code part)

Julia Du

2021-04-14

Load necessary libraries

```
library(tidyverse)
library(lubridate)
library(broom)
library(stats)
library(ivreg)
library(lmtest)
library(sandwich)
library(moderndive)
library(kableExtra)
library(tinytex)

#didn't end up using library(stargazer) or library(lfe)

theme_set(theme_minimal())
```

Question 6d(i): Construct the outcome variables

```
q1 <- read_csv("./DataExercise_IV/pset_iv_data.csv") %>%
  rename(pop = Pop, earnPOW = EarnPOW, man = Manufact)

# 6d(i).A: transform var by taking nat logs
q1 <- q1 %>%
  mutate(lnpay = log(pay),
         lnpop = log(pop),
         lnearnPOW = log(earnPOW), # county earnings
         lndSSI = log(dSSI)) # SSI payments

# 6d(i).B: construct log diff
q1 <- q1 %>%
  group_by(fips) %>%
  mutate(lagpay = lag(lnpay, n = 1),
         lagpop = lag(lnpop, n = 1),
         lagearnPOW = lag(lnearnPOW, n = 1),
         lagdSSI = lag(lndSSI, n = 1)) %>%
  mutate(diff_pay = lnpay - lagpay,
         diff_pop = lnpop - lagpop,
         diff_earnPOW = lnearnPOW - lagearnPOW,
         diff_dSSI = lndSSI - lagdSSI)

# 6d(i).C: construct var that is
# fraction of county earnings from manufacturing in 1969
q1 <-
```

```r
  q1 %>%
  filter(year == 1969) %>%
  mutate(frac_earn = man/earnPOW) %>%
  select(frac_earn, fips) %>%
  left_join(q1, by = "fips")
```

Question 6d(ii): Construct the instruments

```r
# construct coal price
q1 <- q1 %>%
  mutate(coalprice = pcoalyr/pyr)

# construct change in coal reserves
q1 <- q1 %>%
  group_by(fips) %>%
  mutate(ln_coalprice = log(coalprice),
         lag_coalprice = lag(ln_coalprice, n = 1),
         diff_coalprice = ln_coalprice - lag_coalprice) %>%
  mutate(ln_coalres = log(coalres)) %>%
  mutate(valchange_cr =
           if_else(coalres > 0,
                   diff_coalprice * ln_coalres, 0)) %>%
  mutate(valchange_crlag1 = lag(valchange_cr, n = 1),
         valchange_crlag2 = lag(valchange_cr, n = 2))
```

Question 6d(iii).A: OLS structural eqns

```r
# construct OLS regressions
ols_ssi_c <- lm(diff_dSSI ~ diff_earnPOW + as.factor(year)*as.factor(state) +
                  msa + lnpop + diff_pop + frac_earn, data = q1)

ols_ssi_nc <- lm(diff_dSSI ~ diff_earnPOW +
                   as.factor(year)*as.factor(state), data = q1)

ols_di_c <- lm(diff_pay ~ diff_earnPOW + as.factor(year)*as.factor(state) +
                 msa + lnpop + diff_pop + frac_earn, data = q1)

ols_di_nc <- lm(diff_pay ~ diff_earnPOW +
                  as.factor(year)*as.factor(state), data = q1)

# find estimates & robust SEs
coeftest(ols_ssi_c, vcov = vcovHC(ols_ssi_c, type = "HC1")) %>%
  broom::tidy()
```

```
## # A tibble: 101 x 5
##    term                 estimate std.error statistic   p.value
##    <chr>                   <dbl>     <dbl>     <dbl>     <dbl>
##  1 (Intercept)            0.0253   0.00262      9.67 5.23e- 22
##  2 diff_earnPOW          -0.0264   0.0136      -1.94 5.23e-  2
```

```
##  3 as.factor(year)1971 -0.0127    0.00559     -2.28 2.26e-  2
##  4 as.factor(year)1972 -0.0481    0.00584     -8.24 2.07e- 16
##  5 as.factor(year)1973  0.0661    0.00727      9.09 1.21e- 19
##  6 as.factor(year)1974  0.566     0.0174      32.6  4.15e-218
##  7 as.factor(year)1975  0.0447    0.00480      9.32 1.55e- 20
##  8 as.factor(year)1976 -0.00993   0.00382     -2.60 9.38e-  3
##  9 as.factor(year)1977 -0.0267    0.00470     -5.67 1.47e-  8
## 10 as.factor(year)1978 -0.0170    0.00405     -4.20 2.69e-  5
## # ... with 91 more rows
```

```
coeftest(ols_ssi_nc, vcov = vcovHC(ols_ssi_nc, type = "HC1")) %>%
  tidy()
```

```
## # A tibble: 121 x 5
##    term                estimate  std.error statistic p.value
##    <chr>                  <dbl>      <dbl>     <dbl>   <dbl>
##  1 (Intercept)          0.118    0.000794      149.      0
##  2 diff_earnPOW        -0.0200   0.0132         -1.51   0.132
##  3 as.factor(year)1971 -0.0454   0.0000558    -813.      0
##  4 as.factor(year)1972 -0.0571   0.000509     -112.      0
##  5 as.factor(year)1973 -0.0912   0.000731     -125.      0
##  6 as.factor(year)1974  0.309    0.000253     1221.      0
##  7 as.factor(year)1975 -0.00395  0.00000513   -770.      0
##  8 as.factor(year)1976 -0.0789   0.000605     -130.      0
##  9 as.factor(year)1977 -0.0891   0.000596     -149.      0
## 10 as.factor(year)1978 -0.0765   0.000850      -90.1     0
## # ... with 111 more rows
```

```
coeftest(ols_di_c, vcov = vcovHC(ols_di_c, type = "HC1")) %>%
  tidy()
```

```
## # A tibble: 93 x 5
##    term                estimate std.error statistic   p.value
##    <chr>                  <dbl>     <dbl>     <dbl>     <dbl>
##  1 (Intercept)          0.223     0.00959    23.2   6.97e-115
##  2 diff_earnPOW        -0.00216   0.0173     -0.125 9.00e-  1
##  3 as.factor(year)1971 -0.0578    0.0119     -4.83  1.36e-  6
##  4 as.factor(year)1972  0.0807    0.0109      7.38  1.74e- 13
##  5 as.factor(year)1973 -0.0989    0.0105     -9.40  7.46e- 21
##  6 as.factor(year)1974 -0.0134    0.0102     -1.31  1.91e-  1
##  7 as.factor(year)1975 -0.0276    0.0108     -2.56  1.06e-  2
##  8 as.factor(year)1976 -0.0479    0.0109     -4.38  1.22e-  5
##  9 as.factor(year)1977 -0.0772    0.0113     -6.81  1.03e- 11
## 10 as.factor(year)1978 -0.113     0.0105    -10.7   1.16e- 26
## # ... with 83 more rows
```

```
coeftest(ols_di_nc, vcov = vcovHC(ols_di_nc, type = "HC1")) %>%
  tidy()
```

```
## # A tibble: 89 x 5
##    term                estimate std.error statistic   p.value
##    <chr>                  <dbl>     <dbl>     <dbl>     <dbl>
```

```
##  1 (Intercept)          0.215     0.00882   24.3    8.00e-126
##  2 diff_earnPOW         0.00159   0.0166     0.0956 9.24e-   1
##  3 as.factor(year)1971 -0.0577    0.0117    -4.95   7.62e-   7
##  4 as.factor(year)1972  0.0836    0.0109     7.68   1.73e- 14
##  5 as.factor(year)1973 -0.0969    0.0104    -9.31   1.71e- 20
##  6 as.factor(year)1974 -0.0106    0.0102    -1.03   3.01e-   1
##  7 as.factor(year)1975 -0.0251    0.0105    -2.39   1.69e-   2
##  8 as.factor(year)1976 -0.0473    0.0107    -4.44   9.18e-   6
##  9 as.factor(year)1977 -0.0772    0.0110    -7.01   2.57e- 12
## 10 as.factor(year)1978 -0.110     0.0103   -10.7    1.62e- 26
## # ... with 79 more rows
```

NEED TO INTERPRET MAIN ESTIMATES (that's just the estimated coeff for diff_earnPOW aka log diff in county earnings, right?)

Question 6d(iii).B: 1st-stage estimates

```r
first_c <- lm(diff_earnPOW ~ valchange_cr + as.factor(state)*as.factor(year) +
                msa + lnpop + diff_pop + frac_earn, data = q1)

first_nc <- lm(diff_earnPOW ~ valchange_cr + as.factor(state)*as.factor(year),
               data = q1)

#get estimates & normal SE (not robust)
get_regression_table(first_c)
```

```
## # A tibble: 101 x 7
##    term                estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                  <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 intercept              0.047     0.006     7.32    0        0.034    0.06
##  2 valchange_cr           0.022     0.002     9.30    0        0.018    0.027
##  3 as.factor(state)2     -0.008     0.009    -0.894   0.371   -0.026    0.01
##  4 as.factor(state)3     -0.018     0.01     -1.79    0.073   -0.037    0.002
##  5 as.factor(state)4      0.022     0.01      2.07    0.038    0.001    0.042
##  6 as.factor(year)1971    0.008     0.008     0.95    0.342   -0.009    0.025
##  7 as.factor(year)1972    0.07      0.008     8.28    0        0.053    0.087
##  8 as.factor(year)1973    0.058     0.008     6.91    0        0.042    0.075
##  9 as.factor(year)1974    0.065     0.008     7.73    0        0.048    0.081
## 10 as.factor(year)1975   -0.031     0.008    -3.67    0       -0.048   -0.014
## # ... with 91 more rows
```

```r
get_regression_table(first_nc)
```

```
## # A tibble: 97 x 7
##    term                estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                  <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 intercept              0.052     0.006      8.43   0        0.04     0.064
##  2 valchange_cr           0.024     0.002      9.96   0        0.02     0.029
##  3 as.factor(state)2     -0.011     0.009     -1.16   0.244   -0.029    0.007
##  4 as.factor(state)3     -0.021     0.01      -2.14   0.033   -0.041   -0.002
##  5 as.factor(state)4      0.012     0.011      1.09   0.277   -0.009    0.033
```

```
##   6 as.factor(year)1971    0.017     0.008      2.03  0.043    0.001    0.034
##   7 as.factor(year)1972    0.075     0.009      8.77  0        0.058    0.091
##   8 as.factor(year)1973    0.06      0.008      7.11  0        0.044    0.077
##   9 as.factor(year)1974    0.071     0.008      8.45  0        0.055    0.088
##  10 as.factor(year)1975   -0.026     0.008     -3.07  0.002   -0.043   -0.009
## # ... with 87 more rows
```

interpret the main estimates

Question 6d(iii).C: 2SLS structural eqn

```r
# helpful links:
#https://cran.r-project.org/web/packages/ivreg/ivreg.pdf
#https://www.econometrics-with-r.org/12-2-TGIVRM.html

iv_ssi_c <- ivreg(diff_dSSI ~
                    diff_earnPOW + as.factor(year)*as.factor(state) + msa +
                    lnpop + diff_pop + frac_earn
                  | valchange_cr + valchange_crlag1 + valchange_crlag2 +
                    as.factor(year)*as.factor(state) + msa +
                    lnpop + diff_pop + frac_earn, data = q1)


iv_ssi_nc <- ivreg(diff_dSSI ~ diff_earnPOW + as.factor(year)*as.factor(state)
                   | valchange_cr + valchange_crlag1 + valchange_crlag2 +
                    as.factor(year)*as.factor(state), data = q1)

iv_di_c <- ivreg(diff_pay ~ diff_earnPOW + as.factor(year)*as.factor(state) +
                   msa + lnpop + diff_pop + frac_earn
                 | valchange_cr + valchange_crlag1 + valchange_crlag2 +
                   as.factor(year)*as.factor(state) + msa +
                   lnpop + diff_pop + frac_earn, data = q1)

iv_di_nc <- ivreg(diff_pay ~ diff_earnPOW + as.factor(year)*as.factor(state) |
                    valchange_cr + valchange_crlag1 + valchange_crlag2 +
                    as.factor(year)*as.factor(state), data = q1)

# obtain robust SEs
coeftest(iv_ssi_c, vcov = vcovHC(iv_ssi_c, type = "HC1")) %>%
  tidy()
```

```
## # A tibble: 100 x 5
##    term                estimate std.error statistic  p.value
##    <chr>                  <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)           0.0513    0.0121     4.25  2.13e- 5
##  2 diff_earnPOW         -0.862     0.201     -4.29  1.84e- 5
##  3 as.factor(year)1971   0.0116    0.0147     0.793 4.28e- 1
##  4 as.factor(year)1972   0.0234    0.0212     1.11  2.68e- 1
##  5 as.factor(year)1973   0.129     0.0204     6.33  2.59e-10
##  6 as.factor(year)1974   0.645     0.0301    21.5   3.14e-99
##  7 as.factor(year)1975   0.0329    0.0133     2.48  1.33e- 2
```

```
##  8 as.factor(year)1976   0.0625    0.0210     2.98  2.92e- 3
##  9 as.factor(year)1977   0.0441    0.0209     2.11  3.49e- 2
## 10 as.factor(year)1978   0.0326    0.0176     1.85  6.43e- 2
## # ... with 90 more rows
```

```
coeftest(iv_di_c, vcov = vcovHC(iv_di_c, type = "HC1")) %>%
  tidy()
```

```
## # A tibble: 92 x 5
##    term                estimate std.error statistic  p.value
##    <chr>                  <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)           0.242     0.0145    16.7   2.38e-61
##  2 diff_earnPOW         -0.344     0.101     -3.42  6.33e- 4
##  3 as.factor(year)1971 -0.0569     0.0171    -3.33  8.84e- 4
##  4 as.factor(year)1972  0.101      0.0175     5.76  8.70e- 9
##  5 as.factor(year)1973 -0.0824     0.0169    -4.89  1.05e- 6
##  6 as.factor(year)1974  0.00990    0.0177     0.560 5.76e- 1
##  7 as.factor(year)1975 -0.0417     0.0152    -2.74  6.16e- 3
##  8 as.factor(year)1976 -0.0274     0.0176    -1.56  1.19e- 1
##  9 as.factor(year)1977 -0.0574     0.0174    -3.29  1.00e- 3
## 10 as.factor(year)1978 -0.101      0.0161    -6.29  3.33e-10
## # ... with 82 more rows
```

```
coeftest(iv_ssi_nc, vcov = vcovHC(iv_ssi_nc, type = "HC1")) %>%
  tidy()
```

```
## # A tibble: 96 x 5
##    term                estimate std.error statistic   p.value
##    <chr>                  <dbl>     <dbl>     <dbl>     <dbl>
##  1 (Intercept)           0.0668   0.00997      6.70 2.19e- 11
##  2 diff_earnPOW         -0.761    0.160       -4.75 2.06e-  6
##  3 as.factor(year)1971   0.0194   0.0137       1.42 1.56e-  1
##  4 as.factor(year)1972   0.0218   0.0183       1.19 2.33e-  1
##  5 as.factor(year)1973   0.126    0.0177       7.13 1.07e- 12
##  6 as.factor(year)1974   0.645    0.0287      22.5  1.07e-108
##  7 as.factor(year)1975   0.0405   0.0119       3.42 6.36e-  4
##  8 as.factor(year)1976   0.0623   0.0182       3.42 6.31e-  4
##  9 as.factor(year)1977   0.0444   0.0185       2.41 1.61e-  2
## 10 as.factor(year)1978   0.0311   0.0154       2.02 4.37e-  2
## # ... with 86 more rows
```

```
coeftest(iv_di_nc, vcov = vcovHC(iv_di_nc, type = "HC1")) %>%
  tidy()
```

```
## # A tibble: 88 x 5
##    term                estimate std.error statistic  p.value
##    <chr>                  <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)           0.235     0.0140    16.8   5.45e-62
##  2 diff_earnPOW         -0.327     0.0860    -3.80  1.48e- 4
##  3 as.factor(year)1971  -0.0523    0.0172    -3.04  2.39e- 3
##  4 as.factor(year)1972   0.106     0.0173     6.12  1.01e- 9
##  5 as.factor(year)1973  -0.0791    0.0165    -4.79  1.67e- 6
```

```
## 6 as.factor(year)1974  0.0158   0.0176    0.899 3.69e- 1
## 7 as.factor(year)1975 -0.0356   0.0151   -2.37  1.79e- 2
## 8 as.factor(year)1976 -0.0243   0.0172   -1.41  1.58e- 1
## 9 as.factor(year)1977 -0.0547   0.0171   -3.19  1.42e- 3
## 10 as.factor(year)1978 -0.0980  0.0159   -6.17  7.21e-10
## # ... with 78 more rows
```

Question 6d(iii).D: Summary table

As discussed with Gin in office hours, I'm writing my robust standard errors for the main estimates in another column, rather than including them in parentheses in the same cell as the estimates.

```
sum_table <- function(reg_model, reg_name) {
  coeftest(reg_model, vcov = vcovHC(reg_model, type = "HC1")) %>%
    tidy()  %>%
    mutate(reg = reg_name) %>%
    filter(term == "diff_earnPOW") %>%
    select(reg, estimate, std.error, p.value) %>%
    mutate(
      sig_1perc = if_else(p.value < 0.01, "yes", "no"),
      sig_5perc = if_else(p.value < 0.05, "yes", "no"),
      sig_10perc = if_else(p.value < 0.1, "yes", "no"))
}

q6_sum <- sum_table(ols_di_nc, "OLS Disability - No Control") %>%
  bind_rows(sum_table(ols_di_c, "OLS Disability - Control")) %>%
  bind_rows(sum_table(ols_ssi_nc, "OLS SSI - No Control")) %>%
  bind_rows(sum_table(ols_ssi_c, "OLS SSI - Control")) %>%
  bind_rows(sum_table(iv_di_nc, "IV Disability - No Control")) %>%
  bind_rows(sum_table(iv_di_c, "IV Disability - Control")) %>%
  bind_rows(sum_table(iv_ssi_nc, "IV SSI - No Control")) %>%
  bind_rows(sum_table(iv_ssi_c, "IV SSI - Control"))

q6_sum %>%
  knitr::kable(
    caption = "Summary Table of OLS and IV Regressions",
    col.names = c("Regression type",
                  "Main estimate (log diff in county earnings)",
                  "Robust standard error",
                  "P-value",
                  "Sig @ 1%?",
                  "Sig @ 5%?",
                  "Sig @ 10%?"),
    digits = 5,
    linesep = "\\addlinespace") %>%
  kableExtra::column_spec(2:4, width = "2cm") %>%
  kableExtra::column_spec(5:7, width = "1cm")
```

Table 1: Summary Table of OLS and IV Regressions

Regression type	Main estimate (log diff in county earnings)	Robust standard error	P-value	Sig @ 1%?	Sig @ 5%?	Sig @ 10%?
OLS Disability - No Control	0.00159	0.01660	0.92383	no	no	no
OLS Disability - Control	-0.00216	0.01726	0.90049	no	no	no
OLS SSI - No Control	-0.01996	0.01325	0.13192	no	no	no
OLS SSI - Control	-0.02640	0.01360	0.05232	no	no	yes
IV Disability - No Control	-0.32663	0.08604	0.00015	yes	yes	yes
IV Disability - Control	-0.34432	0.10071	0.00063	yes	yes	yes
IV SSI - No Control	-0.76143	0.16028	0.00000	yes	yes	yes
IV SSI - Control	-0.86248	0.20120	0.00002	yes	yes	yes