Exercise_ADH

2022-09-05

Loading data, packages

```
library(readr)
library(tibble)
library(fixest)
library(stringr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(sandwich)
```

```
df = read_csv("adh_data.csv")
```

```
## Rows: 1444 Columns: 208
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (1): city
## dbl (207): czone, statefip, yr, t2, timepwt48, reg_midatl, reg_encen, reg_wn...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(df)
```

```
## # A tibble: 6 x 208
##   czone statefip city         yr    t2 timepwt48 reg_m~1 reg_e~2 reg_w~3 reg_s~4
##   <dbl>    <dbl> <chr>     <dbl> <dbl>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1   100       47 undefined  1990     0  0.00211        0       0       0       0
## 2   100       47 undefined  2000     1  0.00207        0       0       0       0
```

```
## 3   200       47 undefined  1990     0  0.000732       0       0       0       0
## 4   200       47 undefined  2000     1  0.000815       0       0       0       0
## 5   301       47 undefined  1990     0  0.000261       0       0       0       0
## 6   301       47 undefined  2000     1  0.000240       0       0       0       0
## # ... with 198 more variables: reg_escen <dbl>, reg_wscen <dbl>,
## #   reg_mount <dbl>, reg_pacif <dbl>, l_popcount <dbl>,
## #   l_no_workers_totcbp <dbl>, l_shind_manuf_cbp <dbl>, l_sh_popedu_c <dbl>,
## #   l_sh_popfborn <dbl>, l_sh_empl_f <dbl>, l_sh_routine33 <dbl>,
## #   l_task_outsource <dbl>, l_sh_empl <dbl>, l_sh_empl_mfg <dbl>,
## #   l_sh_empl_mfg_edu_c <dbl>, l_sh_empl_mfg_edu_nc <dbl>,
## #   l_sh_empl_mfg_age1634 <dbl>, l_sh_empl_mfg_age3549 <dbl>, ...
```

OLS regression

The core of the paper is looking at what happened to laborer's when theres an increase in us imports from china. Let's try and replicate part of Table 9 - namely the estimate from panel A column 2.

Their y variable is `relchg_avg_hhincwage_pc_pw`.

The important x variable is decadal trade between the us and china `d_tradeusch_pw`.

1. Run that simple regression

```
ols_1 = lm(relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw, data = df)
summary(ols_1)
```

```
##
## Call:
## lm(formula = relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.789  -8.411  -0.663   7.715  49.684
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    16.0720     0.3889   41.33   <2e-16 ***
## d_tradeusch_pw -1.6466     0.1212  -13.59   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.89 on 1442 degrees of freedom
## Multiple R-squared:  0.1135, Adjusted R-squared:  0.1129
## F-statistic: 184.6 on 1 and 1442 DF,  p-value: < 2.2e-16
```

2. Now add heteroskedasticity robust standard (HC1). Hint: Use the `sandwich` and `lmtest` packages

```
# ols_1 = lm(relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw, data = df)
lmtest::coeftest(ols_1, sandwich::vcovHC(ols_1, "HC1"))
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    16.07198    0.57211 28.0923 < 2.2e-16 ***
## d_tradeusch_pw -1.64663    0.28496 -5.7785 9.219e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ols_rob = feols(relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw, data = df, se = "HC1")
summary(ols_rob)
```

```
## OLS estimation, Dep. Var.: relchg_avg_hhincwage_pc_pw
## Observations: 1,444
## Standard-errors: Heteroskedasticity-robust
##                Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)    16.07198   0.572114 28.09229  < 2.2e-16 ***
## d_tradeusch_pw -1.64663   0.284956 -5.77854 9.2187e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 11.9   Adj. R2: 0.112885
```

Now we will start to add extra x variables.

3. Start by adding `t2` - a dummy variable for whether observation is in the second decade. Fit again with HC1 robust standard errors.

```
ols_2 = feols(relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw + t2,
              data = df, vcov = "HC1")
summary(ols_2)
```

```
## OLS estimation, Dep. Var.: relchg_avg_hhincwage_pc_pw
## Observations: 1,444
## Standard-errors: Heteroskedasticity-robust
##                  Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)     21.590958   0.398893  54.12714  < 2.2e-16 ***
## d_tradeusch_pw  -0.883161   0.168044  -5.25555 1.6977e-07 ***
## t2             -13.947688   0.578687 -24.10229  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 9.82092   Adj. R2: 0.393606
```

```
ols_2_hc3 = feols(relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw + t2,
                  data = df, vcov = function (x) sandwich::vcovHC(x, "HC3"))
summary(ols_2_hc3)
```

```
## OLS estimation, Dep. Var.: relchg_avg_hhincwage_pc_pw
## Observations: 1,444
## Standard-errors: vcovHC(x, "HC3")
##                  Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)     21.590958   0.410042  52.65551  < 2.2e-16 ***
## d_tradeusch_pw  -0.883161   0.187696  -4.70526 2.7788e-06 ***
## t2             -13.947688   0.588642 -23.69469  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 9.82092   Adj. R2: 0.393606
```

Clustering

Let us now use clustertered standard errors instead. ADH cluster by `statefip`.

1. Run the basic regression with clustering

```
ols_clust = feols(relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw,
                  data = df, cluster = c("statefip"))
summary(ols_clust)
```

```
## OLS estimation, Dep. Var.: relchg_avg_hhincwage_pc_pw
## Observations: 1,444
## Standard-errors: Clustered (statefip)
##                Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)    16.07198   0.957363 16.78776  < 2.2e-16 ***
## d_tradeusch_pw -1.64663   0.370000 -4.45034 5.2454e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 11.9   Adj. R2: 0.112885
```

2. Add the following controls to your last regression:
    - l_shind_manuf_cbp
    - l_sh_popedu_c
    - l_sh_popfborn
    - l_sh_empl_f
    - l_sh_routine33
    - l_task_outsource

```
ols_clust = feols(relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw + l_shind_manuf_cbp +
                    l_sh_popedu_c + l_sh_popfborn + l_sh_popfborn + l_sh_empl_f +
                    l_sh_routine33 + l_task_outsource,
                  data = df, cluster = c("statefip"))
summary(ols_clust)
```

```
## OLS estimation, Dep. Var.: relchg_avg_hhincwage_pc_pw
## Observations: 1,444
## Standard-errors: Clustered (statefip)
##                     Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)       98.430164  11.191489  8.795091 1.7032e-11 ***
## d_tradeusch_pw    -1.063005   0.287892 -3.692369 5.7754e-04 ***
## l_shind_manuf_cbp -0.172152   0.048286 -3.565261 8.4772e-04 ***
## l_sh_popedu_c     -0.320964   0.068543 -4.682686 2.4379e-05 ***
## l_sh_popfborn     -0.524305   0.094405 -5.553797 1.2659e-06 ***
## l_sh_empl_f        0.067594   0.090310  0.748463 4.5791e-01
## l_sh_routine33    -2.041397   0.287715 -7.095208 5.8625e-09 ***
## l_task_outsource  17.046616   2.083908  8.180119 1.3734e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.4   Adj. R2: 0.317146
```

3. Add region fixed effects to your regression.

    - First find all variables in the dataset that start with `reg_`
    - Add these to your last regression

```
# reg_vars <- df %>% select(starts_with("reg_")) %>% names()
reg_vars <- df %>% colnames() %>% str_subset("reg_")

fe_str = str_c(reg_vars, collapse = " + ")
form_str = str_c(
    "relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw + l_shind_manuf_cbp",
    "l_sh_popedu_c + l_sh_popfborn + l_sh_popfborn + l_sh_empl_f",
    "l_sh_routine33 + l_task_outsource",
    fe_str,
    sep = " + "
)
form <- as.formula(form_str)
```

```
fe_model = feols(form, data = df, cluster = c("statefip"))
summary(fe_model)
```

```
## OLS estimation, Dep. Var.: relchg_avg_hhincwage_pc_pw
## Observations: 1,444
## Standard-errors: Clustered (statefip)
##                     Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)       103.331722  13.422215  7.698560 7.2020e-10 ***
## d_tradeusch_pw     -0.960540   0.288905 -3.324761 1.7221e-03 **
## l_shind_manuf_cbp  -0.139771   0.045212 -3.091447 3.3447e-03 **
## l_sh_popedu_c      -0.297031   0.072550 -4.094181 1.6543e-04 ***
## l_sh_popfborn      -0.548079   0.093430 -5.866216 4.2879e-07 ***
## l_sh_empl_f        -0.229563   0.122405 -1.875429 6.6953e-02 .
## l_sh_routine33     -1.634096   0.319260 -5.118383 5.6310e-06 ***
## l_task_outsource   17.653643   2.200501  8.022556 2.3573e-10 ***
## reg_midatl         -1.139319   2.335920 -0.487739 6.2800e-01
## reg_encen          -0.511749   2.735801 -0.187056 8.5242e-01
## reg_wncen           5.788149   2.668650  2.168943 3.5180e-02 *
## reg_satl           -4.025975   2.784034 -1.446094 1.5479e-01
## reg_escen          -3.523777   2.846048 -1.238130 2.2182e-01
## reg_wscen           1.113610   2.460586  0.452579 6.5293e-01
## reg_mount           1.913916   2.617786  0.731120 4.6834e-01
## reg_pacif          -1.892620   2.649640 -0.714293 4.7858e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.1   Adj. R2: 0.358537
```

```
df = df %>% mutate(region = case_when(
  reg_midatl == 1 ~ 'midatl',
  reg_encen  == 1 ~ 'encen',
  reg_wncen  == 1 ~ 'wncen',
  reg_satl   == 1 ~ 'satl',
  reg_escen  == 1 ~ 'escen',
  reg_wscen  == 1 ~ 'wscen',
  reg_mount  == 1 ~ 'mount',
  reg_pacif  == 1 ~ 'pacif',
  TRUE ~ NA_character_
))

fe_model <- feols(
    relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw + l_shind_manuf_cbp +
        l_sh_popedu_c + l_sh_popfborn + l_sh_popfborn + l_sh_empl_f +
        l_sh_routine33 + l_task_outsource + i(region),
    data = df, cluster = c("statefip")
)
```

```
## NOTE: 32 observations removed because of NA values (RHS: 32).
```

```
summary(fe_model)
```

```
## OLS estimation, Dep. Var.: relchg_avg_hhincwage_pc_pw
## Observations: 1,412
## Standard-errors: Clustered (statefip)
##                     Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)       102.733914  13.411295  7.660253 1.9580e-09 ***
```

```
## d_tradeusch_pw     -0.949723   0.288561 -3.291238 2.0571e-03 **
## l_shind_manuf_cbp  -0.139459   0.045500 -3.065072 3.8397e-03 **
## l_sh_popedu_c      -0.305625   0.073205 -4.174925 1.5148e-04 ***
## l_sh_popfborn      -0.548922   0.094721 -5.795139 8.4614e-07 ***
## l_sh_empl_f        -0.219595   0.123019 -1.785055 8.1652e-02 .
## l_sh_routine33     -1.641218   0.323082 -5.079886 8.6710e-06 ***
## l_task_outsource   17.618681   2.217757  7.944369 7.9153e-10 ***
## region::escen      -3.026390   2.013417 -1.503111 1.4047e-01
## region::midatl     -0.605299   1.579094 -0.383320 7.0346e-01
## region::mount       2.493887   2.296060  1.086159 2.8375e-01
## region::pacif      -1.276165   2.209543 -0.577570 5.6671e-01
## region::satl       -3.511611   2.035516 -1.725170 9.2028e-02 .
## region::wncen       6.270456   2.407715  2.604318 1.2764e-02 *
## region::wscen       1.649646   1.668402  0.988758 3.2858e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.1   Adj. R2: 0.358638
```

```
fe_model <- feols(
    relchg_avg_hhincwage_pc_pw ~ d_tradeusch_pw + l_shind_manuf_cbp +
        l_sh_popedu_c + l_sh_popfborn + l_sh_popfborn + l_sh_empl_f +
        l_sh_routine33 + l_task_outsource | region,
    data = df, cluster = c("statefip")
)
```

```
## NOTE: 32 observations removed because of NA values (Fixed-effects: 32).
```

```
summary(fe_model)
```

```
## OLS estimation, Dep. Var.: relchg_avg_hhincwage_pc_pw
## Observations: 1,412
## Fixed-effects: region: 8
## Standard-errors: Clustered (statefip)
##                    Estimate Std. Error  t value   Pr(>|t|)
## d_tradeusch_pw    -0.949723   0.288561 -3.29124 2.0571e-03 **
## l_shind_manuf_cbp -0.139459   0.045500 -3.06507 3.8397e-03 **
## l_sh_popedu_c     -0.305625   0.073205 -4.17493 1.5148e-04 ***
## l_sh_popfborn     -0.548922   0.094721 -5.79514 8.4614e-07 ***
## l_sh_empl_f       -0.219595   0.123019 -1.78505 8.1652e-02 .
## l_sh_routine33    -1.641218   0.323082 -5.07989 8.6710e-06 ***
## l_task_outsource  17.618681   2.217757  7.94437 7.9153e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.1     Adj. R2: 0.358638
##              Within R2: 0.303106
```

Instrument Variables

1. Instrument `d_tradeusch_pw` with `d_tradeotch_pw_lag` in your last regression

```
iv_reg = feols(
    relchg_avg_hhincwage_pc_pw ~ l_shind_manuf_cbp +
        l_sh_popedu_c + l_sh_popfborn + l_sh_popfborn + l_sh_empl_f +
        l_sh_routine33 + l_task_outsource | 0 | d_tradeusch_pw ~ d_tradeotch_pw_lag,
    data = df, cluster = c("statefip")
)
```

```
summary(iv_reg)
```

```
## TSLS estimation, Dep. Var.: relchg_avg_hhincwage_pc_pw, Endo.: d_tradeusch_pw, Instr.: d_tradeotch_pw
## Second stage: Dep. Var.: relchg_avg_hhincwage_pc_pw
## Observations: 1,444
## Standard-errors: Clustered (statefip)
##                     Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)        84.382112  11.863723  7.112616 5.5171e-09 ***
## fit_d_tradeusch_pw -2.602937   0.714528 -3.642878 6.7110e-04 ***
## l_shind_manuf_cbp  -0.024517   0.065685 -0.373258 7.1063e-01
## l_sh_popedu_c      -0.278367   0.072510 -3.839039 3.6821e-04 ***
## l_sh_popfborn      -0.468990   0.100407 -4.670873 2.5355e-05 ***
## l_sh_empl_f         0.091119   0.099534  0.915459 3.6462e-01
## l_sh_routine33     -1.725561   0.296334 -5.823030 4.9822e-07 ***
## l_task_outsource   14.615441   2.030773  7.196985 4.1110e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.1   Adj. R2: 0.239452
## F-test (1st stage), d_tradeusch_pw: stat = 907.0, p < 2.2e-16, on 1 and 1,436 DoF.
##                         Wu-Hausman: stat = 111.1, p < 2.2e-16, on 1 and 1,435 DoF.
```

2. Weight your regression by `timepwt48`

```
iv_reg_2 = feols(
    relchg_avg_hhincwage_pc_pw ~ l_shind_manuf_cbp +
        l_sh_popedu_c + l_sh_popfborn + l_sh_popfborn + l_sh_empl_f +
        l_sh_routine33 + l_task_outsource | 0 | d_tradeusch_pw ~ d_tradeotch_pw_lag,
    data = df, cluster = c("statefip"),
    weights = ~timepwt48
)
summary(iv_reg_2)
```

```
## TSLS estimation, Dep. Var.: relchg_avg_hhincwage_pc_pw, Endo.: d_tradeusch_pw, Instr.: d_tradeotch_pw
## Second stage: Dep. Var.: relchg_avg_hhincwage_pc_pw
## Observations: 1,444
## Standard-errors: Clustered (statefip)
##                     Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)        73.961602  11.854845  6.238934 1.1694e-07 ***
## fit_d_tradeusch_pw -3.785835   0.570072 -6.640973 2.8672e-08 ***
## l_shind_manuf_cbp   0.169233   0.069850  2.422798 1.9308e-02 *
## l_sh_popedu_c      -0.186881   0.131546 -1.420653 1.6202e-01
## l_sh_popfborn      -0.080177   0.086035 -0.931921 3.5614e-01
## l_sh_empl_f        -0.142948   0.138788 -1.029976 3.0829e-01
## l_sh_routine33     -1.217598   0.352090 -3.458202 1.1655e-03 **
## l_task_outsource   11.234254   1.944024  5.778867 5.8079e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.330337   Adj. R2: 0.185213
## F-test (1st stage), d_tradeusch_pw: stat = 1,117.1, p < 2.2e-16, on 1 and 1,436 DoF.
##                         Wu-Hausman: stat =   154.6, p < 2.2e-16, on 1 and 1,435 DoF.
```

```
iv_reg_2$iv_first_stage
```

```
## $d_tradeusch_pw
## TSLS estimation, Dep. Var.: d_tradeusch_pw, Endo.: d_tradeusch_pw, Instr.: d_tradeotch_pw_lag
## First stage: Dep. Var.: d_tradeusch_pw
```

```
## Observations: 1,444
## Standard-errors: Clustered (statefip)
##                      Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)        -1.832015   1.268263 -1.444508 1.5523e-01
## d_tradeotch_pw_lag  0.746107   0.072278 10.322808 1.1372e-13 ***
## l_shind_manuf_cbp   0.044772   0.008211  5.452934 1.7923e-06 ***
## l_sh_popedu_c       0.022929   0.008207  2.793875 7.5132e-03 **
## l_sh_popfborn       0.014312   0.006246  2.291338 2.6472e-02 *
## l_sh_empl_f         0.005347   0.010529  0.507884 6.1391e-01
## l_sh_routine33     -0.003746   0.041013 -0.091337 9.2761e-01
## l_task_outsource   -0.241380   0.276901 -0.871719 3.8779e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.043006   Adj. R2: 0.562323
## F-test (1st stage): stat = 1,117.1, p < 2.2e-16, on 1 and 1,436 DoF.
```