PROJECT

BUAN6312: APPLIED ECONOMETRICS AND TIME SERIES ANALYSIS

DO GUNS REDUCE CRIME?

Adityan Rajendran

Mariam Coulibaly

Siddharth Govindarajan

Sushanth Chintalapati

Table of Contents

Introduction

There exists an ever-lasting argument whether public can be allowed to carry a gun or not. There are high chances that the person who has gun license bought for his protection may misuse it. Other set of argument goes along, if everyone has a gun, criminals would be afraid to get near the victim. In this dataset there are a total of 13 columns which falls under the category of panel data and in which there are 3 columns which determine the crimes – robbery, violent and murder rate.

Based on the observation made for 23 years on various parameters on the population on how states had laws whether its legal to carry a gun or not, we are going to determine whether having gun laws have reduced the crime rate or not.

Before building the model from the dataset, we should first completely understand the dataset. So, we have built exploratory data analysis to be precise in the direction through which we will be solving our problem.

About the data

a. Summary of dataset

```
##                types nulls       skew          min       median        max
## year         integer     0  0.00000000 1977.0000000000 1988.00000000 1999.00000
## vio          double      0  2.53512569   47.0000000000  443.00000000 2921.80005
## mur          double      0  5.77842841    0.2000000030    6.40000010   80.60000
## rob          double      0  3.87734779    6.4000000954  124.09999847 1635.09998
## incarc_rate integer      0  3.87674656   19.0000000000  187.00000000 1913.00000
## pb1064       double      0  2.34856883    0.2482065558    4.02621317   26.97957
## pw1064       double      0 -2.22045585   21.7804298401   65.06127930   76.52575
## pm1029       double      0  0.26723726   12.2136802673   15.89516926   22.35269
## pop          double      0  2.42752403    0.4027529955    3.27133203   33.14512
## avginc       double      0  0.73331686    8.5548839569   13.40155125   23.64671
## density      double      0  6.68556704    0.0007070804    0.08156896   11.10212
## stateid      integer     0 -0.01926068    1.0000000000   29.00000000   56.00000
## shall        integer     0  1.19710653    0.0000000000    0.00000000    1.00000
##                     mean
## year        1988.0000000
## vio          503.0746806
## mur            7.6651321
## rob          161.8202044
## incarc_rate  226.5797101
## pb1064         5.3362169
## pw1064        62.9454322
## pm1029        16.0811272
## pop            4.8163414
## avginc        13.7247960
## density        0.3520382
## stateid       28.9607843
## shall          0.2429668
```

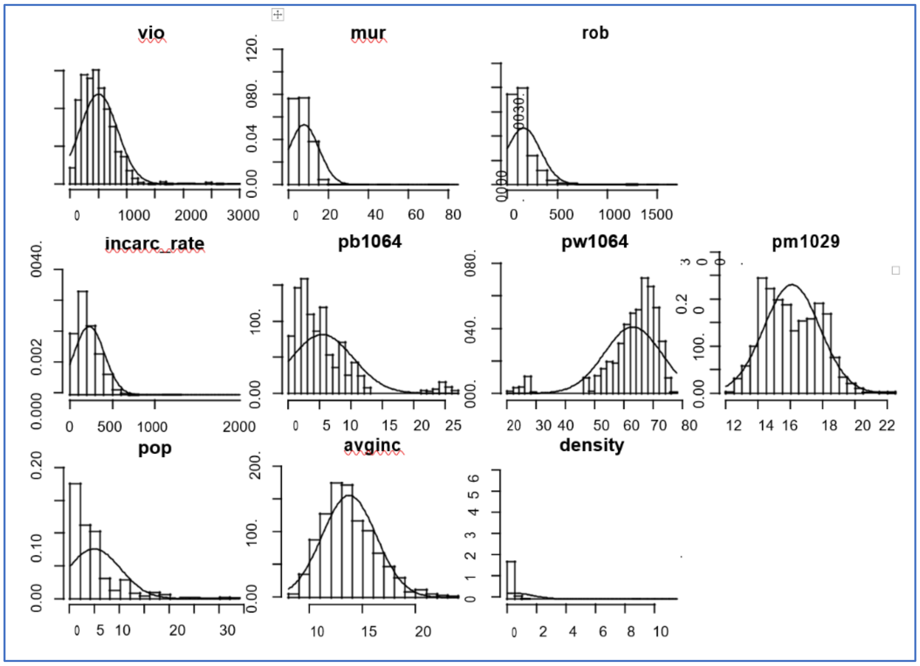
- We don’t have any null values
- All variables are numeric, being either int or double
- Almost all the variables are right or left skewed except ‘avginc’, ‘pm1029’.

b. Histogram of dataset

Based on the figure below, these are the conclusions :-

- Violence is right skewed.
- Murder is right skewed
- Robbery is right skewed.
- Incarceration is right skewed.
- Pb1064 is right skewed
- Pw1064 is left skewed.
- Pm1029 is normal distributed.

- Pop is right skewed.
- Avginc is normal distributed.
- Population density is right skewed.



c. The states with Shall through out

Among the 51 states we have in our dataset, these are the states which always had the shall laws :-

- 18
- 33
- 50
- 53

d. The states without Shall through out

Among the 51 states we have in our dataset, these are the states which never had the shall laws :-

- 1
- 6
- 8
- 9
- 10
- 11
- 15
- 17
- 19
- 20
- 24
- 25
- 26

- 27
- 29
- 31
- 34
- 35
- 36
- 39
- 44
- 55

e. Maximum and Minimum crime rates before and after the shall act

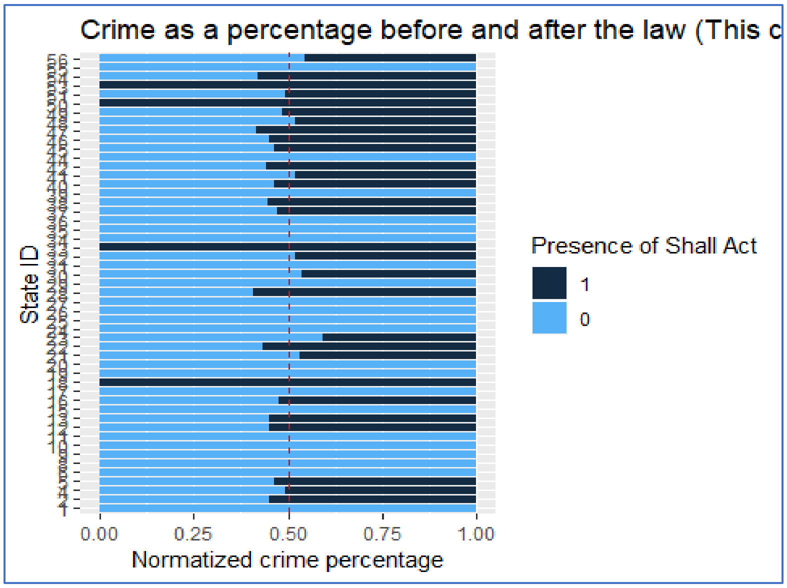
Maximum crime rate before and after the law

stateid	shall	maximum crime rate
12	WithAct	1244.3
11	WithoutAct	2921.8

Minimum crime rate before and after the law

stateid	shall	minimum crime rate
38	WithAct	51.3
38	WithoutAct	47.0

f. Crime as a percentage before and after the shall law

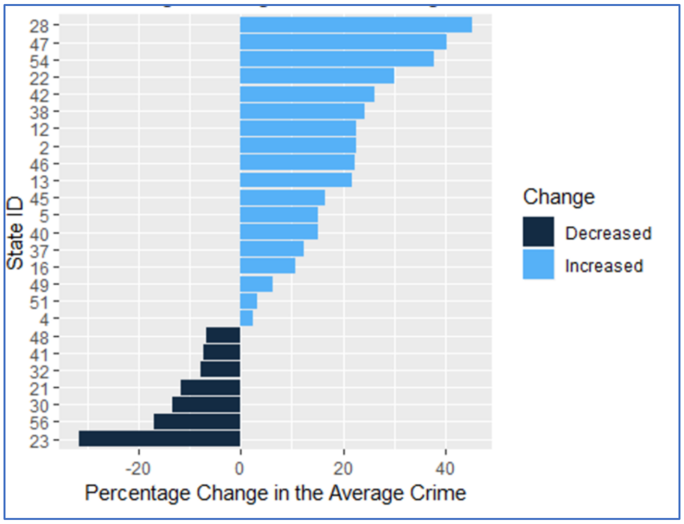


g. Effect on Crime after the Shall Act

After introducing the shall laws to the 25 states which didn't have it before, what happened to the crimes? 7 states had crimes decreased and 18 states had increase in crimes after the introducing of shall laws, but it doesn't necessarily mean that shall laws are the reason as there are many other reasons which contribute to increase in crimes.

Effect of shall laws on crimes	Number of states.
Decreased	7
Increased	18

h. The percentage change in the crime after the introduction of the Shall Act



- 11 had the highest crime rate throughout and didn't introduce the Act.
- Though the crime appears to have decreased with the introduction of shall act, it actually increased in 18 states while it only decreased in 7 states.
- 28 has the highest increase while 23 has the highest decrease in the crime rate with the introduction of Shall Act.

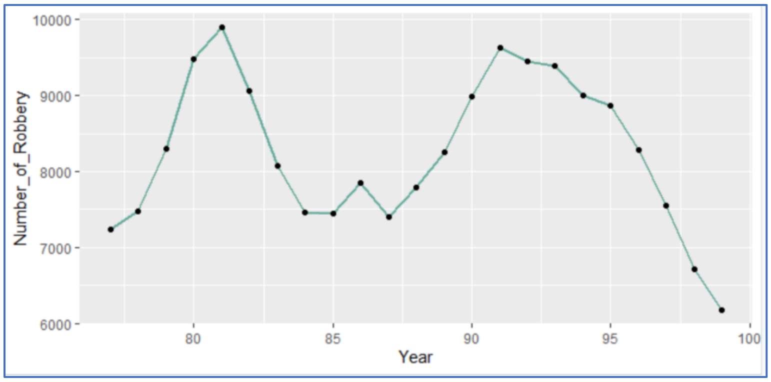
Part I : Exploratory data analysis

a. Aggregation of crimes in the entire nation

1. How the robberies happened in USA all 51 states over the 23 years?

Year wise robberies were aggregated in the data set and random walk has been observed as there is ups and downs with no specific pattern. According the federal reserve history source we researched on, it reported that year 1981 – 1982 had the worst American economic recession after the 1930's of great American depression. When there is an economic recession, people are starving for food and so we see that is the reason why it shows the maximum robberies at that year. The other one was in the year 1991 when there was a large count of robberies and according to Bancroft Berkley this was the time America was facing an economic depression. This depression lasted 8 months in 1990s for 8 months through March 1991.

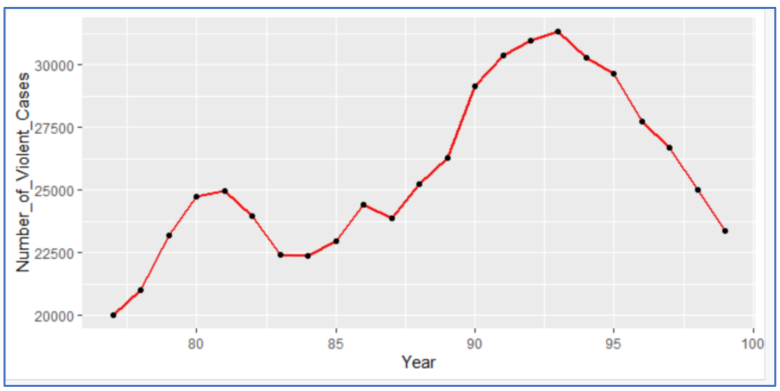
According to value walk, When looked at the year 1985 for instance where the crime was low, value walk states that there was 175,000 private sector jobs created at that time and so there is a mild doubt whether there is an increase in private sector jobs are making the robberies down and economic recession is causing the robberies rate to boom.



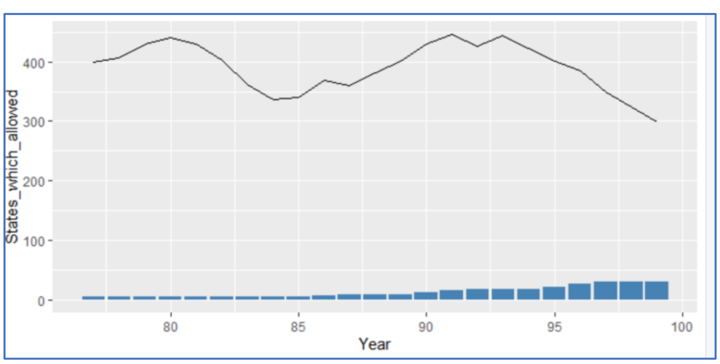
2. **How are the Murder Distribution in USA over 23 years?**

From the year 1977 to 1980, the murder rate has been increasing and after that there is a bit downfall and again a surge until 1993 and it has been falling since then.

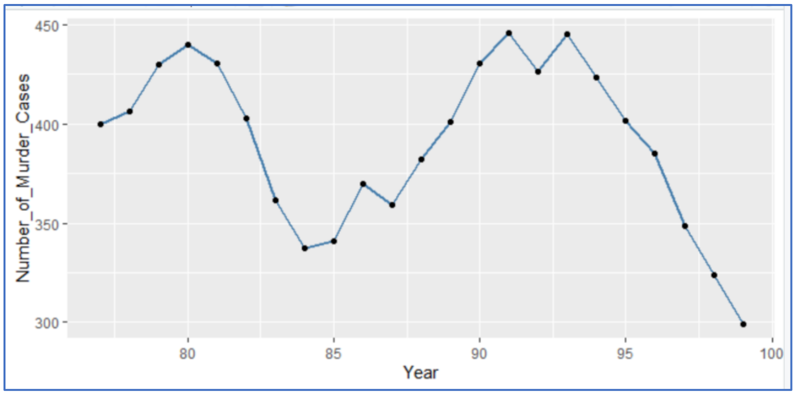
Year 1993 has been observed as the period in which maximum crimes have occurred in USA and after that many of the states have started allowing public carry guns.



To determine whether this rapid decrease in murders after the year 1993 is due to the fact that state allowed guns for the public, lets see the plot below. The bar chart below states the states which have allowed to carry guns and the line chart represents the number of murders. After the year 1993 when many states have started allowing public to carry guns, it is seen that the murders are decreasing.



3. Violent crimes Distribution over 20 years of these crimes.

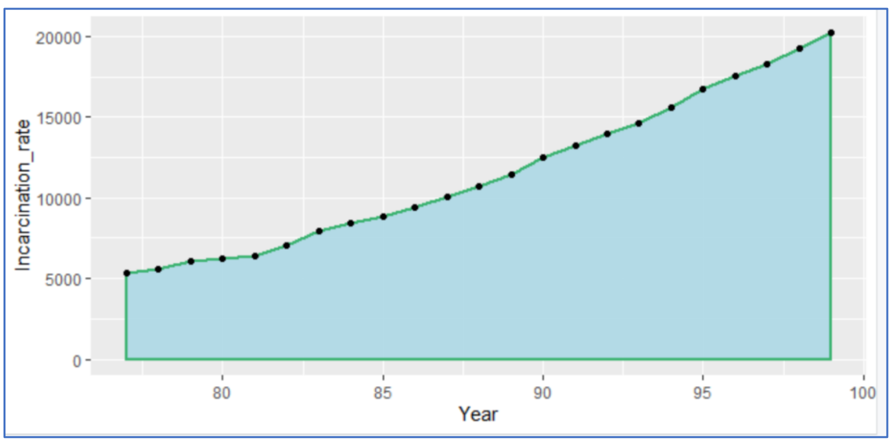


There is a bit of similarity between the pattern shown in violent crimes and murders. There is a fall in violent crimes after the year 1993 just like the murders rate.

4. Incarceration vs Crime rate!

The rate of incarceration has been increasing over the years and in 1999 it has touched close to 20,000 people incarcerated per year and the trend has never been decreasing over these years. Despite the murders, violent and robberies decreasing over the years as we observed in previous cases, it is still seen that the incarceration rate is still increasing over the years.

It contradicts our previous findings as when crimes are low, the rate of incarceration should decrease but it's not.



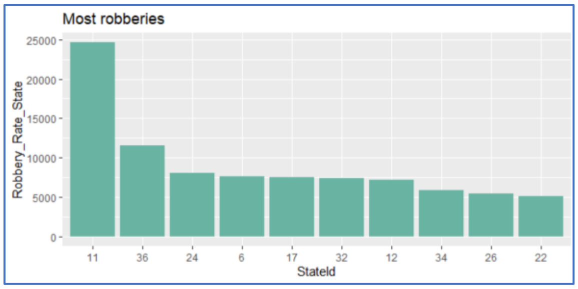
b. Which state has the least and most of the crimes overall 23 years?

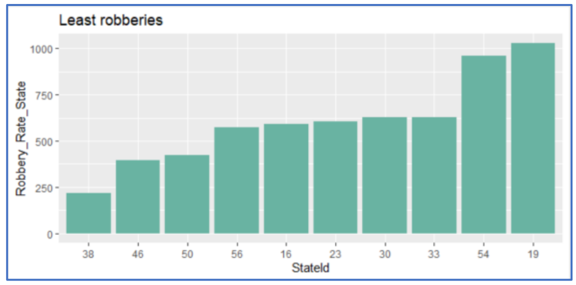
So far we aggregated the crimes nation wise and now let us see which states are having most of the above 3 crimes.

The state wise aggregation of the entire 23 years of crimes has been plotted and it has been observed the stated 11 is having the most number of robberies, murders and violent crimes over almost 2 decades.

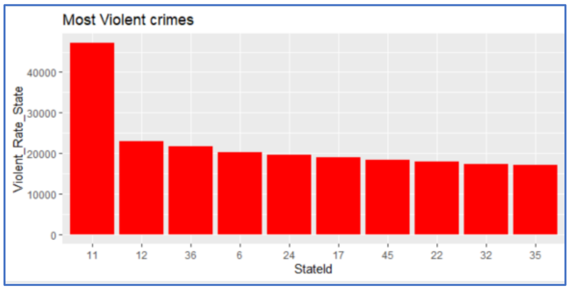
The 3 graphs on the right shows the least crimes which are happening over the years and the state id 38 is having the least number of robberies, murder and violent crimes over these years.

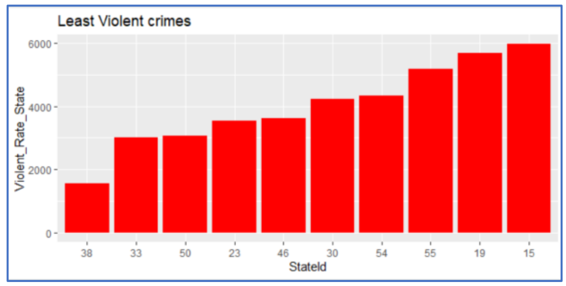
<u>Robberies</u>



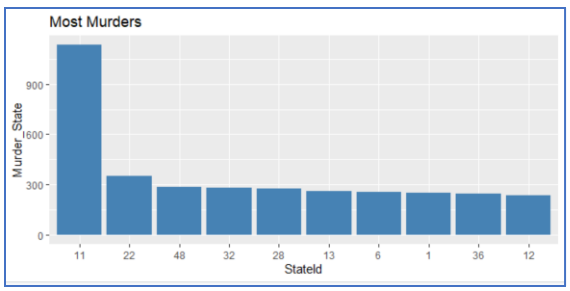


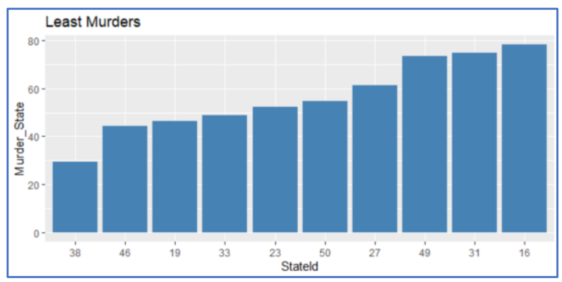
Violent crimes





Murder





So one question arises in our mind right now. *What is the shall laws for the state 11 and state 38?*

It is seen that the state 11 has never had a gun law for the past 23 years. Based on our previous observations we stated that having the shall laws brought down the crimes and in this state, there are no shall laws and crime rates are at its peak.

```
> table(guns_state11$shall)

 0
23
>
```

Before getting to our conclusion let us see about the state 38 – which is the most safest among all other states. The state didn't allow people to carry until the year 1985 and after that it allowed people to carry guns. Since the shall laws were allowed, may be that's the reason why much crimes didn't happen in that state 38.

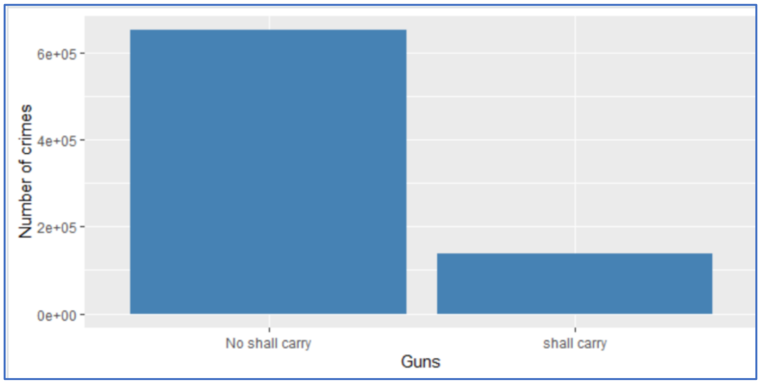
```
> table(guns_state38$shall)

 0  1
 9 14
>
```

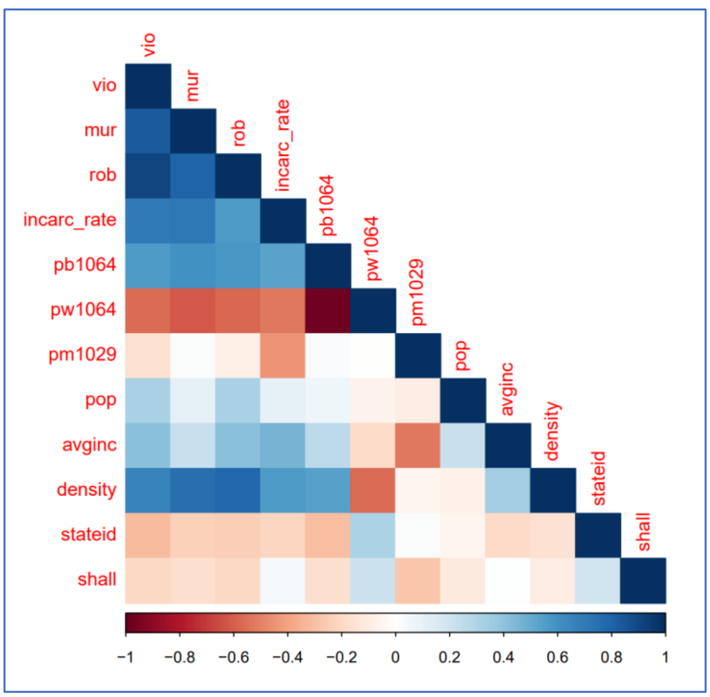
c. How are the crimes when the shall law is 0 and shall law is 1?

Among the 1173 scenarios that we have in our dataset for 51 states * 23 years = 1173, it was seen that 888 times state laws didn't allow shall and 285 times shall laws were issued.

Aggregating the robberies, murders and violent crimes in shall laws issued scenarios and not shall law situations, graph is represented below to bring it to common scale. It is seen that the crime rates are high when the state didn't allow people to carry guns.



d. Correlation plot of dataset



- 'rob' and 'vio' are the only variables with high correlation (0.91)
- 'density' seems to be correlated with 'mur' and 'rob'(0.7 - 0.8)
- States which have implemented Shall-issue laws at some point of time: 2, 4, 5, 12, 13, 16, 21, 22, 23, 28, 30, 32, 37, 38, 40, 41, 42, 45, 46, 47, 48, 49, 51, 54, 56
- States which have never implemented Shall-issue laws: 1, 6, 8, 9, 10, 11, 15, 17, 19, 20, 24, 25, 26, 27, 29, 31, 34, 35, 36, 39, 44, 55

- States with shall-issue laws from the start(1977): 18, 33, 50, 53

These are the questions which arose based on EDA's

- Are Crimes having something to do with each state(entity), year and shall laws?
- Are Shall laws having some impact on the crimes?

To build a model statistically, we have to determine whether any of the coefficient are having an impact on our model and check for entity fixed, time fixed effects on the model.

Part 2 : Building models and hypothesis

For the entire model, we have used violent crimes as the dependent variable and all other variables except robbery and murder as independent variables.

a) <u>Signs expected</u>

1. Crimes should decrease if incarceration rate increases as people would be afraid to do crimes.
2. Having the certain percentage of whites and black can cause raise or fall of crimes.
3. Increase in population, more should be the crimes.
4. More the average income, lesser should be the crimes.
5. Increase in population density, more should be the crimes.
6. Having shall laws, should decrease the crime rate.

NOTE: For point 1, we can have a simultaneous causality bias wherein an increase in crime rate can lead to an increase in incarceration rate. This unobserved heterogeneity that is introducing a bias in the model will looked at later as we proceed in the model implementation.

b) <u>F-Test to determine whether at least one of the coefficients are significant or not</u>

We need to test violence(vio) as dependent variable having independent variables: incarc_rate, pb1064, pw1064, pm1029, pop, avginc, density, shall.

Hypothesis

- NULL hypothesis : Beta coefficient of incarc_rate, pb1064, pw1064, pm1029, pop, avginc, density, shall is zero(0)
- Alternate hypothesis : At least one of the Beta coefficient of incarc_rate, pb1064, pw1064, pm1029, pop, avginc, density, shall is not equal to zero(0)

The F-stat critical value for degree of freedom (8; 1164) is 1.9384. The F-stat obtained is 381.1. Since the F-stat is greater than critical value, at least one of the coefficient is significant.

```
Call:
lm(formula = vio ~ incarc_rate + pb1064 + pw1064 + pm1029 + pop +
    avginc + density + shall, data = guns_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1035.31  -109.36   -27.61   101.39   659.47

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -178.77738  224.01039  -0.798   0.4250
incarc_rate    0.81559    0.04417  18.463  < 2e-16 ***
pb1064        11.29196    6.86442   1.645   0.1002
pw1064         2.70238    3.45360   0.782   0.4341
pm1029         9.31654    4.44137   2.098   0.0361 *
pop           18.50567    1.05486  17.543  < 2e-16 ***
avginc         1.23476    3.20734   0.385   0.7003
density       94.66882    5.42844  17.439  < 2e-16 ***
shall        -92.72125   13.42570  -6.906 8.17e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 176.3 on 1164 degrees of freedom
Multiple R-squared:  0.7237,    Adjusted R-squared:  0.7218
F-statistic: 381.1 on 8 and 1164 DF,  p-value: < 2.2e-16
```

This is a two-tail test and we have taken 95% confidence interval to see which of the coefficient is significant and which is not. T(0.95/2) = (-1.96, 1.96) is the critical values.

Variable	t-value
incarc_rate	18.463
pb1064	1.645
pw1064	0.782
pm1029	2.098
pop	17.543
avginc	0.385
density	17.439
shall	-6.906

The significant values are denoted above in red color.

c) Running Pooled OLS as log – linear function in STATA

```
. *pooled OLS model

. generate lvio=ln(vio)
variable lvio already defined
r(110);

. regress lvio i.shall incarc_rate pb1064 pm1029 pop avginc
pw1064 density
       Source |       SS           df       MS      Number of obs   =
1,173
-------------+----------------------------------   F(8, 1164)      =
188.41
        Model |  275.712977         8  34.4641221   Prob > F        =
0.0000
     Residual |  212.918581     1,164  .182919743   R-squared       =
0.5643
-------------+----------------------------------   Adj R-squared   =
0.5613
        Total |  488.631558     1,172  .416921125   Root MSE        =
.42769
--------------------------------------------------------------------------
-----
```

```
        lvio |      Coef.   Std. Err.      t    P>|t|      [95% Conf.
Interval]
-------------+----------------------------------------------------------
-----
     1.shall |  -.3683869   .0325674   -11.31   0.000    -.4322844   -
.3044895
  incarc rate |   .0016126   .0001072    15.05   0.000     .0014024
.0018229
      pb1064 |   .0808526   .0166514     4.86   0.000     .0481825
.1135227
      pm1029 |   .0088709   .0107737     0.82   0.410    -.0122671
.0300089
         pop |   .0427098   .0025588    16.69   0.000     .0376894
.0477303
      avginc |   .0012051   .0077802     0.15   0.877    -.0140597
.01647
      pw1064 |   .0312005   .0083776     3.72   0.000     .0147636
.0476374
     density |   .0266885    .013168     2.03   0.043     .0008527
.0525242
       cons |   2.981738   .5433938     5.49   0.000     1.915598
4.047879
------------------------------------------------------------------------
-----
```

Comments:

- All the variables are significant except pm1029 and avginc. We are more interested with the variable 'shall' since we need to know the effect of shall on crime rate to respond to our study question.
- The sign expected is negative as the violent crime rate is supposed to decrease in a state with shall. We have a negative sign as expected with a decrease by 36.83 % of crime rate with a 95% confidence that the decrease will fall between 30.44 and 43.22% in state with shall.
- However, we need to check about heteroskedasticity to make sure that our model doesn't suffer of any bias and is correct.

d) <u>Pooled OLS with Cluster-Robust Standard Errors in R</u>

```
##
## Call:
## plm(formula = vio ~ incarc_rate + pb1064 + pw1064 + pm1029 +
##        pop + avginc + density + shall, data = data, model = "pooling",
##        index = c("stateid", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##          Min.       1st Qu.     Median      3rd Qu.       Max.
## -1035.308        -109.361   -27.614       101.385   659.474
##
## Coefficients:
##                        Estimate          Std. Error t-value                    Pr(>|t|)
## (Intercept) -178.777378        224.010387   -0.7981                         0.42499
## incarc_rate          0.815595          0.044174  18.4632 < 0.00000000000000022 ***
## pb1064             11.291962          6.864417   1.6450                         0.10024
## pw1064              2.702383          3.453603   0.7825                         0.43409
## pm1029              9.316543          4.441374   2.0977                         0.03615 *
## pop                18.505673          1.054859  17.5433 < 0.00000000000000022 ***
## avginc              1.234759          3.207344   0.3850                         0.70032
## density            94.668823          5.428439  17.4394 < 0.00000000000000022 ***
## shall             -92.721249         13.425702  -6.9062         0.000000000008168 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:130960000
## Residual Sum of Squares: 36184000
## R-Squared:0.7237
## Adj. R-Squared: 0.7218
## F-statistic: 381.103 on 8 and 1164 DF, p-value: < 0.000000000000000222
```

Running the pooled OLS with cluster robust standard errors gives us the following interpretations:

- Standard errors are significantly different from errors obtained from pooled OLS model without robust errors.
- This confirms the presence of serial correlations and heteroskedasticity
- All the estimates are insignificant at alpha = 5% now because of the corrected bigger standard errors.(except for 'shall')
- Although it can be assumed that the estimates are now right, pooled OLS with robust standard errors doesn't account for endogeneity in the model
- We can test for endogeneity to see if it exists in the model
- The test of endogeneity can be done by running a Hausman test with 'Fixed Effects' and 'Random Effects' model estimates. If they are found to be significantly different, then we can safely conclude the presence of Endogeneity.

e) <u>Why we think pooled OLS is not good?</u>

- We can see that the estimate for 'shall' is negative and the interpretation is that for states with shall issue laws the violent crime rate is 92 units lower than the crime rate for states with no shall issue laws.(units = incidents per 100,000 members of the population)

- We know that pooled OLS has an overstated reliability in the presence of serial correlation and heteroskedasticity.

f) <u>Detecting heteroskedasticity</u>:

Why do we need to detect heteroskedasticity in our model?

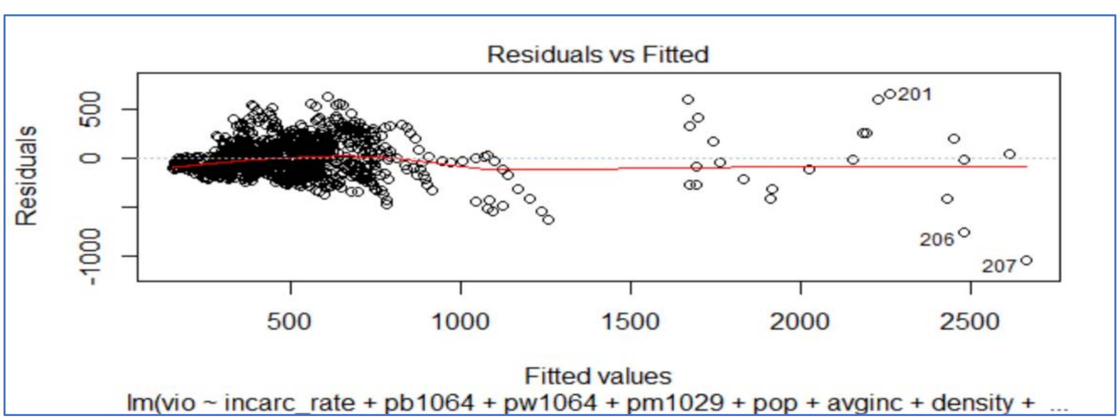
If we build a model ignoring heteroskedasticity:

- The model is going to unbiased and consistent, but inefficient.
- The standard errors and confidence interval are going to be wrong and hypothesis tests will lead us to a wrong prediction.

How to detect?

i. Plotting independent variables against residuals
ii. Using statistical tests

i. Plotting independent variables against residuals



According to this plot, we detect heteroskedasticity but let us confirm it 100% by performing a proper test.

ii. Using statistical tests

- Multiplying R^2 with N gives 0.2296 * 1173 = 269.32.
- The critical value is (0.95, 17) = 124.342

Since the value falls in the critical value we reject NULL hypothesis and say that heteroskedasticity exists in our model and perform the following.

iii. White's test in STATA – log linear function

- White's test for Ho: homoskedasticity
- against Ha: unrestricted heteroskedasticity

Hypothesis

- White's test for Ho: homoskedasticity
- against Ha: unrestricted heteroskedasticity

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	454.02	43	0.0000
Skewness	107.86	8	0.0000
Kurtosis	4.22	1	0.0399
Total	566.10	52	0.0000

Linear regression	
Number of obs =	1,173
F(8, 1164) =	95.67
Prob > F =	0.0000
R-squared =	0.5643
Root MSE =	.42769

lvio	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.shall	-.3683869	.0347879	-10.59	0.000	-.436641	-.3001329
incarc_rate	.0016126	.0001807	8.92	0.000	.0012581	.0019672
pb1064	.0808526	.0199924	4.04	0.000	.0416274	.1200778
pm1029	.0088709	.0120604	0.74	0.462	-.0147917	.0325334
pop	.0427098	.0031466	13.57	0.000	.0365361	.0488836
avginc	.0012051	.0072778	0.17	0.869	-.013074	.0154842
pw1064	.0312005	.0097271	3.21	0.001	.012116	.0502851
density	.0266885	.0143494	1.86	0.063	-.0014651	.054842
_cons	2.981738	.6090198	4.90	0.000	1.786839	4.176638

Comments:

- We can see that the new model has greater standards errors than the previous one and sort a new insignificant variable 'density'.
- Since we are working on cross sections which are states, we are face to unobserved heterogeneity. unobserved individuality. Pooled OLS in this case cannot give us correct estimators. Fixed effects model is used to control unobserved heterogeneity.

g) <u>Fixed Effects Model - Entity Fixed (R)</u>

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = fe_form, data = data, effect = "individual", model = "within",
##     index = c("stateid", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##        Min.      1st Qu.    Median     3rd Qu.      Max.
## -786.3960       -47.6737   -1.4221     44.9161  772.5089
##
## Coefficients:
##                    Estimate     Std. Error t-value          Pr(>|t|)
## incarc_rate        0.096788      0.057367    1.6872        0.0918515 .
## pb1064            12.808031     10.882660    1.1769        0.2394784
## pw1064            10.323253      3.110094    3.3193        0.0009317 ***
## pm1029           -23.863850      3.924761   -6.0803 0.000000001645 ***
## pop               12.244061      5.346751    2.2900        0.0222076 *
## avginc            -4.066645      3.621108   -1.1230        0.2616630
## density         -155.953250     52.117608   -2.9923        0.0028295 **
## shall            -18.575084     11.563239   -1.6064        0.1084714
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:13576000
## Residual Sum of Squares: 10809000
## R-Squared:0.20381
## Adj. R-Squared: 0.16235
## F-statistic: 35.6447 on 8 and 1114 DF, p-value: < 0.000000000000000222
```

Running the Fixed Effects model with Entity Fixed effects, we have the following takeaways:

- The model is significant with an F-statistic of 35.6447 and R-sq = 20.3%
- Estimates for 'pop', 'avginc', 'incarc_rate' are found insignificant at alpha=5%
- Interpreting the estimate for 'shall', for states with shall-issue laws, the crime rate is 4.6% lesser than the crime rate in states with no shall-issue laws

Log – linear function STATA

```
. *Entity Fixed Effect model
|
. xtset stateid
       panel variable:  stateid (balanced)

. xtreg lvio i.shall density avginc pop pm1029 pw1064 pb1064 incarc_rate,
fe cl
> uster (stateid)

Fixed-effects (within) regression               Number of obs      =
1,173
Group variable: stateid                         Number of groups   =
51

R-sq:                                           Obs per group:
     within   = 0.2178                                        min =
23
     between = 0.0033                                         avg =
23.0
     overall = 0.0001                                         max =
23

                                                F(8,50)            =
34.10
corr(u_i, Xb)    = -0.3687                      Prob > F           =
```

```
0.0000

                                 (Std. Err. adjusted for 51 clusters in
stateid)
------------------------------------------------------------------------------
-----
             |               Robust
        lvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-----
     1.shall |  -.0461415   .0417616    -1.10   0.275    -.1300223
.0377392
     density |  -.1722901   .1376129    -1.25   0.216    -.4486936
.1041135
      avginc |  -.0092037   .0129649    -0.71   0.481    -.0352445
.016837
         pop |   .0115247    .014224     0.81   0.422    -.0170452
.0400945
      pm1029 |  -.0502725   .0206949    -2.43   0.019    -.0918394    -
.0087057
      pw1064 |   .0408611   .0134585     3.04   0.004     .0138289
.0678932
      pb1064 |   .1042804   .0326849     3.19   0.002     .0386308
.1699301
 incarc_rate |   -.000071   .0002504    -0.28   0.778    -.0005739
.0004318
       _cons |   3.866017   .7701057     5.02   0.000     2.319214
5.412819
-------------+----------------------------------------------------------------
-----
     sigma_u |  .68024951
     sigma_e |  .16072287
         rho |  .94712779   (fraction of variance due to u_i)
```

Comments:

- With this model most of the variables are insignificant we only have 3 significant variables pm1029, pw1064, pb1064. Our variable of interest 'shall' is insignificant at 10 % significance level with an estimator of -0.04
- Showing that a state with 'shall' has a decrease of 4% of the violent crime rate. this result is significantly lower than the pooled OLS estimate.as we suspected pooled OLS are positively biased and the bias is quite big.
- **transition: as the omitted variable can vary across entities, it can also vary over time and not across entities.in this case we need to use the time fix ed effect model. In our case, we will run a entity and time fixed effects model.
- Estimates for ‘pb1064’, ‘avginc’, ‘shall’ are found insignificant at alpha=5%

- Interpreting the estimate for ‘shall’, for states with shall-issue laws, the number of crime incidents per 100k members is 18 lesser than the number of crime incidents per 100k members in states with no shall-issue laws.

h) Fixed Effects Model - Entity Fixed + Cluster Robust Standard Errors (RStudio)

Fixed Effects Model (Entity Fixed) + Robust Errors

```
##
## t test of coefficients:
##
##               Estimate   Std. Error t value Pr(>|t|)
## incarc_rate -0.000071008  0.000247883 -0.2865 0.774581
## pb1064       0.104280401  0.032362813  3.2222 0.001309 **
## pw1064       0.040861070  0.013325836  3.0663 0.002219 **
## pm1029      -0.050272536  0.020490909 -2.4534 0.014303 *
## pop          0.011524661  0.014083855  0.8183 0.413367
## avginc      -0.009203729  0.012837107 -0.7170 0.473547
## density     -0.172289988  0.136256609 -1.2645 0.206332
## shall       -0.046141523  0.041350058 -1.1159 0.264716
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comments:

- Though we have run a Fixed Effects model, we run the Fixed Effects model with Robust Standard errors to account for heteroskedasticity and serial correlation to get the correct standard errors
- We see that, the estimates for ‘density’ and ‘shall’ are no more significant in addition to the variables that were earlier deemed insignificant (‘pop’, ‘avginc’, ‘incarc_rate’)
- Since only standard errors are corrected, the coefficient estimates remain the same.

i) Fixed Effects Model - Time Fixed (RStudio)

```
## Oneway (time) effect Within Model
##
## Call:
## plm(formula = fe_form, data = data, effect = "time", model = "within",
##     index = c("stateid", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##      Min.    1st Qu.    Median  3rd Qu.     Max.
## -1047.669   -104.863   -23.885   97.618   689.040
##
## Coefficients:
##              Estimate Std. Error t-value              Pr(>|t|)
## incarc_rate  0.971689   0.045565  21.3251 < 0.00000000000000022 ***
## pb1064      14.867441   7.268267   2.0455             0.0410316 *
## pw1064       4.278799   3.637759   1.1762             0.2397528
## pm1029      -9.149798   6.983295  -1.3102             0.1903778
## pop         17.486780   1.006259  17.3780 < 0.00000000000000022 ***
## avginc       9.163959   3.173222   2.8879             0.0039513 **
## density     76.774741   5.560903  13.8062 < 0.00000000000000022 ***
## shall      -49.962260  13.426401  -3.7212             0.0002079 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:126260000
## Residual Sum of Squares: 31675000
## R-Squared:0.74912
## Adj. R-Squared: 0.74253
## F-statistic: 426.247 on 8 and 1142 DF, p-value: < 0.000000000000000222
```

Comments:

- Running a Time Fixed effect model, we have the below interpretations:
- The model is highly siginificant with an F-statistic of 426.247 and R-sq = 74.9%
- The estimates of all the variables are significant with alpha=5% save 'pw1064' and 'pm1029'
- The estimate for 'shall' is -49 which seems realistic and reasonable
- Interpreting the estimate, for states with shall-issue laws, as compared to states with no shall-issue laws, the no. of crime incidents per 100k members is 49 lesser.
- Other coefficient estimates are also reasonable and make sense

STATA – as log linear

```
. regress lvio 1.shall density avginc pop pm1029 pw1064 pb1064 incarc_rate i.year,
vce(robust)

Linear regression                    Number of obs     =      1,173
                                     F(30, 1142)       =      31.88
                                     Prob > F          =     0.0000
                                     R-squared         =     0.5922
                                     Root MSE          =     .41772

------------------------------------------------------------------------------
             |               Robust
        lvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.shall |  -.2877692   .0368363    -7.81   0.000    -.3600438   -.2154947
     density |  -.0088742   .0185827    -0.48   0.633    -.0453342    .0275858
      avginc |   .0128565   .0080061     1.61   0.109    -.0028518    .0285647
         pop |   .0408235   .0029273    13.95   0.000      .03508     .0465669
      pm1029 |  -.0444208   .0206037    -2.16   0.031    -.0848462   -.0039953
      pw1064 |   .0400798   .0103422     3.88   0.000     .0197881    .0603716
      pb1064 |   .0999892    .021424     4.67   0.000     .0579544     .142024
 incarc_rate |   .0019305   .0002224     8.68   0.000     .0014941     .002367
             |
        year |
         78  |   .0219046   .0772058     0.28   0.777    -.1295765    .1733857
         79  |   .0912726    .077681     1.17   0.240    -.0611409    .2436861
         80  |   .1235947   .0804656     1.54   0.125    -.0342822    .2814716
         81  |    .095128   .0831957     1.14   0.253    -.0681056    .2583616
         82  |   .0247982   .0833665     0.30   0.766    -.1387705    .1883668
         83  |  -.0753135   .0825574    -0.91   0.362    -.2372946    .0866677
         84  |  -.0988051   .0828634    -1.19   0.233    -.2613866    .0637764
         85  |  -.1013227   .0871838    -1.16   0.245    -.2723811    .0697356
         86  |  -.0693822   .0873848    -0.79   0.427    -.2408351    .1020706
         87  |  -.1229694   .0900179    -1.37   0.172    -.2995885    .0536497
         88  |    -.11757   .0955339    -1.23   0.219    -.3050117    .0698717
         89  |  -.1233524    .100415    -1.23   0.220     -.320371    .0736662
         90  |  -.1339978   .1086623    -1.23   0.218    -.3471979    .0792023
         91  |  -.1203355   .1149338    -1.05   0.295    -.3458406    .1051695
         92  |   -.135282   .1178565    -1.15   0.251    -.3665216    .0959576
         93  |  -.1597547   .1223309    -1.31   0.192    -.3997732    .0802637
         94  |  -.2336142   .1274336    -1.83   0.067    -.4836443     .016416
         95  |  -.2854855   .1333553    -2.14   0.033    -.5471344   -.0238367
```

```
         96  |   -.363436   .1395039    -2.61   0.009    -.6371487   -.0897232
         97  |   -.412783   .1423534    -2.90   0.004    -.6920865   -.1334795
         98  |  -.5240673   .1470749    -3.56   0.000    -.8126347     -.2355
         99  |  -.6361456   .1462722    -4.35   0.000     -.923138   -.3491533
             |
       _cons |   3.093731   .6001287     5.16   0.000     1.916253     4.27121
------------------------------------------------------------------------------
```

Comments

- Variables are significant at 5% significance level except time variables from 78 to 94 and density.
- Our variable of interest 'shall' coefficient is -.2877 showing a decrease of 28.77 % of violent crime rate in a state with 'shall' as compared to states with no ‘shall-issue’ laws.
- This is so far the best model, considering the 'shall' variable with the best effect on crime rate and significance.

j) Fixed Effects Model - Time Fixed + Cluster Robust Standard Errors (RStudio)

Fixed Effects Model (Time Fixed) + Robust Errors

```
coeftest(fe.model.tf, vcov = vcovHC(fe.model.tf, type='HC1', cluster='group'))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value  Pr(>|t|)
## incarc_rate  0.00193051  0.00075928  2.5426 0.0111353 *
## pb1064       0.09998919  0.08126541  1.2304 0.2187997
## pw1064       0.04007982  0.03864586  1.0371 0.2999065
## pm1029      -0.04442077  0.07771991 -0.5715 0.5677397
## pop          0.04082345  0.01086395  3.7577 0.0001801 ***
## avginc       0.01285649  0.02837513  0.4531 0.6505700
## density     -0.00887417  0.05697323 -0.1558 0.8762494
## shall       -0.28776925  0.12183579 -2.3619 0.0183465 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comments:

Running a Time Fixed effect model + Cluster Robust errors, we account for heteroskedasticity and have the below interpretations:

- In addition to ‘density’ & ‘avginc’, the variables ‘pm1029’, ‘pw1064’ and ‘pb1064’ are deemed significant at alpha=5%
- The estimate for ‘shall’ is -0.28 which seems realistic and reasonable
- The estimates remain consistent as only standard errors are corrected.
- The ‘shall’ coefficient estimate is reasonable and significant.

k) Fixed Effects model - Entity + Time fixed effects (R)

```
## Twoways effects Within Model
##
## Call:

8

## plm(formula = fe_form, data = data, effect = "twoways", model = "within",
##       index = c("stateid", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##          Min.        1st Qu.       Median        3rd Qu.        Max.
## -636.333085    -42.492632    -0.099103    41.499666  641.548959
##
## Coefficients:
##                      Estimate Std. Error t-value          Pr(>|t|)
## incarc_rate        0.191571    0.054741    3.4996       0.0004847 ***
## pb1064           -36.092871  13.758543   -2.6233       0.0088295 **
## pw1064           -10.936465    4.766677   -2.2944       0.0219590 *
## pm1029            47.171050    9.466965    4.9827  0.0000007286 ***
## pop                2.057769    4.770199    0.4314       0.6662773
## avginc             6.887878    3.901603    1.7654       0.0777765 .
## density          -95.203341  46.251071   -2.0584       0.0397883 *
## shall             -0.839059  10.403066   -0.0807       0.9357312
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:8871100
## Residual Sum of Squares: 7871200
## R-Squared:0.11272
## Adj. R-Squared: 0.047716
## F-statistic: 17.3406 on 8 and 1092 DF, p-value: < 0.000000000000000222
```

Running a fixed effects model with both time and entity fixed effects gives the above results and is subject to following interpretations:

- The model is significant with an F-statistic of 17.34 and R-sq = 11%
- Estimates for all the variables are significant at alpha = 10% except for 'shall'
- This is not an accurate model as all estimates are insignificant

STATA

```
. xtreg lvio i.shall density avginc pop pm1029 pw1064 pb1064
incarc_rate i.year
> , fe vce(cluster stateid)

Fixed-effects (within) regression               Number of obs      =
1,173
Group variable: stateid                         Number of groups   =
51

R-sq:                                           Obs per group:
     within  = 0.4180                                          min =
23
     between = 0.0419                                          avg =
23.0
     overall = 0.0009                                          max =
23

                                                F(30,50)           =
56.86
corr(u_i, Xb)  = -0.2929                        Prob > F           =
0.0000

                              (Std. Err. adjusted for 51 clusters in
stateid)
--------------------------------------------------------------------------
-----
              |               Robust
         lvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
```

```
-----------------------------
      1.shall |  -.0279935   .0407168    -0.69   0.495    -.1097757
.0537886
      density |   -.091555   .1238622    -0.74   0.463    -.3403396
.1572296
       avginc |   .0009587   .0164931     0.06   0.954    -.0321688
.0340861
          pop |  -.0047544   .0152294    -0.31   0.756    -.0353436
.0258347
       pm1029 |   .0733254   .0524733     1.40   0.168    -.0320704
.1787211
       pw1064 |   .0092501   .0237564     0.39   0.699    -.0384659
.0569662
       pb1064 |   .0291862   .0495407     0.59   0.558    -.0703192
.1286916
  incarc_rate |    .000076   .0002079     0.37   0.716    -.0003416
.0004935
              |
         year |
           78  |   .0585261   .0161556     3.62   0.001     .0260767
.0909755
           79  |   .1639486   .0244579     6.70   0.000     .1148233
.2130738
           80  |   .2170759   .0334184     6.50   0.000     .1499531
.2841987
           81  |   .2172551   .0391956     5.54   0.000     .1385284
.2959819
           82  |   .1946328   .0465743     4.18   0.000     .1010856
.28818
           83  |    .158645   .0593845     2.67   0.010     .0393676
.2779223
           84  |   .1929883   .0770021     2.51   0.015     .0383251
.3476515
           85  |   .2444764   .0922217     2.65   0.011     .0592438
.4297091
           86  |   .3240904   .1089181     2.98   0.004     .1053219
.5428589
           87  |    .324365   .1249881     2.60   0.012      .073319
.5754111
           88  |   .3867412   .1397074     2.77   0.008     .1061305
.6673518
           89  |   .4422143   .1535358     2.88   0.006     .1338286
.7505999
           90  |   .5430478   .1960859     2.77   0.008     .1491976
.936898
           91  |   .5959456   .2040685     2.92   0.005     .1860618
1.005829
           92  |   .6275171   .2170306     2.89   0.006     .1915982
1.063436
           93  |   .6497414   .2246177     2.89   0.006     .1985834
1.100899
           94  |   .6354187   .2332437     2.72   0.009     .1669349
1.103903
```

```
           95  |   .6276831   .2423607     2.59   0.013     .1408874
1.114479
           96  |   .5713423   .2534067     2.25   0.029       .06236
1.080325
           97  |   .5501153   .2613516     2.10   0.040     .0251751
1.075055
           98  |   .4932904   .2746546     1.80   0.079    -.0583697
1.04495
           99  |   .4328776   .2862197     1.51   0.137    -.1420117
1.007767
               |
          cons |   3.765525   1.152108     3.27   0.002     1.451448
6.079603
-----------------------------
       sigma_u |  .6663043
       sigma_e |  .1400264
           rho |  .95770338   (fraction of variance due to u_i)
--------------------------------------------------------------------------
-----
```

Comments

- All the time variables are significant with an increasing effect
- over time on violent crime rate and slightly decreasing late 90. however, the other variables are insignificant with a 10% significance level.
- the estimator 'shall' in this model still has a negative sign with a coefficient of - 0.027 showing that a state with shall has a decrease in violent crime rate of 2.7%
- quite lower effect than the one in entity fixed model. with this model we can say that we removed most source of biases, we can expect a quite small bias.

l) Fixed Effects model - Entity + Time fixed effects + Cluster Robust Standard Errors (RStudio)

Fixed Effects model (Entity + Time) + Cluster Standard Errors

```
coeftest(fe.model.eftf, vcov = vcovHC(fe.model.eftf, type='HC1', cluster='group'))
```

```
##
## t test of coefficients:
##
##                Estimate   Std. Error t value Pr(>|t|)
## incarc_rate  0.000075995  0.000203884  0.3727   0.7094
## pb1064       0.029186170  0.048586680  0.6007   0.5482
## pw1064       0.009250117  0.023298878  0.3970   0.6914
## pm1029       0.073325371  0.051462838  1.4248   0.1545
## pop         -0.004754443  0.014936144 -0.3183   0.7503
## avginc       0.000958650  0.016175537  0.0593   0.9528
## density     -0.091554914  0.121476997 -0.7537   0.4512
## shall       -0.027993527  0.039932732 -0.7010   0.4834
```

Comments

Running a Fixed Effects model (Time + Entity) with Cluster Robust Standard Errors is subject to interpretations:

- All estimates are now deemed insignificant with corrected standard errors.
- This model, on the whole, is deemed insignificant as all estimates are insignificant.

m) Random effects model

```
## Oneway (individual) effect Random Effect Model
##     (Swamy-Arora's transformation)
##
## Call:
## plm(formula = re_form, data = data, model = "random", index = c("stateid",
##       "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Effects:
##                       var    std.dev share
## idiosyncratic     9703.24     98.51 0.354
## individual       17694.57    133.02 0.646
## theta: 0.8474
##
## Residuals:
##        Min.   1st Qu.Median   3rd Qu.Max.
```

```
## -609.2264      -52.5582       -7.2925   44.6080    842.6913
##
## Coefficients:
##                          Estimate Std. Error z-value                    Pr(>|z|)
## (Intercept) -54.701888 231.127874 -0.2367                             0.812910
## incarc_rate        0.393153      0.039789    9.8808 < 0.0000000000000022 ***
## pb1064           20.610769      7.196501   2.8640                     0.004183 **
## pw1064             8.026554      3.093081   2.5950                     0.009459 **
## pm1029            -9.640604      3.623266  -2.6607                     0.007797 **
## pop               13.158514      3.074760   4.2795          0.00001872924799 ***
## avginc            -6.025157      3.514785  -1.7142                     0.086486 .
## density         104.716646     15.978408   6.5536          0.00000000005615 ***
## shall            -37.193442     11.602500  -3.2056                     0.001348 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:16310000
## Residual Sum of Squares: 12145000
## R-Squared:0.25536
## Adj. R-Squared: 0.25025
## Chisq: 399.177 on 8 DF, p-value: < 0.000000000000000222
```

Comments

- Though running a 'Random Effects' model doesn't make sense because we don't have random samples, we run it to understand if endogeneity is present in the model.
- The estimates of the RE model is compared with the estimates of the FE model using an Hausman test wherein the null hypothesis states no endogeneity and the alternate hypothesis roots for the presence of endogeneity
- In other words, the estimates for both models are expected to be the same in the absence of endogeneity
- The RE model is significant with a chi-sq statistic of 399 and R-sq of 25%
- All the coefficient estimates are significant at alpha = 10%

n) Test for Endogeneity - Hausman Test

```
##
##   Hausman Test
##
## data:  fe_form
## chisq = 64.486, df = 8, p-value = 0.00000000006098
## alternative hypothesis: one model is inconsistent

##
##   Hausman Test
##
## data:  fe_form
## chisq = 644.49, df = 8, p-value < 0.00000000000000022
## alternative hypothesis: one model is inconsistent

##
##   Hausman Test
##
|

## data:  fe_form
## chisq = 554.09, df = 8, p-value < 0.00000000000000022
## alternative hypothesis: one model is inconsistent
```

Conducting the Hausman Test for endogeneity, the following points can be concluded:

Hausman test is conducted for the following combinations

1. Random Effects Vs Fixed Effects (Entity fixed)
2. Random Effects Vs Fixed Effects (Time Fixed)
3. Random Effects Vs Fixed Effects (Entity Fixed and Time Fixed)

Results:

- All the three tests provide the same result - Endogeneity is present
- The second test has a very high chi-sq statistic of 644 which implies that the estimates are very different for each of the model

o) Fixed Effects (Time) model with interaction terms + Cluster Errors

```
## Oneway (time) effect Within Model
##
## Call:
## plm(formula = fe_form2, data = data, effect = "time", model = "within",
##     index = c("stateid", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.574234 -0.232063  0.042893  0.284732  1.266076
##
## Coefficients:
##                     Estimate  Std. Error  t-value              Pr(>|t|)
## incarc_rate       0.00151316  0.00011730  12.9000 < 0.00000000000000022 ***
## pb1064            0.11361801  0.01754034   6.4775       0.0000000001383 ***

## pw1064            0.04976878  0.00880582   5.6518       0.0000000200453 ***
## pm1029           -0.05350015  0.01682621  -3.1796              0.001515 **
## pop               0.03879830  0.00242949  15.9697 < 0.00000000000000022 ***
## avginc            0.01433280  0.00763616   1.8770              0.060778 .
## density           0.03175494  0.01398226   2.2711              0.023327 *
## shall            -0.72932870  0.05469640 -13.3341 < 0.00000000000000022 ***
## incarc_rate:shall  0.00192093  0.00019202  10.0041 < 0.00000000000000022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    474.53
## Residual Sum of Squares: 183.2
## R-Squared:      0.61394
## Adj. R-Squared: 0.60345
## F-statistic: 201.61 on 9 and 1141 DF, p-value: < 0.000000000000000222

##
## t test of coefficients:
##
##                     Estimate  Std. Error t value   Pr(>|t|)
## incarc_rate       0.00151316  0.00062136  2.4352  0.0150339 *
## pb1064            0.11361801  0.07806682  1.4554  0.1458352
## pw1064            0.04976878  0.03728062  1.3350  0.1821501
## pm1029           -0.05350015  0.07935814 -0.6742  0.5003456
## pop               0.03879830  0.00925225  4.1934 0.00002961 ***
## avginc            0.01433280  0.02584079  0.5547  0.5792373
## density           0.03175494  0.05107242  0.6218  0.5342219
## shall            -0.72932870  0.20185145 -3.6132  0.0003156 ***
## incarc_rate:shall  0.00192093  0.00060301  3.1856  0.0014838 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The Fixed Effects model with Time fixed effects is supposed to be the right model because of the significance of the model and also thre reasonable estimate of ‘shall’
- We use this model as the base and explore with ‘shall’ as an interaction term with ‘incarc_rate’
- The estimate for ‘shall’ is -0.729 which suggests that for states with shall-issue laws the crime rate is 72.9% lower than crime rate in states without shall-issue laws
- Also, interpreting the interaction term, for states with shall issue laws, the return on incarceration rate is 0.19% higher than for states with no shall-issue laws.

p) Fixed Effects (Time) model with quadratic term + Cluster Errors

```
## Oneway (time) effect Within Model
##
## Call:
## plm(formula = fe_form3, data = data, effect = "time", model = "within",
##     index = c("stateid", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##       Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.368628 -0.253902  0.031288  0.258616  1.265925


##
## Coefficients:
##                       Estimate     Std. Error  t-value              Pr(>|t|)
## incarc_rate       0.00434427579  0.00018474772  23.5146 < 0.00000000000000022
## pb1064            0.05487780303  0.01676968063   3.2724              0.001098
## pw1064            0.02399081555  0.00833295284   2.8790              0.004063
## pm1029           -0.04118976099  0.01587762931  -2.5942              0.009603
## pop               0.03270859259  0.00234479144  13.9495 < 0.00000000000000022
## avginc            0.03400972846  0.00733772419   4.6349   0.000003982195899936
## density           0.02666588661  0.01284161167   2.0765              0.038069
## shall            -0.24064149159  0.03067028875  -7.8461   0.000000000000009822
## I(incarc_rate^2) -0.00000186439  0.00000011816 -15.7791 < 0.00000000000000022
##
## incarc_rate      ***
## pb1064           **
## pw1064           **
## pm1029           **
## pop              ***
## avginc           ***
## density          *
## shall            ***
## I(incarc_rate^2) ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    474.53
## Residual Sum of Squares: 163.57
## R-Squared:      0.65529
## Adj. R-Squared: 0.64593
## F-statistic: 241.008 on 9 and 1141 DF, p-value: < 0.000000000000000222


##
## t test of coefficients:
##
##                       Estimate     Std. Error t value        Pr(>|t|)
## incarc_rate       0.00434427579  0.00069037495  6.2926 0.0000000004441 ***
## pb1064            0.05487780303  0.06968303781  0.7875         0.43113
## pw1064            0.02399081555  0.03385641448  0.7086         0.47871
## pm1029           -0.04118976099  0.07174534933 -0.5741         0.56601
## pop               0.03270859259  0.00757303675  4.3191 0.0000170301331 ***
## avginc            0.03400972846  0.02385338694  1.4258         0.15420
## density           0.02666588661  0.03067563645  0.8693         0.38487
## shall            -0.24064149159  0.11238352982 -2.1413         0.03247 *
## I(incarc_rate^2) -0.00000186439  0.00000035864 -5.1985 0.0000002379152 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- In this model, we explore the effect of the presence of a quadratic term for incarc_rate
- ‘incarc_rate, ’incarc_rateˆ2’, ‘pop’ and ‘shall’ are significant at alpha=5%
- Interpreting ‘shall’ estimate, states with shall-issue laws have 24% lower crime rate than states without shall-issue laws.
- There is a negative sign on incarc_rateˆ2 which suggests that past a point, incarc_rate decreases as crime rate increases

q) Fixed Effects model (Time) with quadratic, interaction term + Cluster Errors

```
## Oneway (time) effect Within Model
##
## Call:
## plm(formula = fe_form4, data = data, effect = "time", model = "within",
##     index = c("stateid", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -1.42007 -0.25068  0.02654  0.25681  1.31347
##
## Coefficients:
##                      Estimate     Std. Error  t-value             Pr(>|t|)
## incarc_rate        0.00380303471  0.00020554255  18.5024 < 0.00000000000000022
## pb1064             0.06827682486  0.01670977711   4.0860          0.00004695144
## pw1064             0.03148953265  0.00832525119   3.7824              0.0001634
## pm1029            -0.04671433990  0.01569332639  -2.9767              0.0029752
## pop                0.03259691909  0.00231322539  14.0915 < 0.00000000000000022
## avginc             0.03215859976  0.00724597557   4.4381          0.00000995394
## density            0.04504655254  0.01307290975   3.4458              0.0005901
## shall             -0.49536762786  0.05399830158  -9.1738 < 0.00000000000000022
## I(incarc_rate^2)  -0.00000162794  0.00000012373 -13.1567 < 0.00000000000000022
## incarc_rate:shall  0.00108214179  0.00019000545   5.6953          0.00000001566
##
## incarc_rate       ***
## pb1064            ***
## pw1064            ***
## pm1029            **
## pop               ***
## avginc            ***
## density           ***
## shall             ***
## I(incarc_rate^2)  ***
## incarc_rate:shall ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    474.53
## Residual Sum of Squares: 159.05
## R-Squared:      0.66483
## Adj. R-Squared: 0.65542
## F-statistic: 226.127 on 10 and 1140 DF, p-value: < 0.000000000000000222

##
## t test of coefficients:
##
##                      Estimate     Std. Error t value      Pr(>|t|)
## incarc_rate        0.00380303471  0.00068487395  5.5529 0.00000003494 ***
## pb1064             0.06827682486  0.07044062361  0.9693      0.332610
## pw1064             0.03148953265  0.03407740931  0.9241      0.355651
## pm1029            -0.04671433990  0.07417159321 -0.6298      0.528942
## pop                0.03259691909  0.00727842678  4.4786 0.00000826984 ***
## avginc             0.03215859976  0.02375018266  1.3540      0.175993
## density            0.04504655254  0.03131280959  1.4386      0.150539
## shall             -0.49536762786  0.18101042722 -2.7367      0.006303 **
## I(incarc_rate^2)  -0.00000162794  0.00000037108 -4.3870 0.00001255738 ***
## incarc_rate:shall  0.00108214179  0.00044723719  2.4196      0.015693 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With a model incorporating both quadratic and interaction terms with Cluster Robust standard errors, we have below interpretations:

- ‘shall’, ‘incarc_rate’, ‘incarc_rateˆ2’, ’incarc_rate*shall’ are significant at alpha=5%.
- For states with shall-issue laws the crime rate is 49.5% lower than crime rate in states without shall-issue laws.
- For states with shall-issue laws, for 1 unit increase in incarceration rate the crime rate increases by 0.38% approximate.

Part 3: Conclusion/Recommendations

From our analysis we conclude/recommend the following:

1. The Time Fixed Effects model is the most accurate model. All the estimates for the variables are as expected and reasonable.
2. Using it as the base, we can say that for states with shall-issue laws the crime rate is lower than the crime rate for states without shall-issue laws.
3. Hence, to answer the problem statement - ’Yes, shall-issue laws do reduce crime.
4. Implementation of shall-issue laws is advised for the states without shall-issue laws, as its implementation will lead to a decrease in crime rate

Code

Adityan Rajendran (RStudio – Rmd)

```
---
title: "ECONOMETRICS Project"
author: "Adityan Rajendran"
date: "26/11/2019"
output: pdf_document
---
```{r setup, include=FALSE}
if(!require('pacman')) install.packages('pacman')
pacman::p_load(ggplot2, foreign, e1071, psych, tidyverse, reshape, corrplot, lubridate, plm, lmtest)
```
```{r data import, echo=FALSE}
options(scipen = 999)
data <- as.data.frame(read.dta('guns.dta'))
data$year <- as.integer(data$year + 1900)
eda_custom <- function(df) {
 sprintf('%f X %f',dim(df)[1], dim(df)[2])
 eda_df <- data.frame(matrix(nrow=NCOL(df), ncol=0))
 row.names(eda_df) <- colnames(df)
   eda_df$types <- sapply(df, typeof)
 eda_df$nulls <- sapply(df, function(x) {sum(is.na(x))})
 eda_df$skew <- sapply(df, skewness)
 eda_df$min <- sapply(df, min)
 eda_df$median <- sapply(df, median)
 eda_df$max <- sapply(df, max)
 eda_df$mean <- sapply(df, mean)
 return(eda_df)
}
(eda_custom(data))
```
```{r, fig.height = 6, echo=FALSE}
multi.hist(data)
```
Histograms of the variables
```{r outlier detection,fig.height = 6, echo=FALSE}
boxplot(data[-1])
```
```{r correlation, fig.height = 6, echo=FALSE}
corr <- round(cor(data[-1]),2)
corrplot(corr, method = 'color', type='lower')
```
```{r eda, echo=FALSE}
shall <- data[data$shall == 1,]
```
```

````
no_shall <- data[data$shall == 0,]
states_shall <- unique(shall$stateid)
states_noshall <- unique(no_shall$stateid)
commons <- intersect(states_shall, states_noshall)
states_n <- setdiff(states_noshall, commons)
states_s <- setdiff(states_shall, commons)
```
# Model Implementation
## Pooled OLS
```{r data processing for Pooled OLS, echo=FALSE}
data$year <- as.factor(data$year)
data$stateid <- as.factor(data$stateid)
proc_data <- data
proc_data <- proc_data[-c(1,3,4,12)]
```
```{r Pooled OLS, echo=FALSE}
pooled.ols <- plm(log(vio) ~ incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall, data =
data, model = 'pooling', index = c("stateid","year"))
summary(pooled.ols)
ggplot(data, aes(shall, vio)) + geom_bar(stat='identity') +
  xlab('Shall-issue laws') + ylab('Violent Crime Rate') +
  ggtitle('Violent Crime Rate (incidents/100k members) VS Shall-issue laws') + theme_bw()
```
## Pooled OLS with Cluster-Robust Standard Errors
```{r Pooled OLS with robust errrors, echo=FALSE}
#Standard Erros - clustered by States
coeftest(pooled.ols, vcov = vcovHC(pooled.ols, type='HC1', cluster='group'))
```
## Fixed Effects Model - Entity Fixed
```{r Fixed Effects model - Entity Fixed, echo=FALSE}
fe_form <- log(vio) ~ incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall
fe.model.ef <- plm(fe_form, data, model = 'within', index = c("stateid","year"), effect = 'individual')
summary(fe.model.ef)
```
## Fixed Effects Model (Entity Fixed) + Robust Errors
```{r Fixed Effects + Robust Errors, echo=FALSE}
coeftest(fe.model.ef, vcov = vcovHC(fe.model.ef, type='HC1', cluster='group'))
```
## Fixed Effects Model - Time Fixed
```{r Fixed Effects model - time fixed effects, echo=FALSE}
fe_form_tf <- log(vio) ~ incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall + factor(year)

fe.model.tf <- plm(fe_form_tf, data, model = 'within', index = c("stateid","year"), effect = 'time')
summary(fe.model.tf)
````

````
```
## Fixed Effects Model (Time Fixed) + Robust Errors
```{r Fixed Effects Model(Time) + Robust errors}
coeftest(fe.model.tf, vcov = vcovHC(fe.model.tf, type='HC1', cluster='group'))
```
## Fixed Effects model - Entity + Time fixed effects
```{r Fixed Effects model - entity plus time fixed effects, echo=FALSE}
fe_form <- log(vio) ~ incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall
fe.model.eftf <- plm(fe_form, data, model = 'within', index = c("stateid","year"), effect = 'twoways')
summary(fe.model.eftf)
```
## Fixed Effects model (Entity + Time) + Cluster Standard Errors
```{r Fixed (Entity + Time) + Cluster Errors}
coeftest(fe.model.eftf, vcov = vcovHC(fe.model.eftf, type='HC1', cluster='group'))
```
## Random Effects Model
```{r Random Effects model, echo=FALSE}
re_form <- log(vio) ~ incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall
re.model <- plm(re_form, data, model = 'random', index = c("stateid","year"))
summary(re.model)
```
## Test for Endogeneity - Hausman Test
```{r Test for endogeneity, echo=FALSE}
phtest(fe.model.ef, re.model)
phtest(fe.model.tf, re.model)
phtest(fe.model.eftf, re.model)
```
## Fixed Effects model with interaction terms + Cluster Errors
```{r Fixed interactions, echo=FALSE}
fe_form2 <- log(vio) ~
incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall+(shall*incarc_rate)
fe.model.tf2 <- plm(fe_form2, data, model = 'within', index = c("stateid","year"), effect = 'time')
summary(fe.model.tf2)
coeftest(fe.model.tf2, vcov = vcovHC(fe.model.tf2, type='HC1', cluster='group'))
```
## Fixed Effects model with quadratic term + Cluster Errors
```{r Fixed quadratic, echo=FALSE}
fe_form3 <- log(vio) ~
incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall+I(incarc_rate**2)
fe.model.tf3 <- plm(fe_form3, data, model = 'within', index = c("stateid","year"), effect = 'time')
summary(fe.model.tf3)
coeftest(fe.model.tf3, vcov = vcovHC(fe.model.tf3, type='HC1', cluster='group'))
```
## Fixed Effects model with quadratic, interaction term + Cluster Errors
````
````

```{r Fixed interactions + quadratic, echo=FALSE}
fe_form4 <- log(vio) ~
incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall+I(incarc_rate**2)+(shall*incarc_rate)
fe.model.tf4 <- plm(fe_form4, data, model = 'within', index = c("stateid","year"), effect = 'time')
summary(fe.model.tf4)
coeftest(fe.model.tf4, vcov = vcovHC(fe.model.tf4, type='HC1', cluster='group'))
```

###

Mariam Coulibaly (STATA)

```
log using "C:\Users\smc180005\Downloads\new project econo.log"

use "C:\Users\smc180005\Downloads\guns.dta"

*summary statistics

summarize, allbaselevels

*pooled OLS model

generate lvio=ln(vio)

regress lvio i.shall incarc_rate pb1064 pm1029 pop avginc pw1064 density

*check multicolinearity

vif

*check heteroskedasticity

estat imtest, white

*new pooled OLS model with robust standard errors

regress lvio i.shall incarc_rate pb1064 pm1029 pop avginc pw1064 density, vce(robust)

*Entity Fixed Effect model

xtset stateid

xtreg lvio i.shall density avginc pop pm1029 pw1064 pb1064 incarc_rate, fe cluster (stateid)

*Entity and Time Fixed Effect model

xtreg lvio i.shall density avginc pop pm1029 pw1064 pb1064 incarc_rate i.year, fe vce(cluster stateid)

*Fixed Time Effect model

regress lvio 1.shall density avginc pop pm1029 pw1064 pb1064 incarc_rate i.year, vce(robust)
```

###

Siddharth Govindarajan (RStudio)

```
#Import required packages
list.of.packages <- c("foreign","ggplot2")
```

```
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)
library(foreign)
library(ggplot2)

#Read dta file
guns_data <- read.dta("guns.dta")

#Question 1(A) : Robbery
robbery_aggregation <- aggregate(guns_data$rob,
            by = list(guns_data$year),
            FUN = sum)
colnames(robbery_aggregation) <- c("Year","Number_of_Robbery")

#ggplot(guns_data)+geom_line(aes(year, rob))
ggplot(robbery_aggregation, aes(x = Year, y = Number_of_Robbery),
     ) +
 geom_line()+ xlim(c(77,99))+
 geom_line( color="#69b3a2", size=1, alpha=0.9)+
 geom_point()

#Question 1(B) : Violent
robbery_aggregation <- aggregate(guns_data$vio,
                         by = list(guns_data$year),
                         FUN = sum)
colnames(robbery_aggregation) <- c("Year","Number_of_Violent_Cases")

#ggplot(guns_data)+geom_line(aes(year, rob))
ggplot(robbery_aggregation, aes(x = Year, y = Number_of_Violent_Cases),
) +
 geom_line()+ xlim(c(77,99))+
 geom_line( color="red", size=1, alpha=0.9)+
 geom_point()

#Question 1(C) : Murder
robbery_aggregation <- aggregate(guns_data$mur,
                         by = list(guns_data$year),
                         FUN = sum)
colnames(robbery_aggregation) <- c("Year","Number_of_Murder_Cases")

#ggplot(guns_data)+geom_line(aes(year, rob))
ggplot(robbery_aggregation, aes(x = Year, y = Number_of_Murder_Cases),
) +
 geom_line()+ xlim(c(77,99))+
```

```
  geom_line( color="#4682B4", size=1, alpha=0.9)+
  geom_point()

#Question 2: Incarcination rate
incarcination_rate_aggregation <- aggregate(guns_data$incarc_rate,
                              by = list(guns_data$year),
                              FUN = sum)
colnames(incarcination_rate_aggregation) <- c("Year","Incarcination_rate")

#ggplot(guns_data)+geom_line(aes(year, rob))
ggplot(incarcination_rate_aggregation, aes(x = Year, y = Incarcination_rate),
) + xlim(c(77,99))+
  geom_area( color="#3CB371", size=1, alpha=0.9,fill = "lightblue")+
  geom_point()

#Question 3 a) Robbery: Which state has most of the above crimes?
state_crime_robbery <- aggregate(guns_data$rob,
                                    by = list(guns_data$stateid),
                                    FUN = sum)
colnames(state_crime_robbery) <- c("StateId","Robbery_Rate_State")

state_crime_robbery_highest <- head(state_crime_robbery[order(-
state_crime_robbery$Robbery_Rate_State),],10)
state_crime_robbery_highest$StateId <- as.character(state_crime_robbery_highest$StateId)
ggplot(state_crime_robbery_highest,  aes(x = reorder(StateId, -Robbery_Rate_State) , y =
Robbery_Rate_State))+geom_bar( fill="#69b3a2",stat="identity")+xlab("StateId")+
  ggtitle("Most robberies")

#Question 3 b) Robbery: Which state has least of the above crimes?
state_crime_robbery_lowest <-
head(state_crime_robbery[order(state_crime_robbery$Robbery_Rate_State),],10)
state_crime_robbery_lowest$StateId <- as.character(state_crime_robbery_lowest$StateId)
ggplot(state_crime_robbery_lowest,  aes(x = reorder(StateId, Robbery_Rate_State) , y =
Robbery_Rate_State))+geom_bar( fill="#69b3a2",stat="identity")+xlab("StateId")+
  ggtitle("Least robberies")

#Question 4 a) Violent crimes: Which state has most of the above crimes?
state_crime_violent <- aggregate(guns_data$vio,
                              by = list(guns_data$stateid),
                              FUN = sum)
colnames(state_crime_violent) <- c("StateId","Violent_Rate_State")
state_crime_violent_highest <- head(state_crime_violent[order(-
state_crime_violent$Violent_Rate_State),],10)
state_crime_violent_highest$StateId <- as.character(state_crime_violent_highest$StateId)
```

```
ggplot(state_crime_violent_highest,  aes(x = reorder(StateId,
                                      -Violent_Rate_State) , y = Violent_Rate_State))+geom_bar(
fill="red",stat="identity")+xlab("StateId")+ggtitle("Most Violent crimes")

#Question 4 b) Violent crimes: Which state has least of the above crimes?
state_crime_violent_lowest <-
head(state_crime_violent[order(state_crime_violent$Violent_Rate_State),],10)
state_crime_violent_lowest$StateId <- as.character(state_crime_violent_lowest$StateId)
ggplot(state_crime_violent_lowest,  aes(x = reorder(StateId, Violent_Rate_State) , y =
Violent_Rate_State))+geom_bar( fill="red",stat="identity")+xlab("StateId")+
  ggtitle("Least Violent crimes")

#Question 5 a) Murder: Which state has most of the above crimes?
state_crime_murder <- aggregate(guns_data$mur,
                        by = list(guns_data$stateid),
                        FUN = sum)
colnames(state_crime_murder) <- c("StateId","Murder_State")
state_crime_murder_highest <- head(state_crime_murder[order(-
state_crime_murder$Murder_State),],10)
state_crime_murder_highest$StateId <- as.character(state_crime_murder_highest$StateId)
ggplot(state_crime_murder_highest,  aes(x = reorder(StateId,
                                      -Murder_State) , y = Murder_State))+geom_bar(
fill="#4682B4",stat="identity")+xlab("StateId")+
  ggtitle("Most Murders")

#Question 5 b) Murder: Which state has least of the above crimes?
state_crime_murder <- aggregate(guns_data$mur,
                       by = list(guns_data$stateid),
                       FUN = sum)
colnames(state_crime_murder) <- c("StateId","Murder_State")
state_crime_murder_highest <-
head(state_crime_murder[order(state_crime_murder$Murder_State),],10)


state_crime_murder_highest$StateId <- as.character(state_crime_murder_highest$StateId)
ggplot(state_crime_murder_highest,  aes(x = reorder(StateId,
                                     Murder_State) , y = Murder_State))+geom_bar(
fill="#4682B4",stat="identity")+xlab("StateId")+
  ggtitle("Least Murders")

#Question 6 : Crimes count with arms and without arms
table(guns_data$shall)
gun_crimes_rate <- aggregate(guns_data$mur + guns_data$rob + guns_data$vio ,
                       by = list(guns_data$shall),
```

```
                        FUN = sum)
colnames(gun_crimes_rate) <- c("Guns","Number of crimes")
gun_crimes_rate$totalinstance <- c(table(guns_data$shall))
gun_crimes_rate$crime_per_instance <- gun_crimes_rate$`Number of
crimes`/gun_crimes_rate$totalinstance
gun_crimes_rate$Guns <-  as.character(gun_crimes_rate$Guns)
ggplot(gun_crimes_rate,  aes(x = Guns , y = `Number of crimes`))+geom_bar(
fill="#4682B4",stat="identity") + scale_x_discrete(labels=c("0" = "No shall carry", "1" = "shall carry"))

#Question 7: Each year, states which has issued shall laws.
gun_crimes_allowed_year <- aggregate(guns_data$shall ,
                      by = list(guns_data$year),
                      FUN = sum)
colnames(gun_crimes_allowed_year) = c("Year", "States_which_allowed")
ggplot(gun_crimes_allowed_year,  aes(x = Year , y = States_which_allowed))+geom_bar(
fill="#4682B4",stat="identity")+
  geom_line(robbery_aggregation,  mapping = aes(x = Year , y = Number_of_Murder_Cases))

#8
guns_state11 <- guns_data[guns_data$stateid == 11,]
table(guns_state11$shall)
guns_state38 <- guns_data[guns_data$stateid == 38,]
table(guns_state38$shall)
########################End of EDA #############################
#####Start of 8 tasks ######################
#(1)
lm_f_test_1_model <- lm(formula = vio ~
incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall, data = guns_data)
plot(lm_f_test_1_model)
summary(lm_f_test_1_model)

#(2) check for heteroskedasticity - informal
guns_data$resi <- lm_f_test_1_model$residuals
guns_data$resi_sq <- guns_data$resi * guns_data$resi

guns_data$incarc_rate_sq <- guns_data$incarc_rate * guns_data$incarc_rate
guns_data$pb1064_sq <- guns_data$pb1064 * guns_data$pb1064
guns_data$pw1064_sq <- guns_data$pw1064 * guns_data$pw1064
guns_data$pm1029_sq <- guns_data$pm1029 * guns_data$pm1029
guns_data$pop_sq <- guns_data$pop * guns_data$pop
guns_data$avginc_sq <- guns_data$avginc * guns_data$avginc
guns_data$density_sq <- guns_data$density * guns_data$density
guns_data$shall_sq <- guns_data$shall * guns_data$shall
var.func <- lm(resi_sq ~
```

```
              incarc_rate+incarc_rate_sq+
              pb1064+pb1064_sq+
              pw1064+pw1064_sq+
              pm1029+pm1029_sq+
              pop+pop_sq+
              avginc+avginc_sq+
              density+density_sq+
              shall, data = guns_data)
summary(var.func)

#(3) White - detect heteroskedasticity
lm_f_test_1_model <- lm(formula = vio ~
incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall, data = guns_data)
confint(lm_f_test_1_model)
#########################################################################################
```

Sushanth Chintalapati (R Markdown)

```
---
title: "EDA"
author: "Sushanth Chintalapati"
date: "11/20/2019"
output:
  word_document: default
  pdf_document: default
---
```{r setup, include=FALSE, echo = FALSE, warning = FALSE, message = FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
```{r loading libraries, echo = FALSE, warning = FALSE, message = FALSE}
library("tidyverse")
library("ggplot2")
library("dplyr")
library("foreign")
library("readxl")
library("corrplot")
library("knitr")
```
Data Set Structure
and the correlations between the variables.
```{r loading the data, echo = FALSE, warning = FALSE, message = FALSE}
guns <- read.dta("guns.dta")
glimpse(guns)
cor(guns) %>% kable(caption = "Correlations among the variables")
corrplot(cor(guns), method = "color")
guns$stateid <- guns$stateid %>% as.factor()
```
```

```
About states that had or didn't have shall act through out the given duration:
```{r States and shall, echo = FALSE, warning = FALSE, message = FALSE}
guns %>% mutate(shall = ifelse(shall == 0, "WithoutAct", "WithAct")) %>% group_by(stateid, shall) %>% summarise(Count = n()) %>% filter(Count == 23) %>% arrange(shall, stateid) %>% filter(shall == "WithAct") %>% select(-c("shall", "Count")) %>%
  kable(caption = "The states with Shall through out")
guns %>% mutate(shall = ifelse(shall == 0, "WithoutAct", "WithAct")) %>% group_by(stateid, shall) %>% summarise(Count = n()) %>% filter(Count == 23) %>% arrange(shall, stateid) %>% filter(shall != "WithAct") %>% select(-c("shall", "Count")) %>%
  kable(caption = "The states without Shall through out")

```
Maximum and Minimum crime rates before and after the shall act:
```{r Crime rate Max and Min, echo = FALSE, warning = FALSE, message = FALSE}
guns %>% select("shall", "vio") %>% mutate(shall = ifelse(shall == 0, "WithoutAct", "WithAct")) %>% group_by(shall) %>% summarise(`maximum crime rate` = max(vio)) %>% left_join((guns %>% select("stateid", "vio")), by = c("maximum crime rate" = "vio")) %>% select("stateid", "shall", "maximum crime rate") %>%
  kable(caption = "Maximum crime rate before and after the law")
guns %>% select("shall", "vio") %>% mutate(shall = ifelse(shall == 0, "WithoutAct", "WithAct")) %>% group_by(shall) %>% summarise(`minimum crime rate` = min(vio)) %>% left_join((guns %>% select("stateid", "vio")), by = c("minimum crime rate" = "vio")) %>% select("stateid", "shall", "minimum crime rate") %>%
  kable(caption = "Minimum crime rate before and after the law")
```
All about the Crime Rate and the act:
```{r Crime Forencics with visualizations, echo = FALSE, warning = FALSE, message = FALSE}
guns %>%
  ggplot(aes(x = year, y = vio))+geom_point()+facet_wrap(~as.factor(shall))

kable(guns %>% filter(vio > 1500) %>% select(c("stateid", "year", "pb1064", "pw1064", "pm1029", "pop", "avginc", "shall")), caption = "Max Crime")
guns %>% filter(stateid != 11) %>%
  ggplot(aes(x = year, y = vio, color = shall))+geom_point()+facet_wrap(~as.factor(shall))+
  theme(legend.position = "none")+labs(x = "Year", y = "Crime Rate", title = "Crime Rate distribution over time")
guns %>% group_by(stateid, shall) %>% summarise(avio = sum(vio)/n(), count = n()) %>%
  ggplot(aes(x = stateid, y = avio, fill = factor(shall, levels = c("1", "0"))))+geom_bar(stat = "identity", position = "fill")+coord_flip()+geom_hline(yintercept = 0.5, color = "red", linetype = "dashed")+scale_fill_manual(values = c("#132b43", "#56b1f7"))+labs(y = "Normatized crime percentage", x = "State ID", title = "Crime as a percentage before and after the law", fill = "Presence of Shall Act")

````
g <- guns %>% select("stateid", "shall", "vio") %>% group_by(stateid, shall) %>% summarise(crime =
sum(vio)/n()) %>% mutate(shall = ifelse(shall == 0, "WithoutAct", "WithAct")) %>% spread(shall, crime)
%>% na.omit() %>% mutate(response = ifelse(WithoutAct < WithAct, "Increased", "Decreased"))
kable(table(g$response), caption = "Effect on Crime after the Shall Act")
g %>% ggplot(aes(x = as.factor(reorder(stateid, WithAct))))+
  geom_bar(aes(y = WithAct, fill = "With the Act"), stat = "identity")+
  geom_bar(aes(y = WithoutAct, alpha = 0.7, fill = "Without the Act"), stat =
"identity")+scale_fill_manual(values = c("#132b43", "#56b1f7"))+labs(y = "Avg Crime Rate", x = "State
ID", title = "Crime with and without the Sahll Act (Not a stacked chart)")+coord_flip()
g <- g %>% mutate(`% Change` = (WithAct - WithoutAct)/WithoutAct*100) %>% arrange(desc(`%
Change`))

g %>% select(c("stateid", "% Change")) %>% kable(caption = "The percentage change in the crime after
the introduction of the Shall Act")
g %>% ggplot(aes(y = `% Change`, x = reorder(stateid, `% Change`)))+geom_bar(aes(fill = response), stat
= "identity")+scale_fill_manual(values = c("#132b43", "#56b1f7"))+labs(y = "Percentage Change in the
Average Crime", x = "State ID", title = "Percentage Change in the Average Crime with the introduction of
the Shall Act", fill = "Change")+coord_flip()
```
```{r setup, include=FALSE, echo = FALSE, warning = FALSE, message = FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

```{r loading libraries, echo = FALSE, warning = FALSE, message = FALSE}
library("tidyverse")
library("ggplot2")
library("dplyr")
library("foreign")
library("readxl")
library("corrplot")
library("knitr")
library("lmtest")
library("broom")
library("car")
library("plm")
options(scipen = 999)
```
Data Set Structure and the correlations between the variables.
```{r loading the data, warning = FALSE, message = FALSE}
guns <- read.dta("guns.dta")
guns <- guns %>% select(-c("mur", "rob"))
guns <- guns %>% select(c(10, 1, 11, seq(3,9, by = 1), 2)) %>% arrange(stateid, year)
cor(guns) %>% kable(caption = "Correlations among the variables")
cor(guns) %>% corrplot(method = "color",
````

```
        type="upper", order="hclust",
        addCoef.col = "black",
        tl.col="black", tl.srt=45,
        sig.level = 0.01, insig = "blank",
        diag=FALSE
        )
guns$stateid <- guns$stateid %>% as.factor()
guns$year <- guns$year %>% as.factor()
guns$shall <- guns$shall %>% as.factor()
glimpse(guns)
```
```
## **Heteroscadacity**
```{r Hetroscadacity, echo = FALSE, warning = FALSE, message = FALSE}
fit <- lm(log(vio) ~ ., data = guns %>% select(-c("stateid", "year")))
g <- data.frame("res" = fit$residuals, "fit" = fit$fitted.values)
ggplot(data = g, aes(y = res, x = fit)) + geom_point(col = 'blue') + geom_abline(slope = 0)
bptest(fit)
kable(tidy(bptest(fit)),
caption="Breusch-Pagan heteroskedasticity test")
```
Though the plot doesn't give an accurate evidence of the presence of hetroscadacity.
The BP test gives a very significant evidence of it.
Hense we must use the white robust standard errors.
Correcting Standard Errors
```{r, Standard Errors, echo = FALSE, warning = FALSE, message = FALSE}
fit <- lm(log(vio)~.,data=guns %>% select(-c("stateid", "year")))
summary(fit)
kable(tidy(fit),caption=
 "Regular standard errors in the 'guns' equation")
cov1 <- hccm(fit, type="hc1")
fit.hc1 <- coeftest(fit, vcov.=cov1)
kable(tidy(fit.hc1),caption=
  "Robust (HC1) standard errors in the 'fit' equation")
```
Pooled OLS
```{r Pooled OLS, echo = FALSE, warning = FALSE, message = FALSE}
fit <- lm(log(vio)~., data=guns %>% select(-c("stateid", "year")))
summary(fit)
ggplot(data = fit, aes(y = fit$residuals, x = fit$fitted.values)) +
  geom_point(col = 'blue') +
  geom_abline(slope = 0, color = "red", linetype = "dashed") +
  labs(x = "Fitted Values", y = "Residuals", title = "Residual Plot")
```
Fixed Entity Effects model
```

```
```{r Fixed Entity, echo = FALSE, warning = FALSE, message = FALSE}
fit <- plm(log(vio)~shall + incarc_rate + pb1064 + pw1064 + pm1029 + pop + avginc + density, data=guns, method = "within", index = c("stateid", "year"))
summary(fit)
```
```

Fixed Time Effects Model

```
```{r, Fixed Time, echo = FALSE, warning = FALSE, message = FALSE}
fit <- plm(log(vio)~shall + incarc_rate + pb1064 + pw1064 + pm1029 + pop + avginc + density + year, data=guns, method = "within", index = c("stateid", "year"))
summary(fit)
```
```

###