The University of Queensland
Australia
CREATE CHANGE

ECON2300 - Introductory Econometrics

Tutorial 8: Regression with Panel Data

Tutor: Francisco Tavares Garcia

R-Exercise 4 is available!

Posted on: Monday, 18 September 2023 06:00:00 o'clock AEST

Dear ECON2300 Students,

R-Exercise 4 is now available in the "R-Exercises: Analysis of Data and Short Report" folder, which you can access via the Assessment tab.

The due date for R-Exercise 4 is **Friday, September 22, 2023, 4pm**

Please read all instructions carefully before commencing the R-Exercise. For convenience, a copy of the R-Exercise instructions has been presented below.

===

Instructions:

Please pay close attention to the number of decimal places required (if any) for each answer. The required number of decimal places may differ from question to question.

Avoid rounding during intermediate calculations where possible.

This R-Exercise is not timed. This means that you can open the R-Exercise and return to it as many times as you need to (provided that you do not click submit).

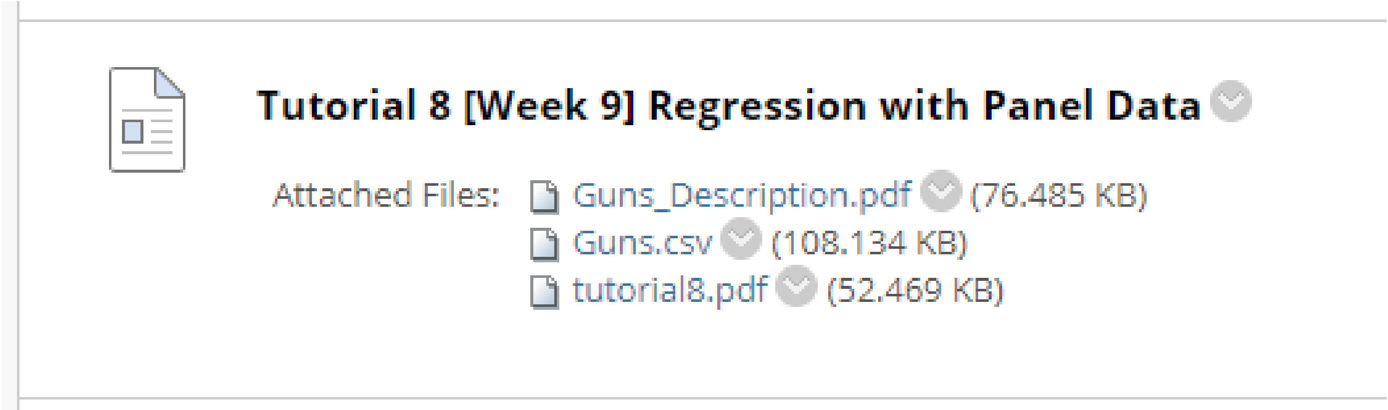
There is only one attempt for this R-Exercise.

The R-Exercise is marked out of 7, but will contribute 10% towards your final grade if it is among the highest 3 of your 5 R-Exercise scores across the semester.

The closing time for this R-Exercise is **4pm on Friday, September 22, 2023**. Please make sure that you have submitted your answers by this time. Remember that you must click submit before the deadline for your R-Exercise to be marked.

Please Note: If you encounter any technical issues with the R-Exercise, please email the CML coordinator at cml.2300@uq.edu.au. Do not email R-Exercise issues to the Course Coordinator or Course Administrator. Otherwise there may be a delay in responding to your enquiry.

- Download the files for tutorial 08 from Blackboard,
- save them into a folder for this tutorial.



- Copy the code from Codeshare,
- https://codeshare.io/tut08
- Paste the code in a new script in RStudio,
- Save the script in the same folder as the data.

E10.1 Some U.S. states have enacted laws that allow citizens to carry concealed weapons. These laws are known as "shall-issue" laws because they instruct local authorities to issue a concealed weapons permit to all applicants who are citizens, are mentally competent, and have not been convicted of a felony. (Some states have some additional restrictions.) Proponents argue that if more people carry concealed weapons, crime will decline because criminals will be deterred from attacking other people. Opponents argue that crime will increase because of accidental or spontaneous use of the weapons. In this exercise, you will analyze the effect of concealed weapons laws on violent crimes, using the data file `Guns.csv`, which contains a balanced panel of data from the 50 U.S. states plus the District of Columbia for the years 1977 through 1999. A detailed description is given in `Guns_Description.pdf`.

Variable	Definition
vio	violent crime rate (incidents per 100,000 members of the population)
rob	robbery rate (incidents per 100,000)
mur	murder rate (incidents per 100,000)
shall	= 1 if the state has a shall-carry law in effect in that year = 0 otherwise
incarc_rate	incarceration rate in the state in the previous year (sentenced prisoners per 100,000 residents; value for the previous year)
density	population per square mile of land area, divided by 1000
avginc	real per capita personal income in the state, in thousands of dollars
pop	state population, in millions of people
pm1029	percent of state population that is male, ages 10 to 29
pw1064	percent of state population that is white, ages 10 to 64
pb1064	percent of state population that is black, ages 10 to 64
stateid	ID number of states (Alabama = 1, Alaska = 2, etc.)
year	Year (1977-1999)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	year	vio	mur	rob	incarc_rat	pb1064	pw1064	pm1029	pop	avginc	density	stateid	shall
2	77	414.4	14.2	96.8	83	8.384873	55.12291	18.17441	3.780403	9.563149	0.074552	1	0
3	78	419.1	13.3	99.1	94	8.352101	55.14367	17.99408	3.831838	9.932	0.075567	1	0
4	79	413.3	13.2	109.5	144	8.329575	55.13586	17.83934	3.866248	9.877028	0.076245	1	0
5	80	448.5	13.2	132.1	141	8.408386	54.91259	17.7342	3.900368	9.541428	0.076829	1	0
6	81	470.5	11.9	126.5	149	8.483435	54.92513	17.67372	3.918531	9.548351	0.077187	1	0
7	82	447.7	10.6	112	183	8.514	54.89621	17.51052	3.925229	9.478919	0.077319	1	0
8	83	416	9.2	98.4	215	8.545608	54.83936	17.35089	3.934103	9.783	0.077493	1	0
9	84	431.2	9.4	96.1	243	8.559511	54.77876	17.11902	3.951826	10.3572	0.077842	1	0
10	85	457.5	9.8	105.4	256	8.562801	54.67899	16.85875	3.97252	10.72586	0.07825	1	0

```r
library(readr)      # package for fast read rectangular data
library(dplyr)      # package for data manipulation
library(estimatr)   # package for commonly used estimators with robust SE
library(texreg)     # package converting R regression output to LaTeX/HTML tables
library(plm)        # package for estimating linear panel data models
library(dummies)    # package for creating dummy/indicator variables
```

SW E10.1

```r
rm(list = ls())
setwd("/Users/uqdkim7/Dropbox/Teaching/R tutorials/Data")
Guns <- read_csv("Guns.csv") %>%
  mutate(lvio = log(vio), lrob = log(rob), lmur = log(mur))
attach(Guns)
```

(a) Estimate (1) a regression of ln (vio) against shall and (2) a regression of ln (vio) against shall, incarc_rate, density, avginc, pop, pb1064, pw1064, and pm1029.

The solutions for (a)–(c) will reference regression results summarized in Table 1 (See page 2)[1].

```r
# fit pooled OLS using cluster and heteroskedasticity robust SE
pols1 = lm_robust(lvio ~ shall, data = Guns, se_type = "stata", clusters = stateid)
pols2 = lm_robust(lvio ~ shall + incarc_rate + density + avginc +
                    pop + pb1064 + pw1064 + pm1029,
                  data = Guns, se_type = "stata", clusters = stateid)
# fit fixed effects model
fe1 = plm(lvio ~ shall + incarc_rate + density + avginc +
            pop + pb1064 + pw1064 + pm1029,
          data = Guns, model = "within", index = c("stateid", "year"))
# fit fixed effects model with time effects
fe2 = plm(lvio ~ shall + incarc_rate + density + avginc +
            pop + pb1064 + pw1064 + pm1029,
          data = Guns, model = "within", index = c("stateid", "year"),
          effect = "twoway")
# or equivalently use function factor() to include dummies
fe3 = plm(lvio ~ shall + incarc_rate + density + avginc +
            pop + pb1064 + pw1064 + pm1029 + factor(year),
          data = Guns, model = "within", index = c("stateid", "year"))
```

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

(a) Estimate (1) a regression of ln (vio) against shall and (2) a regression of ln (vio) against shall, incarc_rate, density, avginc, pop, pb1064, pw1064, and pm1029.

To my knowledge, there is no option for plm that can help computing cluster robust SE. Here we compute them using the vcovHC function as follows:

```
# compute cluster robust SE for FE estimator
SE.fe1 <- sqrt(diag(vcovHC(fe1, type="sss", cluster="group")))
SE.fe2 <- sqrt(diag(vcovHC(fe2, type="sss", cluster="group")))
```

To use texreg, all SEs and p-values should be customized.

```
# extract SE of pooled OLS estimator
SE.pols1 <- pols1$std.error
SE.pols2 <- pols2$std.error
# compute p-values
p.pols1 <- 2*(1 - pnorm(abs(pols1$coefficients/SE.pols1)))
p.pols2 <- 2*(1 - pnorm(abs(pols2$coefficients/SE.pols2)))
p.fe1 <- 2*(1 - pnorm(abs(fe1$coefficients/SE.fe1)))
p.fe2 <- 2*(1 - pnorm(abs(fe2$coefficients/SE.fe2)))
# generate LaTeX code for Table 1
texreg(list(pols1, pols2, fe1, fe2), include.ci = F, caption.above = T, digits = 3,
          override.se = list(SE.pols1,SE.pols2,SE.fe1,SE.fe2),
       override.pvalues = list(p.pols1, p.pols2, p.fe1, p.fe2),
       caption = "Violent Crime Rate and Shall-Carry Law",
       custom.model.names = c("(1) Pooled OLS (1)", "(2) Pooled OLS",
                              "(3) Fixed Effects", "(4) Fixed Effects & Time Effects"))
```

(a) Estimate (1) a regression of ln (vio) against shall and (2) a regression of ln (vio) against shall, incarc_rate, density, avginc, pop, pb1064, pw1064, and pm1029.

	(1) Pooled OLS 1	(2) Pooled OLS 2
(Intercept)	6.135***	2.982
	(0.079)	(2.167)
shall	−0.443**	−0.368**
	(0.157)	(0.114)
incarc_rate		0.002**
		(0.001)
density		0.027
		(0.041)
avginc		0.001
		(0.024)
pop		0.043***
		(0.012)
pb1064		0.081
		(0.071)
pw1064		0.031
		(0.034)
pm1029		0.009
		(0.034)
R^2	0.087	0.564
Adj. R^2	0.086	0.561
Num. obs.	1173	1173
RMSE	0.617	0.428
N Clusters	51	51

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

i. Interpret the coefficient on `shall` in regression (2). Is this estimate large or small in a "real-world" sense?

ii. Does adding the control variables in regression (2) change the estimated effect of a shall-carry law in regression (1) as measured by statistical significance? As measured by the "real-world" significance of the estimated coefficient?

iii. Suggest a variable that varies across states but plausibly varies little – or not at all – over time and that could cause omitted variable bias in regression (2)

i The coefficient is -0.368, which suggests that shall-issue laws reduce violent crime by 36%. This is a large effect.

ii The coefficient in (1) is -0.443, while in (2) it is -0.368. Both are highly statistically significant. Adding the control variables results in a small drop in the coefficient.

iii There are several examples. Here are two: Attitudes towards guns and crime, and quality of police and other crime-prevention programs.

(b) Do the results change when you add fixed state effects? If so, which set of regression results is more credible, and why?

Fixed Effects Regression SW Section 10.3

- What if you have more than 2 time periods ($T > 2$)?

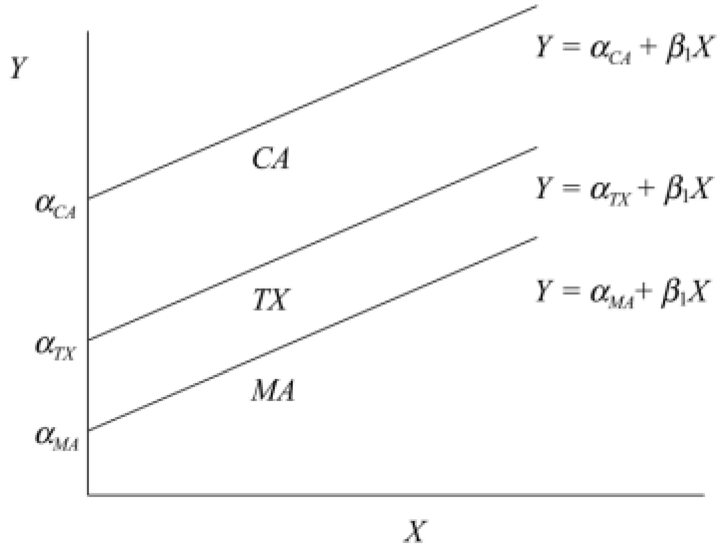
$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it}, \quad i = 1, \ldots, n, \quad T = 1, \ldots, T$$

- We can rewrite this in two equivalent ways:
  - "$n - 1$ binary regressor" regression model
  - "Fixed Effects" regression model
- We first rewrite this in "fixed effects" form. Suppose we have $n = 3$ states: California (CA), Texas (TX), and Massachusetts (MA).
- For $i = CA$, we rewrite the model above as follow;

$$Y_{CA,t} = \underbrace{\beta_0 + \beta_2 Z_{CA}}_{=\alpha_{CA}} + \beta_1 X_{CA,t} + u_{CA,t}$$
$$= \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t}$$

- So, α_{CA} 'picks up' Z_{CA}, unobserved factors like 'traffic density' and 'driving/drinking culture' in CA, which may cause omitted variable bias.

The regression lines for each state in a picture



- Recall that we can re-write the fixed effect form using binary regressors;

$$Y_{it} = \beta_0 + \gamma_{TX} DTX_i + \gamma_{CA} DCA_i + \beta_1 X_{it} + u_{it}$$

where DTX_i is the dummy for TX and DCA_i is for CA.

- **Question:** Why DMA not included?

(b) Do the results change when you add fixed state effects? If so, which set of regression results is more credible, and why?

Table 1: Violent Crime Rate and Shall-Carry

	(1) Pooled OLS (1)	(2) Pooled OLS	(3) Fixed Effects
(Intercept)	6.135***	2.982	
	(0.079)	(2.167)	
shall	−0.443**	−0.368**	−0.046
	(0.157)	(0.114)	(0.042)
incarc_rate		0.002**	−0.000
		(0.001)	(0.000)
density		0.027	−0.172
		(0.041)	(0.138)
avginc		0.001	−0.009
		(0.024)	(0.013)
pop		0.043***	0.012
		(0.012)	(0.014)
pb1064		0.081	0.104**
		(0.071)	(0.033)
pw1064		0.031	0.041**
		(0.034)	(0.013)
pm1029		0.009	−0.050*
		(0.034)	(0.021)
R^2	0.087	0.564	0.218
Adj. R^2	0.086	0.561	0.177
Num. obs.	1173	1173	1173
RMSE	0.617	0.428	
N Clusters	51	51	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

In (3) the coefficient on `shall` falls to −0.046, a large reduction in the coefficient from (2). Evidently there was important omitted variable bias in (2). The estimate is not statistically significantly different from zero.

(c) Do the results change when you add fixed time effects? If so, which set of regression results is more credible, and why?

Time fixed effects only

- If there was no entity FE, the model would be given as

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_3 S_t + u_{it}$$

- That is, the **time fixed effects regression model** is

$$Y_{it} = \beta_1 X_{it} + \lambda_t + u_{it}$$

  where $\lambda_1, \ldots, \lambda_T$ are known as time fixed effects.
- This model can be equivalently written with $T - 1$ time dummies

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \cdots + \delta_T BT_t + u_{it}$$

  where $B2_t = 1$ if t is 2, otherwise it is zero, etc.
- Estimation and inference is parallel to the entity FE case above.
  1. "$T - 1$" binary regressor" OLS regressions
  2. "time-demeaned" OLS regression

Estimation with both entity and time fixed effects

- We may have both entity FEs and time FEs. Then, the **entity and time fixed effects regression model** is

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}$$

- When $T = 2$, computing the first difference and including an intercept is equivalent to including entity and time fixed effects.
- When $T > 2$, there are a number of alternative algorithms to estimate this model;
  - entity demeaning & $T - 1$ time indicators
  - time demeaning & $n - 1$ entity indicators
  - $T - 1$ time indicators & $n - 1$ entity indicators
  - entity & time demeaning

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

(c) Do the results change when you add fixed time effects? If so, which set of regression results is more credible, and why?

Table 1: Violent Crime Rate and Shall-Carry Law

	(1) Pooled OLS (1)	(2) Pooled OLS	(3) Fixed Effects	(4) Fixed Effects and Time Effects
(Intercept)	6.135***	2.982		
	(0.079)	(2.167)		
shall	−0.443**	−0.368**	−0.046	−0.028
	(0.157)	(0.114)	(0.042)	(0.040)
incarc_rate		0.002**	−0.000	0.000
		(0.001)	(0.000)	(0.000)
density		0.027	−0.172	−0.092
		(0.041)	(0.138)	(0.123)
avginc		0.001	−0.009	0.001
		(0.024)	(0.013)	(0.016)
pop		0.043***	0.012	−0.005
		(0.012)	(0.014)	(0.015)
pb1064		0.081	0.104**	0.029
		(0.071)	(0.033)	(0.049)
pw1064		0.031	0.041**	0.009
		(0.034)	(0.013)	(0.024)
pm1029		0.009	−0.050*	0.073
		(0.034)	(0.021)	(0.052)
R^2	0.087	0.564	0.218	0.056
Adj. R^2	0.086	0.561	0.177	−0.013
Num. obs.	1173	1173	1173	1173
RMSE	0.617	0.428		
N Clusters	51	51		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

(c) Do the results change when you add fixed time effects? If so, which set of regression results is more credible, and why?

```r
# test time effects
pFtest(fe2, fe1)
```

```
> pFtest(fe2, fe1)

        F test for twoways effects

data:  lvio ~ shall + incarc_rate + density + avginc + pop + pb1064 +  ...
F = 17.075, df1 = 22, df2 = 1092, p-value < 2.2e-16
alternative hypothesis: significant effects
```

The coefficient in (4) falls further to -0.028. The coefficient is insignificantly different from zero. The time effects are jointly statistically significant (p-value ≈ 0), so this regression seems better specified than (3).

pFtest {plm} R Documentation

F Test for Individual and/or Time Effects

Description

Test of individual and/or time effects based on the comparison of the `within` and the `pooling` model.

Usage

```r
pFtest(x, ...)

## S3 method for class 'formula'
pFtest(x, data, ...)

## S3 method for class 'plm'
pFtest(x, z, ...)
```

The University of Queensland
AUSTRALIA

(d) Repeat the analysis using ln (rob) and ln (mur) in place of ln (vio).

```
pols1 = lm_robust(lrob ~ shall, data = Guns, se_type = "stata", clusters = stateid)
pols2 = lm_robust(lrob ~ shall + incarc_rate + density + avginc +
                    pop + pb1064 + pw1064 + pm1029,
                  data = Guns, se_type = "stata", clusters = stateid)
fe1 = plm(lrob ~ shall + incarc_rate + density + avginc +
            pop + pb1064 + pw1064 + pm1029,
          data = Guns, model = "within", index = c("stateid", "year"))
fe2 = plm(lrob ~ shall + incarc_rate + density + avginc +
            pop + pb1064 + pw1064 + pm1029,
          data = Guns, model = "within", index = c("stateid", "year"),
          effect = "twoway")
SE.pols1 <- pols1$std.error
SE.pols2 <- pols2$std.error
SE.fe1 <- sqrt(diag(vcovHC(fe1, type="sss", cluster="group")))
SE.fe2 <- sqrt(diag(vcovHC(fe2, type="sss", cluster="group")))
p.pols1 <- 2*(1 - pnorm(abs(pols1$coefficients/SE.pols1)))
p.pols2 <- 2*(1 - pnorm(abs(pols2$coefficients/SE.pols2)))
p.fe1 <- 2*(1 - pnorm(abs(fe1$coefficients/SE.fe1)))
p.fe2 <- 2*(1 - pnorm(abs(fe2$coefficients/SE.fe2)))

texreg(list(pols1, pols2, fe1, fe2), include.ci = F, caption.above = T, digits = 3,
       override.se = list(SE.pols1,SE.pols2,SE.fe1,SE.fe2),
       override.pvalues = list(p.pols1, p.pols2, p.fe1, p.fe2),
       caption = "Robbery Rate and Shall-Carry Law",
       custom.model.names = c("(1) Pooled OLS (1)", "(2) Pooled OLS",
                              "(3) Fixed Effects", "(4) Fixed Effects & Time Effects"))
```

(d) Repeat the analysis using ln `(rob)` and ln `(mur)` in place of ln `(vio)`.

Table 2: Robbery Rate and Shall-Carry Law

	(1) Pooled OLS (1)	(2) Pooled OLS	(3) Fixed Effects	(4) Fixed Effects and Time Effects
(Intercept)	4.873***	0.904		
	(0.116)	(3.061)		
shall	−0.773***	−0.529**	−0.008	0.027
	(0.225)	(0.161)	(0.055)	(0.052)
incarc_rate		0.001	−0.000	0.000
		(0.001)	(0.000)	(0.000)
density		0.091*	−0.186	−0.045
		(0.046)	(0.166)	(0.196)
avginc		0.041	−0.018	0.014
		(0.028)	(0.022)	(0.025)
pop		0.078***	0.016	0.000
		(0.023)	(0.028)	(0.026)
pb1064		0.102	0.112*	0.014
		(0.089)	(0.051)	(0.083)
pw1064		0.028	0.027	−0.013
		(0.045)	(0.016)	(0.032)
pm1029		0.027	0.011	0.105
		(0.042)	(0.029)	(0.072)
R^2	0.121	0.596	0.037	0.049
Adj. R^2	0.120	0.593	−0.014	−0.021
Num. obs.	1173	1173	1173	1173
RMSE	0.895	0.609		
N Clusters	51	51		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

(d) Repeat the analysis using ln (rob) and ln (mur) in place of ln (vio).

```
pols1 = lm_robust(lmur ~ shall, data = Guns, se_type = "stata", clusters = stateid)
pols2 = lm_robust(lmur ~ shall + incarc_rate + density + avginc +
                    pop + pb1064 + pw1064 + pm1029,
                  data = Guns, se_type = "stata", clusters = stateid)
fe1 = plm(lmur ~ shall + incarc_rate + density + avginc +
            pop + pb1064 + pw1064 + pm1029,
          data = Guns, model = "within", index = c("stateid", "year"))
fe2 = plm(lmur ~ shall + incarc_rate + density + avginc +
            pop + pb1064 + pw1064 + pm1029,
          data = Guns, model = "within", index = c("stateid", "year"),
          effect = "twoway")
SE.pols1 <- pols1$std.error
SE.pols2 <- pols2$std.error
SE.fe1 <- sqrt(diag(vcovHC(fe1, type="sss", cluster="group")))
SE.fe2 <- sqrt(diag(vcovHC(fe2, type="sss", cluster="group")))
p.pols1 <- 2*(1 - pnorm(abs(pols1$coefficients/SE.pols1)))
p.pols2 <- 2*(1 - pnorm(abs(pols2$coefficients/SE.pols2)))
p.fe1 <- 2*(1 - pnorm(abs(fe1$coefficients/SE.fe1)))
p.fe2 <- 2*(1 - pnorm(abs(fe2$coefficients/SE.fe2)))

texreg(list(pols1, pols2, fe1, fe2), include.ci = F, caption.above = T, digits = 3,
       override.se = list(SE.pols1,SE.pols2,SE.fe1,SE.fe2),
       override.pvalues = list(p.pols1, p.pols2, p.fe1, p.fe2),
       caption = "Murder Rate and Shall-Carry Law",
       custom.model.names = c("(1) Pooled OLS (1)", "(2) Pooled OLS",
                              "(3) Fixed Effects", "(4) Fixed Effects & Time Effects"))
```

Table 3: Murder Rate and Shall-Carry Law

	(1) Pooled OLS (1)	(2) Pooled OLS	(3) Fixed Effects	(4) Fixed Effects and Time Effects
(Intercept)	1.898***	−2.486		
	(0.093)	(1.992)		
shall	−0.473**	−0.313**	−0.061	−0.015
	(0.149)	(0.099)	(0.037)	(0.038)
incarc_rate		0.002***	−0.000	−0.000
		(0.000)	(0.000)	(0.000)
density		0.040	−0.671	−0.544
		(0.040)	(0.396)	(0.316)
avginc		−0.077**	0.024	0.057***
		(0.027)	(0.016)	(0.016)
pop		0.042***	−0.026	−0.032
		(0.012)	(0.020)	(0.021)
pb1064		0.131*	0.031	0.022
		(0.061)	(0.078)	(0.075)
pw1064		0.047	0.010	−0.000
		(0.029)	(0.013)	(0.020)
pm1029		0.066	0.039	0.069
		(0.036)	(0.022)	(0.041)
R^2	0.083	0.606	0.153	0.116
Adj. R^2	0.083	0.603	0.109	0.051
Num. obs.	1173	1173	1173	1173
RMSE	0.674	0.443		
N Clusters	51	51		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Tables 2–3 (See pp. 4–5) show the coefficient on `shall` in the regression specifications (1)–(4) using ln (`rob`) and ln (`mur`) as dependent variables, respectively. The quantitative results are similar to the results using violent crimes: there is a large estimated effect of concealed weapons laws in specifications (1) and (2). This effect is spurious and is due to omitted variable bias as specification (3) and (4) show.

(e) In your view, what are the most important remaining threats to the internal validity of this regression analysis?

There is potential two-way causality between this year's incarceration rate and the number of crimes. Because this year's incarceration rate is much like last year's rate, there is a potential two-way causality problem. There are similar two-way causality issues relating crime and shall.

(f) Based on your analysis, what conclusions would you draw about the effects of concealed weapons laws on these crime rate?

The most credible results are given by regression (4). The 95% confidence interval for β_{Shall} is -11.0% to 5.3%. This includes $\beta_{Shall} = 0$. Thus, there is no statistically significant evidence that concealed weapons laws have any effect on crime rates.

Thank you

Francisco Tavares Garcia | Academic Tutor
School of Economics

Reference

Stock, J. H., & Watson, M. W. (2019). Introduction to econometrics (Fourth edition, global edition.). Pearson Education Limited.