Replication study of DiD from Minimum Wages and Employment

Replication study of 'Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania,' American Economic Review (1994„ vol. 84: 772-793)

Preliminaries

```
library(tidyverse)
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.3
## v tibble  3.0.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(haven)
library(estimatr)
minwage = read.csv('minwage.csv')
minwage = minwage %>%
  filter(sample == 1) %>%
  rename(treat = state)%>%
mutate(state = case_when(
  treat == 0 ~ 'PA',
  treat == 1 ~ 'NJ'
),
low_wage = 1*(wage_st<5)
)
summary(minwage)
```

```
##      sheet           chain          co_owned          treat
##  Min.   :  1.0   Min.   :1.000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:118.5   1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:1.000
##  Median :234.0   Median :2.000   Median :0.0000   Median :1.000
##  Mean   :246.9   Mean   :2.091   Mean   :0.3533   Mean   :0.812
##  3rd Qu.:377.5   3rd Qu.:3.000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :522.0   Max.   :4.000   Max.   :1.0000   Max.   :1.000
```

```
##      empft           emppt          wage_st          empft2
##  Min.   : 0.0   Min.   : 0.00   Min.   :4.250   Min.   : 0.000
##  1st Qu.: 2.0   1st Qu.:11.00   1st Qu.:4.250   1st Qu.: 2.000
##  Median : 6.0   Median :17.00   Median :4.500   Median : 6.000
##  Mean   : 8.4   Mean   :18.82   Mean   :4.621   Mean   : 8.355
##  3rd Qu.:12.0   3rd Qu.:25.00   3rd Qu.:5.000   3rd Qu.:12.000
##  Max.   :60.0   Max.   :60.00   Max.   :5.750   Max.   :40.000
##      emppt2         wage_st2          fte             fte2
##  Min.   : 0.00   Min.   :4.250   Min.   : 3.00   Min.   : 3.50
##  1st Qu.:11.00   1st Qu.:5.050   1st Qu.:11.88   1st Qu.:11.50
##  Median :17.00   Median :5.050   Median :16.50   Median :17.00
##  Mean   :18.62   Mean   :4.995   Mean   :17.81   Mean   :17.66
##  3rd Qu.:25.00   3rd Qu.:5.050   3rd Qu.:21.00   3rd Qu.:22.50
##  Max.   :60.00   Max.   :6.250   Max.   :80.00   Max.   :55.50
##       dfte              gap               dw              sample
##  Min.   :-43.500   Min.   :0.00000   Min.   :-0.7500   Min.   :1
##  1st Qu.: -3.625   1st Qu.:0.00000   1st Qu.: 0.0500   1st Qu.:1
##  Median :  0.000   Median :0.06316   Median : 0.4300   Median :1
##  Mean   : -0.146   Mean   :0.08490   Mean   : 0.3744   Mean   :1
##  3rd Qu.:  4.125   3rd Qu.:0.18824   3rd Qu.: 0.8000   3rd Qu.:1
##  Max.   : 26.000   Max.   :0.18824   Max.   : 2.0000   Max.   :1
##     state              low_wage
##  Length:351         Min.   :0.0000
##  Class :character   1st Qu.:0.0000
##  Mode  :character   Median :1.0000
##                     Mean   :0.7464
##                     3rd Qu.:1.0000
##                     Max.   :1.0000
```

Baseline difference in differences estimate: starting wages

```
DinD_wage = minwage %>%
  group_by(state) %>%
  summarize(mean_wage_st = mean(wage_st),
            mean_wage_st2 = mean(wage_st2))%>%
              mutate(diff = mean_wage_st2 - mean_wage_st)
DinD_wage
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 4
##   state mean_wage_st mean_wage_st2    diff
##   <chr>        <dbl>         <dbl>   <dbl>
## 1 NJ            4.61          5.08  0.469
## 2 PA            4.65          4.62 -0.0348
```

PA's within-state time-difference = -0.034848 NJ's within-state time-difference = 0.469158 DiD = 0.469

–(-0.0348) = 0.504 It assumes parallel trend between NJ and PA. That means every factor that affects the wage have a parallel trend besides the minimum wage law.

Baseline DID Estimate: full time equivalent employment.

```
DinD_employment = minwage %>%
  group_by(state) %>%
  summarize(mean_fte = mean(fte),
            mean_fte2 = mean(fte2))%>%
  mutate(diff = mean_fte2 - mean_fte)
DinD_employment
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 4
##   state mean_fte mean_fte2   diff
##   <chr>    <dbl>     <dbl>  <dbl>
## 1 NJ        17.3      17.6  0.287
## 2 PA        20.1      18.1 -2.02
```

PA's within-state time-difference = -2.01515 NJ's within-state time-difference = 0.286842 Difference in differences = 0.287-(-2.02) = 2.89 It assumes parallel trend between NJ and PA. That means every factor that affects the wage have a parallel trend besides the minimum wage law.

Reshape minwage for DID regression estimation:

```
wave1 = minwage %>%
  select(state, treat, wage_st, fte, chain, co_owned, low_wage) %>%
  mutate(post = 0)
wave2 = minwage %>%
  select(state, treat, wage_st2, fte2, chain, co_owned, low_wage) %>%
  mutate(post = 1) %>%
  rename(wage_st = wage_st2, fte = fte2)
both_waves = bind_rows(wave1,wave2)
```

DID Regression Estimates:

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```r
reg_wage1 = lm(wage_st ~ treat + post + treat:post, both_waves)
reg_emp1 = lm(fte ~ treat +post + treat:post, both_waves)
stargazer(reg_wage1, reg_emp1, type = 'text', digits =2,
          dep.var.labels =c('Staring wage','Full-time Equiv. Employment')
          )
```

```
##
## ========================================================================
##                                          Dependent variable:
##                                ----------------------------------------
##                                Staring wage Full-time Equiv. Employment
##                                    (1)                  (2)
## ------------------------------------------------------------------------
## treat                             -0.04               -2.84**
##                                   (0.04)               (1.22)
##
## post                              -0.03                -2.02
##                                   (0.05)               (1.56)
##
## treat:post                       0.50***               2.30
##                                   (0.05)               (1.73)
##
## Constant                         4.65***             20.11***
##                                   (0.03)               (1.10)
##
## ------------------------------------------------------------------------
## Observations                       702                  702
## R2                                0.40                 0.01
## Adjusted R2                       0.40                 0.004
## Residual Std. Error (df = 698)    0.28                 8.97
## F Statistic (df = 3; 698)       156.91***              1.87
## ========================================================================
## Note:                                       *p<0.1; **p<0.05; ***p<0.01
```

DiD for wage = 0.5. DiD for employment is 2.3. The treatment is not statistically significant and not close to the number in the difference in differences table above. There is no robust evidence that the employment would go down (the classic demand and supply model) or up (monopsonist model).

```r
reg_wage2 = lm(wage_st ~ treat + post + treat:post + co_owned + as.factor(chain), both_waves)
reg_emp2 = lm(fte ~ treat + post + treat:post + co_owned + as.factor(chain), both_waves)
stargazer(reg_wage2, reg_emp2, type = 'text', digits =2,
          dep.var.labels = c('Starting Wage', 'Full-time Equiv. Employment')
          )
```

```
##
## =========================================================================
##                                           Dependent variable:
##                                -----------------------------------------
##                                Starting Wage Full-time Equiv. Employment
##                                     (1)                  (2)
## -------------------------------------------------------------------------
## treat                              -0.04               -2.14*
```

```
##                                 (0.04)              (1.10)
##
## post                             -0.03               -2.02
##                                 (0.05)              (1.40)
##
## co_owned                        0.07***             -1.01
##                                 (0.02)              (0.72)
##
## as.factor(chain)2                0.02              -10.16***
##                                 (0.03)              (0.84)
##
## as.factor(chain)3                0.05               -1.35
##                                 (0.03)              (0.86)
##
## as.factor(chain)4               0.12***             -1.37
##                                 (0.03)              (0.97)
##
## treat:post                      0.50***              2.30
##                                 (0.05)              (1.55)
##
## Constant                        4.59***            22.56***
##                                 (0.04)              (1.05)
##
## ---------------------------------------------------------------------
## Observations                      702                 702
## R2                               0.43                0.21
## Adjusted R2                      0.42                0.20
## Residual Std. Error (df = 694)   0.27                8.03
## F Statistic (df = 7; 694)       73.61***            26.22***
## =====================================================================
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

The conclusion for wage is the same as the sample variance and coefficient is the same. For employment, the DID coefficient is still the same. But the sample variance goes down. There is no effect on employment.

Probing the DID Assumption:

```r
nj_only = both_waves %>%
  filter(state == 'NJ')
pa_only = both_waves %>%
  filter(state == 'PA')
nj_wage1 = lm(wage_st ~ low_wage + post + low_wage:post, nj_only)
nj_emp1 = lm(fte ~ low_wage + post + low_wage:post, nj_only)
stargazer(nj_wage1, nj_emp1, type = 'text', digits = 2,
          dep.var.labels = c('Staring Wage', 'Full-time Equiv. Employment')
          )
```

```
##
## ====================================================================
##                                        Dependent variable:
##                                ---------------------------------------
##                                Staring Wage Full-time Equiv. Employment
##                                    (1)                 (2)
## --------------------------------------------------------------------
```

```
## low_wage                          -0.65***             -2.23*
##                                    (0.02)              (1.21)
##
## post                               -0.004               -2.25
##                                    (0.03)              (1.50)
##
## low_wage:post                      0.62***              3.30*
##                                    (0.03)              (1.71)
##
## Constant                           5.11***             18.99***
##                                    (0.02)              (1.06)
##
## ------------------------------------------------------------------------
## Observations                         570                  570
## R2                                   0.78                0.01
## Adjusted R2                          0.77                0.002
## Residual Std. Error (df = 566)       0.16                8.62
## F Statistic (df = 3; 566)         653.18***              1.44
## ========================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

Control:

```
nj_wage2 = lm(wage_st ~ low_wage + post + low_wage:post + co_owned + as.factor(chain), nj_only)
nj_emp2 = lm(fte ~ low_wage + post + low_wage:post + co_owned + as.factor(chain), nj_only)
stargazer(nj_wage2, nj_emp2, type = 'text', digits = 2,
          dep.var.labels = c('Staring Wage', 'Full-time Equiv. Employment')
          )
```

```
##
## ========================================================================
##                                          Dependent variable:
##                                    -------------------------------------
##                                    Staring Wage Full-time Equiv. Employment
##                                        (1)                  (2)
## ------------------------------------------------------------------------
## low_wage                             -0.64***             -2.44**
##                                       (0.02)              (1.09)
##
## post                                  -0.004              -2.25*
##                                       (0.03)              (1.34)
##
## co_owned                               0.005               -0.74
##                                       (0.02)              (0.76)
##
## as.factor(chain)2                      0.004              -9.37***
##                                       (0.02)              (0.87)
##
## as.factor(chain)3                     0.04**               0.01
##                                       (0.02)              (0.90)
##
## as.factor(chain)4                     0.08***              -1.04
##                                       (0.02)              (1.08)
```

```
##
## low_wage:post                     0.62***            3.30**
##                                    (0.03)             (1.53)
##
## Constant                          5.08***            21.64***
##                                    (0.02)             (1.07)
##
## ----------------------------------------------------------------------------
## Observations                        570                570
## R2                                 0.78               0.21
## Adjusted R2                        0.78               0.20
## Residual Std. Error (df = 562)     0.16               7.72
## F Statistic (df = 7; 562)       288.51***           21.53***
## ============================================================================
## Note:                                          *p<0.1; **p<0.05; ***p<0.01
```

Parallel trend occurs in low-wage and high-wage restaurants in NJ. After the minimum wage law is passed, workers would go to restaurants with better working condition but offering same wage.

```
pa_wage1 = lm(wage_st ~ low_wage + post + low_wage:post, pa_only)
pa_emp1 = lm(fte~ low_wage + post + low_wage:post, pa_only)
stargazer(pa_wage1, pa_emp1, type = 'text', digits = 2,dep.var.labels = c('Starring Wage', 'Full-time Eq
          )
```

```
##
## =============================================================================
##                                              Dependent variable:
##                                 ---------------------------------------------
##                                 Starring Wage Full-time Equiv. Employment
##                                      (1)                  (2)
## -----------------------------------------------------------------------------
## low_wage                          -0.63***              -0.89
##                                    (0.07)               (2.67)
##
## post                              -0.27***              -3.85
##                                    (0.08)               (3.04)
##
## low_wage:post                      0.35***               2.81
##                                    (0.10)               (3.77)
##
## Constant                           5.07***              20.70***
##                                    (0.06)               (2.15)
##
## -----------------------------------------------------------------------------
## Observations                        132                  132
## R2                                 0.43                 0.01
## Adjusted R2                        0.41                -0.01
## Residual Std. Error (df = 128)     0.28                 10.32
## F Statistic (df = 3; 128)       31.61***                0.63
## =============================================================================
## Note:                                            *p<0.1; **p<0.05; ***p<0.01
```

Control:

```r
pa_wage2 = lm(wage_st ~ low_wage + post + low_wage:post + co_owned + as.factor(chain), pa_only)
pa_emp2 = lm(fte ~ low_wage + post + low_wage:post + co_owned + as.factor(chain), pa_only)
stargazer(pa_wage2, pa_emp2, type = 'text', digits = 2,
          dep.var.labels = c('Staring Wage', 'Full-time Equiv. Employment')
          )
```

```
##
## ==========================================================================
##                                           Dependent variable:
##                                 -----------------------------------------
##                                 Staring Wage Full-time Equiv. Employment
##                                     (1)                (2)
## --------------------------------------------------------------------------
## low_wage                          -0.60***           -4.77**
##                                    (0.07)             (2.38)
##
## post                              -0.27***            -3.85
##                                    (0.08)             (2.57)
##
## co_owned                          0.24***             -1.73
##                                    (0.07)             (2.27)
##
## as.factor(chain)2                  -0.11            -14.42***
##                                    (0.08)             (2.73)
##
## as.factor(chain)3                 -0.18**            -9.12***
##                                    (0.08)             (2.57)
##
## as.factor(chain)4                  -0.11              -3.15
##                                    (0.08)             (2.49)
##
## low_wage:post                     0.35***              2.81
##                                    (0.10)             (3.18)
##
## Constant                          5.03***            28.63***
##                                    (0.07)             (2.22)
##
## --------------------------------------------------------------------------
## Observations                        132                132
## R2                                  0.48               0.32
## Adjusted R2                         0.45               0.28
## Residual Std. Error (df = 124)      0.27               8.71
## F Statistic (df = 7; 124)         16.20***            8.33***
## ==========================================================================
## Note:                                         *p<0.1; **p<0.05; ***p<0.01
```

DID assumption may not hold in this case. After the minimum wage law, restaurant with low-wage in PA raises their wages. Low wage restaurants are affected by the minimum wage law in their decision as the overall wage in PA decreases.

——————————Case2: Minimum Legal Drinking Age (MLDA) revisit———————- Estimate the effect of legal on mrate including state andyear effects:

```
library(tidyverse)
library(haven)
library(estimatr)
mlda = read.csv('deaths.csv')
mlda = mlda %>%
  filter(year<=1983, agegr== '18-20 yrs' , dtype=='all')%>%
  mutate(year_factor=factor(year), state=factor(state))
summary(mlda)
```

```
##       year          state       legal1820                 dtype
##  Min.   :1970   1      : 14   Min.   :0.0000   all           :714
##  1st Qu.:1973   2      : 14   1st Qu.:0.0000   homicide      :  0
##  Median :1976   4      : 14   Median :0.6777   internal      :  0
##  Mean   :1976   5      : 14   Mean   :0.5571   MVA           :  0
##  3rd Qu.:1980   6      : 14   3rd Qu.:1.0000   other external:  0
##  Max.   :1983   8      : 14   Max.   :1.0000   suicide       :  0
##                 (Other):630
##        agegr         count             pop               age
##  15-17 yrs:  0   Min.   :  23.0   Min.   :  17317   Min.   :18.92
##  18-20 yrs:714   1st Qu.:  83.0   1st Qu.:  58872   1st Qu.:18.99
##  21-24 yrs:  0   Median : 221.0   Median : 168903   Median :19.00
##                  Mean   : 309.6   Mean   : 243764   Mean   :19.00
##                  3rd Qu.: 371.8   3rd Qu.: 310276   3rd Qu.:19.02
##                  Max.   :1893.0   Max.   :1368730   Max.   :19.15
##
##      legal           beertaxa         beerpercap      winepercap
##  Min.   :0.0000   Min.   :0.02008   Min.   :0.600   Min.   :0.0600
##  1st Qu.:0.0000   1st Qu.:0.11503   1st Qu.:1.130   1st Qu.:0.1600
##  Median :0.6777   Median :0.18923   Median :1.300   Median :0.2500
##  Mean   :0.5571   Mean   :0.28695   Mean   :1.311   Mean   :0.2983
##  3rd Qu.:1.0000   3rd Qu.:0.34711   3rd Qu.:1.440   3rd Qu.:0.3900
##  Max.   :1.0000   Max.   :1.98454   Max.   :2.280   Max.   :1.0800
##                   NA's   :14
##   spiritpercap     totpercap         mrate         year_factor
##  Min.   :0.480   Min.   :1.380   Min.   : 52.47   1970   : 51
##  1st Qu.:0.810   1st Qu.:2.203   1st Qu.:111.99   1971   : 51
##  Median :1.050   Median :2.690   Median :130.70   1972   : 51
##  Mean   :1.165   Mean   :2.773   Mean   :136.44   1973   : 51
##  3rd Qu.:1.260   3rd Qu.:3.060   3rd Qu.:149.90   1974   : 51
##  Max.   :4.450   Max.   :6.920   Max.   :331.89   1975   : 51
##                                                   (Other):408
```

Setting weights = pop:

```
reg1 = lm_robust(mrate~legal+state+year_factor-1, data=mlda, clusters=state, se_type='stata')
reg2 = lm_robust(mrate~legal+state+year_factor+state:year-1, data=mlda, clusters=state, se_type='stata')
reg3<-lm_robust(mrate~legal+state+year-1, data=mlda,
weights=pop, clusters=state, se_type='stata')
estimates = c(coef(reg1)[1], coef(reg2)[1], coef(reg3)[1])
std_errors = c(reg1$std.error[1], reg2$std.error[1], reg3$std.error[1])
results = cbind(estimates, std_errors)
row.names(results)<-paste0('reg',1:3)
results
```

```
##      estimates std_errors
## reg1 10.804141   4.592205
## reg2  8.466624   5.097812
## reg3 12.000347   3.346856
```

Allow for state-specific e??ects by including an interaction between state and year:

```
reg4 = lm_robust(mrate~legal+state+year+state:year-1, data=mlda, weights =pop,
                 clusters=state, se_type='stata')
estimates2 = c(coef(reg1)[1], coef(reg2)[1], coef(reg3)[1], coef(reg4)[1])
std_errors2 = c(reg1$std.error[1], reg2$std.error[1], reg3$std.error[1],reg4$std.error[1])
results2 = cbind(estimates2, std_errors2)
row.names(results2)<-paste0('reg',1:4)
results2
```

```
##      estimates2 std_errors2
## reg1  10.804141    4.592205
## reg2   8.466624    5.097812
## reg3  12.000347    3.346856
## reg4  10.104667    3.455301
```

Reg1 and Reg3 both indicates that the legal effect is significant and robust. However, there is no interaction term as DiD does. The causal effect is not sure. Reg4 includes an interaction between state and factor(year). Reg2 includes the interaction between state and year. It is hard to interpret the result.

Control for beer taxes:

```
reg5 = lm_robust(mrate~legal+state+year_factor+beertaxa-1, data=mlda, clusters=state, se_type='stata')
reg6 = lm_robust(mrate~legal+state+year_factor+state:year+beertaxa-1, data=mlda, clusters=state, se_typ
reg7<-lm_robust(mrate~legal+state+year-1+beertaxa, data=mlda,
weights=pop, clusters=state, se_type='stata')
estimates3 = c(coef(reg5)[1], coef(reg6)[1], coef(reg7)[1])
std_errors3 = c(reg5$std.error[1], reg6$std.error[1], reg7$std.error[1])
results3 = cbind(estimates3, std_errors3)
row.names(results3)<-paste0('reg',5:7)
results3
```

```
##      estimates3 std_errors3
## reg5   10.98272    4.691734
## reg6   10.02933    4.915832
## reg7   12.29245    3.283094
```

The beer tax is significant. The estimates are higher and the standard errors are smaller, the t values are higher, although it is not DiD.