HW4 Econometrics

Tianyu Gao

2022-11-02

1

a)

```
.33/.08
```

```
## [1] 4.125
```

```
.35/.38
```

```
## [1] 0.9210526
```

```
1.53/.13
```

```
## [1] 11.76923
```

```
qnorm(.975)
```

```
## [1] 1.959964
```

$$\because\ .33/.08 = 4.13 > 1.96$$
$$.35/.38 = .92 < 1.96$$
$$1.53/.13 = 11.77 > 1.96$$

$$\therefore \text{ At 95\%Confidence Level, coefficient for} \frac{M_{it}}{W_{it}} \text{ and } uram_{it} \text{are significant, while for } SHY_{it} \text{ it is not significant}$$

b)

$$.07/.10 = .7 < 1.96$$
$$.19/.22 = .86 < 1.96$$
$$.54/.11 = 4.91$$

```
.07/.1
```

```
## [1] 0.7
```

```
.19/.22
```

```
## [1] 0.8636364
```

```
.54/.11
```

```
## [1] 4.909091
```

The coefficient for $\frac{M_{it}}{W_{it}}$ becomes not significant. This is because

$$\mathrm{corr}(E_{it}, D_i) \neq 0, \mathrm{corr}(D_i, \frac{M_{it}}{W_{it}}) \neq 0$$

In other words, the ratio $\frac{M_{it}}{W_{it}}$ is different in different states at a given period. When the state fixed effects is not included in the model, there is an omitted variable bias, in other words,

$$\mathrm{corr}(\frac{M_{it}}{W_{it}}, u_{it}) \neq 0$$

$$\because \hat{\beta_i} \xrightarrow{p} \frac{\sigma_X}{\sigma_u}\varrho_{Xu}$$

$$\therefore \mathrm{corr}(\frac{M_{it}}{W_{it}}, D_i) > 0$$

And the coefficient seems larger when not taking state fixed effect into account.

c)

This result rejects the idea that all state fixed effect can be removed from the model. To test this hypothesis, I will need 50 dummy variables($D_1, \ldots, D_{50}$) (indicating 50 states) to run a regression including the time fixed effects, and the test for removing state fixed effects is as follows:

$$H_0 : D_1 = \cdots = D_{50}$$
$$H_A : \exists i \in [1, 50], D_i \neq 0$$

If H_0 is rejected, we cannot remove state fixed effects.

2

a)

```
120700-30100
```

```
## [1] 90600
```

For places far away from the treatment plant, $\mathrm{nearplant}_i = 0, \mathrm{rprice}_i = 120700$

For places near the treatment plant, $\mathrm{nearplant}_i = 1, \mathrm{rprice}_i = 120700 - 30100 = 90600$

b)

```
.457 - .063
```

```
## [1] 0.394
```

year81 controls the year fixed effect, and the coefficient means for places near the plant, the price is 39.4% higher in 1981 than in 1978, while for places far from the plant, the price is 45.7% higher in 1981 than in 1978.

```
-.34 - .063
```

```
## [1] -0.403
```

$\text{nearplant}_{\text{it}}$ is the variable of interest. From the regression model, we know that for places near the desired plant location (this could be the result of construction of the plant) in 1978, the price is 34.0% lower than places far from the location, while in 1981, the price is 40.3% lower.

c)

```
.063/.023
```

```
## [1] 2.73913
```

The treatment plant suppresses the the price of houses. I am sure about the causal effect, because the coefficient of the interaction term, which represents the drop of the home price near the plant from 1978 to 1981, is significant.

3

a)

Model (ii) is most appropriate. Because international trade is how development in South Africa can impact other countries' GDP, and $\text{GDP} = \text{C} + \text{I} + \text{G} + \text{NX}$, which means SAFit, therefore, trade needs to be added to the model as a control variable.

b) & c)

The net effect is not a constant, it is $.62 + .21\text{TRADE}_{\text{it}}$

4

a)

```
setwd('/Users/kevintsukuyo/Documents/Course Files/2022F/Applied Econometrics/HW4')
guns = read.csv('guns.csv')
```

```
colnames(guns)
```

```
##  [1] "year"        "vio"         "mur"         "rob"         "incarc_rate"
##  [6] "pb1064"      "pw1064"      "pm1029"      "pop"         "avginc"
## [11] "density"     "stateid"     "shall"
```

(1)

```
library(lmtest)
```

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

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

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

```
library(sandwich)
library(plm)
```

To minimize the omitted variable bias, we use panel linear model here

```
pla1 = plm(log(vio) ~ shall, data = guns, index = c('year', 'stateid'), model = 'within')
summary(pla1)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log(vio) ~ shall, data = guns, model = "within",
##     index = c("year", "stateid"))
##
## Balanced Panel: n = 23, T = 51, N = 1173
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -2.145382 -0.436980  0.053547  0.405558  1.684921
##
## Coefficients:
##        Estimate Std. Error t-value  Pr(>|t|)
## shall -0.598122   0.044464 -13.452 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    474.53
## Residual Sum of Squares: 409.97
## R-Squared:      0.13606
## Adj. R-Squared: 0.11876
## F-statistic: 180.948 on 1 and 1149 DF, p-value: < 2.22e-16
```

```
coeftest(pla1, vcovHC(pla1, type = 'HC1', cluster = 'group'))
```

```
##
## t test of coefficients:
##
##        Estimate Std. Error t value  Pr(>|t|)
## shall -0.598122   0.045679 -13.094 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(2)

```
pla2 = plm(log(vio)~shall + incarc_rate + density + avginc + pop + pb1064 + pw1064 + pm1029,
data = guns, index = c('year', 'stateid'), model = 'within')
summary(pla2)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log(vio) ~ shall + incarc_rate + density + avginc +
##     pop + pb1064 + pw1064 + pm1029, data = guns, model = "within",
##     index = c("year", "stateid"))
##
## Balanced Panel: n = 23, T = 51, N = 1173
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.799779 -0.253651  0.054167  0.299845  1.157193
##
## Coefficients:
##                Estimate  Std. Error t-value  Pr(>|t|)
## shall       -0.28776925  0.03367595 -8.5452 < 2.2e-16 ***
## incarc_rate  0.00193051  0.00011429 16.8919 < 2.2e-16 ***
## density     -0.00887417  0.01394780 -0.6362   0.52475
## avginc       0.01285649  0.00795904  1.6153   0.10652
## pop          0.04082345  0.00252389 16.1748 < 2.2e-16 ***
## pb1064       0.09998919  0.01823019  5.4848 5.094e-08 ***
## pw1064       0.04007982  0.00912419  4.3927 1.224e-05 ***
## pm1029      -0.04442077  0.01751542 -2.5361   0.01134 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    474.53
## Residual Sum of Squares: 199.27
## R-Squared:      0.58008
## Adj. R-Squared: 0.56904
## F-statistic: 197.192 on 8 and 1142 DF, p-value: < 2.22e-16
```

```
coeftest(pla2, vcovHC(pla2, type = 'HC1', cluster = 'group'))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error  t value  Pr(>|t|)
## shall       -0.28776925  0.02556436 -11.2567 < 2.2e-16 ***
## incarc_rate  0.00193051  0.00018572  10.3946 < 2.2e-16 ***
## density     -0.00887417  0.01517937  -0.5846  0.558918
## avginc       0.01285649  0.00441282   2.9134  0.003644 **
## pop          0.04082345  0.00196696  20.7546 < 2.2e-16 ***
## pb1064       0.09998919  0.02283499   4.3788 1.303e-05 ***
## pw1064       0.04007982  0.01184967   3.3824  0.000743 ***
## pm1029      -0.04442077  0.01596066  -2.7831  0.005472 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The violent crime rate is 28% lower instate having a shall-carry law in effect than states without. And the result is statistically significant.

b)

Adding control variables changes the estimated effect. And the coefficient is statistically significant.

c)

If we do not use panel regression, suppose the variable is $\mathrm{U_i}$ for $\mathrm{i^{th}}$ observation, and the vector of the shall low variable for $\mathrm{i^{th}}$ observation is $\mathrm{shall_i}$

$\mathrm{U_i}$ varies across states and $\mathrm{shall_i}$ varies across states→ $\mathrm{corr(U_i|Shall_i)} \neq 0$

Also, $\mathrm{corr(shall_i, ln(vio_i)} \neq 0$

d)

After adding the control variables which are important but regardless of the panel, we can run following regression:

```
ln = lm(log(vio)~shall + incarc_rate + density + avginc + pop + pb1064 + pw1064 + pm1029, da
ta = guns)
summary(ln)
```

```
##
## Call:
## lm(formula = log(vio) ~ shall + incarc_rate + density + avginc +
##     pop + pb1064 + pw1064 + pm1029, data = guns)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.72300 -0.26620  0.04767  0.30478  1.05998
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.9817381  0.5433938   5.487 5.01e-08 ***
## shall       -0.3683869  0.0325674 -11.312  < 2e-16 ***
## incarc_rate  0.0016126  0.0001072  15.050  < 2e-16 ***
## density      0.0266885  0.0131681   2.027 0.042915 *
## avginc       0.0012051  0.0077802   0.155 0.876931
## pop          0.0427098  0.0025588  16.691  < 2e-16 ***
## pb1064       0.0808526  0.0166514   4.856 1.36e-06 ***
## pw1064       0.0312005  0.0083776   3.724 0.000205 ***
## pm1029       0.0088709  0.0107737   0.823 0.410458
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4277 on 1164 degrees of freedom
## Multiple R-squared:  0.5643, Adjusted R-squared:  0.5613
## F-statistic: 188.4 on 8 and 1164 DF,  p-value: < 2.2e-16
```

```
coeftest(ln, vcovHC(ln, type = 'HC1'))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept)  2.98173807  0.60901978   4.8960 1.116e-06 ***
## shall       -0.36838695  0.03478791 -10.5895 < 2.2e-16 ***
## incarc_rate  0.00161263  0.00018069   8.9246 < 2.2e-16 ***
## density      0.02668848  0.01434939   1.8599  0.063151 .
## avginc       0.00120512  0.00727782   0.1656  0.868510
## pop          0.04270983  0.00314664  13.5731 < 2.2e-16 ***
## pb1064       0.08085261  0.01999243   4.0442 5.597e-05 ***
## pw1064       0.03120051  0.00972707   3.2076  0.001375 **
## pm1029       0.00887088  0.01206040   0.7355  0.462160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we only consider the state fixed effects, we should run the following regression model:

```
pld = plm(log(vio)~shall + incarc_rate + density + avginc + pop + pb1064 + pw1064 + pm1029,
data = guns, index = c('stateid'), model = 'within')
summary(pld)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log(vio) ~ shall + incarc_rate + density + avginc +
##     pop + pb1064 + pw1064 + pm1029, data = guns, model = "within",
##     index = c("stateid"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.5518825 -0.1016341  0.0094813  0.1053023  0.5572337
##
## Coefficients:
##                Estimate  Std. Error t-value  Pr(>|t|)
## shall       -4.6142e-02  1.8867e-02 -2.4456   0.01461 *
## incarc_rate -7.1008e-05  9.3602e-05 -0.7586   0.44824
## density     -1.7229e-01  8.5036e-02 -2.0261   0.04299 *
## avginc      -9.2037e-03  5.9083e-03 -1.5578   0.11957
## pop          1.1525e-02  8.7239e-03  1.3210   0.18676
## pb1064       1.0428e-01  1.7756e-02  5.8728 5.649e-09 ***
## pw1064       4.0861e-02  5.0745e-03  8.0522 2.074e-15 ***
## pm1029      -5.0273e-02  6.4037e-03 -7.8505 9.684e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    36.789
## Residual Sum of Squares: 28.777
## R-Squared:      0.21779
## Adj. R-Squared: 0.17707
## F-statistic: 38.7715 on 8 and 1114 DF, p-value: < 2.22e-16
```

```
coeftest(pld, vcovHC(pld, type = 'HC1', cluster = 'group'))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value Pr(>|t|)
## shall       -4.6142e-02  4.1350e-02 -1.1159 0.264716
## incarc_rate -7.1008e-05  2.4788e-04 -0.2865 0.774581
## density     -1.7229e-01  1.3626e-01 -1.2645 0.206332
## avginc      -9.2037e-03  1.2837e-02 -0.7170 0.473547
## pop          1.1525e-02  1.4084e-02  0.8183 0.413367
## pb1064       1.0428e-01  3.2363e-02  3.2222 0.001309 **
## pw1064       4.0861e-02  1.3326e-02  3.0663 0.002219 **
## pm1029      -5.0273e-02  2.0491e-02 -2.4534 0.014303 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we only consider state fixed effects in the model, the coefficients will change compared to the regression without panel effects. Intuitively, the panel model is more credible, because it takes the difference among states into account, like the agresiveness of the local people.

e)

```
ple = plm(log(vio)~shall + incarc_rate + density + avginc + pop + pb1064 + pw1064 + pm1029,
data = guns, index = c('year'), model = 'within')
summary(ple)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log(vio) ~ shall + incarc_rate + density + avginc +
##     pop + pb1064 + pw1064 + pm1029, data = guns, model = "within",
##     index = c("year"))
##
## Balanced Panel: n = 23, T = 51, N = 1173
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.799779 -0.253651  0.054167  0.299845  1.157193
##
## Coefficients:
##                Estimate  Std. Error t-value  Pr(>|t|)
## shall       -0.28776925  0.03367595 -8.5452 < 2.2e-16 ***
## incarc_rate  0.00193051  0.00011429 16.8919 < 2.2e-16 ***
## density     -0.00887417  0.01394780 -0.6362   0.52475
## avginc       0.01285649  0.00795904  1.6153   0.10652
## pop          0.04082345  0.00252389 16.1748 < 2.2e-16 ***
## pb1064       0.09998919  0.01823019  5.4848 5.094e-08 ***
## pw1064       0.04007982  0.00912419  4.3927 1.224e-05 ***
## pm1029      -0.04442077  0.01751542 -2.5361   0.01134 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    474.53
## Residual Sum of Squares: 199.27
## R-Squared:      0.58008
## Adj. R-Squared: 0.56904
## F-statistic: 197.192 on 8 and 1142 DF, p-value: < 2.22e-16
```

```
coeftest(ple, vcovHC(ple, type = 'HC1', cluster = 'group'))
```

```
##
## t test of coefficients:
##
##               Estimate  Std. Error  t value  Pr(>|t|)
## shall       -0.28776925  0.02556436 -11.2567 < 2.2e-16 ***
## incarc_rate  0.00193051  0.00018572  10.3946 < 2.2e-16 ***
## density     -0.00887417  0.01517937  -0.5846  0.558918
## avginc       0.01285649  0.00441282   2.9134  0.003644 **
## pop          0.04082345  0.00196696  20.7546 < 2.2e-16 ***
## pb1064       0.09998919  0.02283499   4.3788 1.303e-05 ***
## pw1064       0.04007982  0.01184967   3.3824  0.000743 ***
## pm1029      -0.04442077  0.01596066  -2.7831  0.005472 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

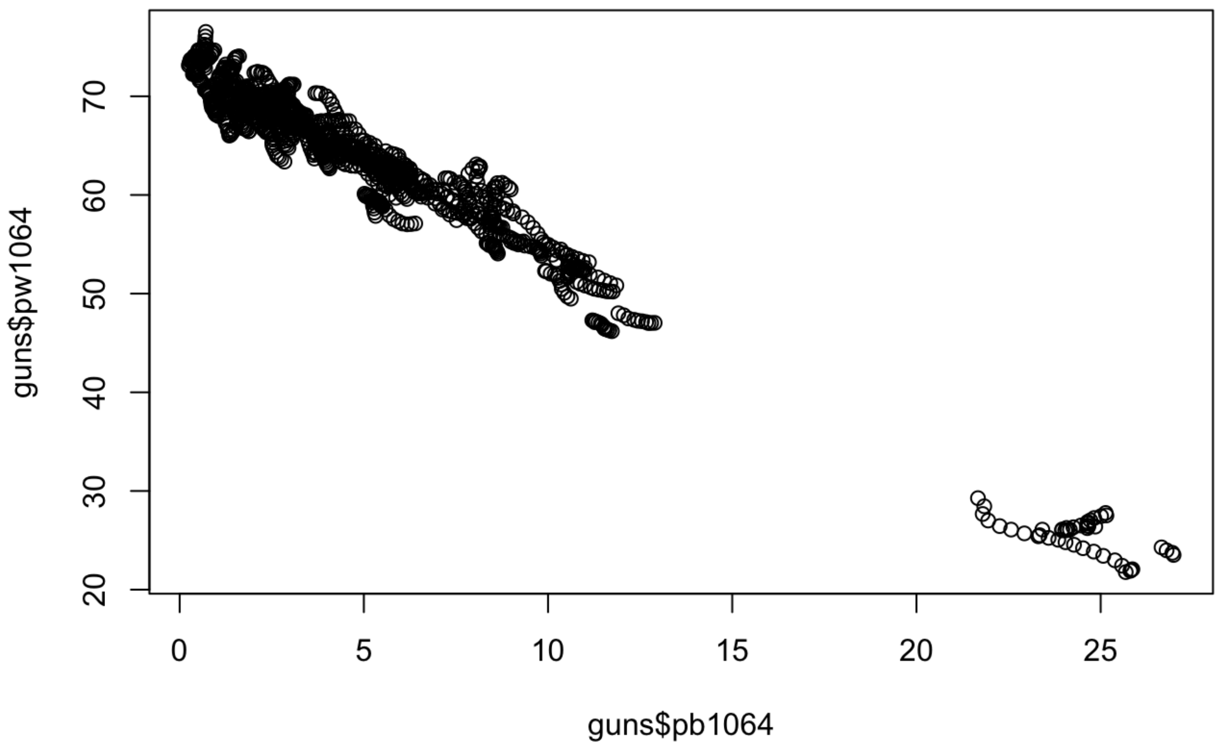
```
coeftest(ln, vcovHC(ln, type = 'HC1'))
```

```
##
## t test of coefficients:
##
##               Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept)  2.98173807  0.60901978   4.8960 1.116e-06 ***
## shall       -0.36838695  0.03478791 -10.5895 < 2.2e-16 ***
## incarc_rate  0.00161263  0.00018069   8.9246 < 2.2e-16 ***
## density      0.02668848  0.01434939   1.8599  0.063151 .
## avginc       0.00120512  0.00727782   0.1656  0.868510
## pop          0.04270983  0.00314664  13.5731 < 2.2e-16 ***
## pb1064       0.08085261  0.01999243   4.0442 5.597e-05 ***
## pw1064       0.03120051  0.00972707   3.2076  0.001375 **
## pm1029       0.00887088  0.01206040   0.7355  0.462160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The coefficients are different. And the panel one is more credible. Because it takes into account the difference among different periods, like some regional or national political regulation about law enforcement.

f)

```
plot(guns$pb1064, guns$pw1064)
```



The biggest concern of mine is the multicollinearity between pw1064 and pb1064. As shown above, the two variable are negatively correlated, this is because population of black and population of white make up the majority of the population, and increase in one ethnic group leads to decrease in another one.

g)

Because the coefficient is significantly negative, the laws suppresses the violation.