Final project

Guns in 51 US states

Eugenia Lou (EHL180001)
I-Ting Chao (IXC171230)
Rishabh Daga (RXD180021)
Kaiyue Liu (KXL180021)
Mehraj SK (MXS180076)

Contents

1. Objective

The objective of this project is to understand how shall-issues laws affect crime rate across 51 states in U.S from 1977 to 1999.

2. Variable definitions

- The data total observations are 1173, across 51 states within 23 years with no missing value so this is a balanced panel data.
- Dependent variable identified – violent(vio)
- Expected explanatory variables – rob, mur, shall, incarc_rate, density, avginc, pop, pm1029, pw1064, pb1064, stateid, year

Variable	type	Definition
vio	*numerical*	violent crime rate (incidents per 100,000 members of the population)
rob	*numerical*	robbery rate (incidents per 100,000)
mur	*numerical*	murder rate (incidents per 100,000)
shall	*binary*	= 1 if the state has a shall-carry law in effect in that year = 0 otherwise
incarc_rate	*numerical*	incarceration rate in the state in the previous year (sentenced prisoners per 100,000 residents; value for the previous year)
density	*numerical*	population per square mile of land area, divided by 1000
avginc	*numerical*	real per capita personal income in the state, in thousands of dollars
pop	*numerical*	state population, in millions of people
pm1029	*numerical*	percent of state population that is male, ages 10 to 29
pw1064	*numeri*	percent of state population that is white, ages 10 to 64

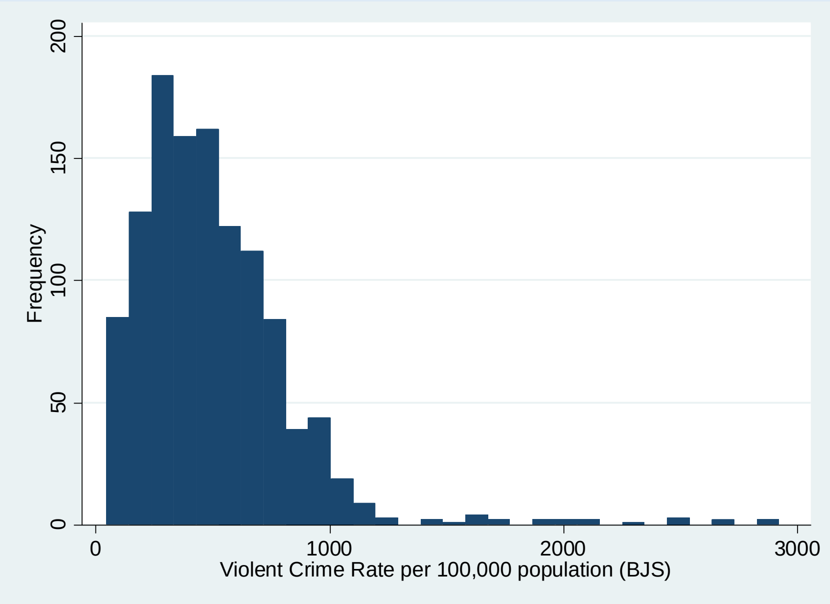
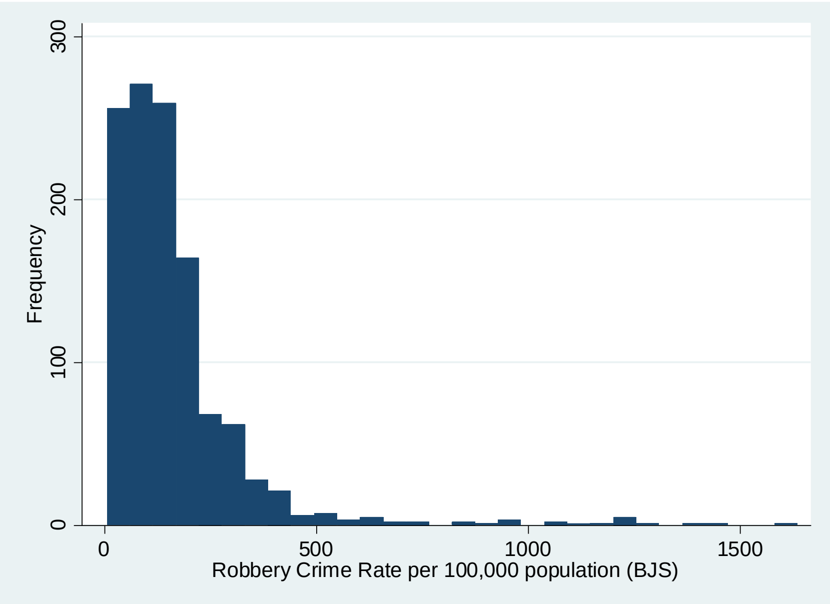
	cal	
pb1064	*numerical*	percent of state population that is black, ages 10 to 64
stateid	*character*	ID number of states (Alabama = 1, Alaska = 2, etc.)
year	*character*	Year (1977-1999)

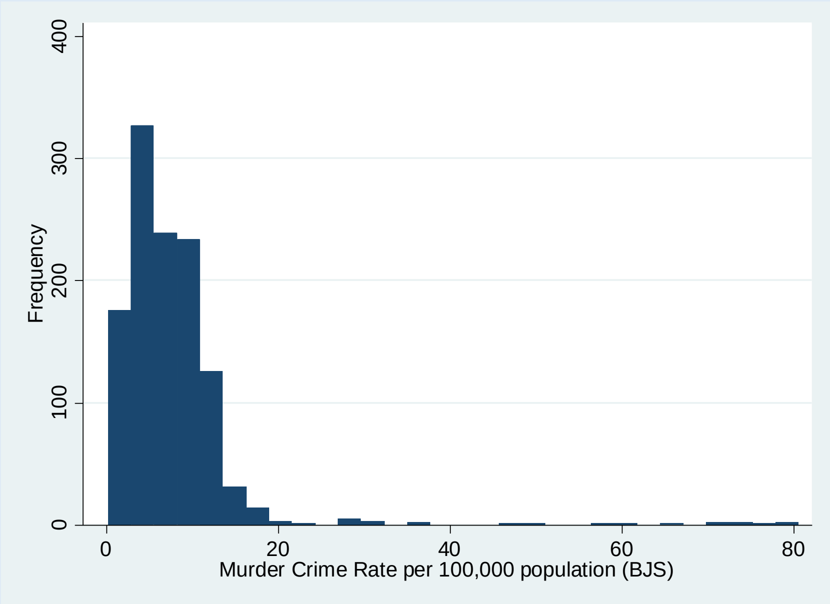
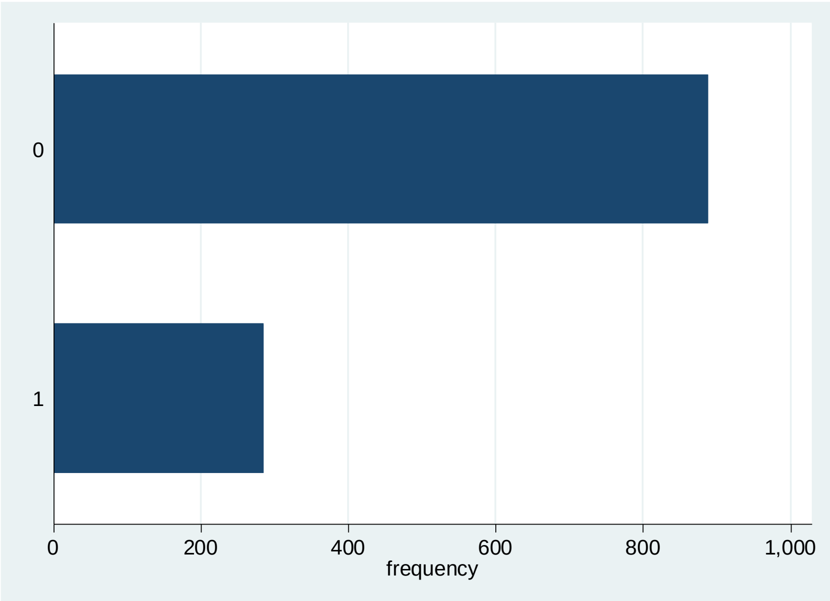
3. Summary statistic and distributions

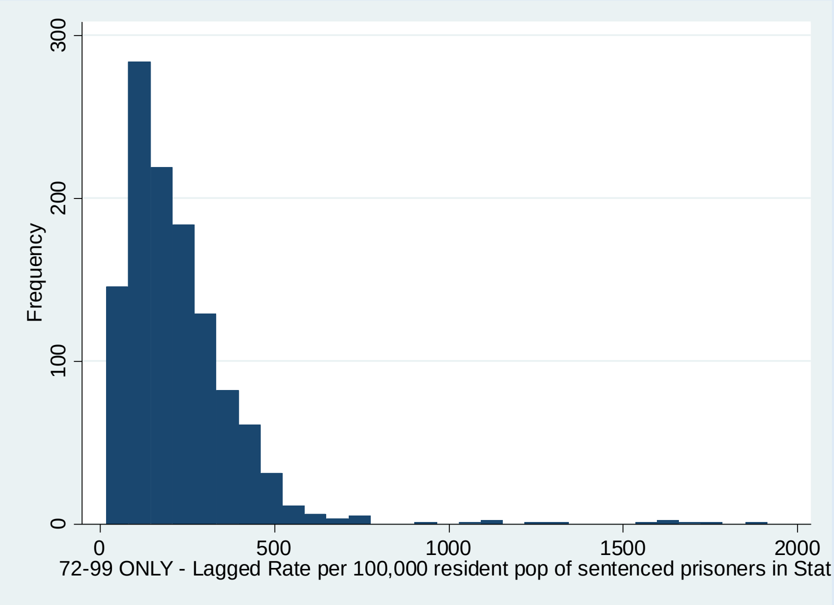
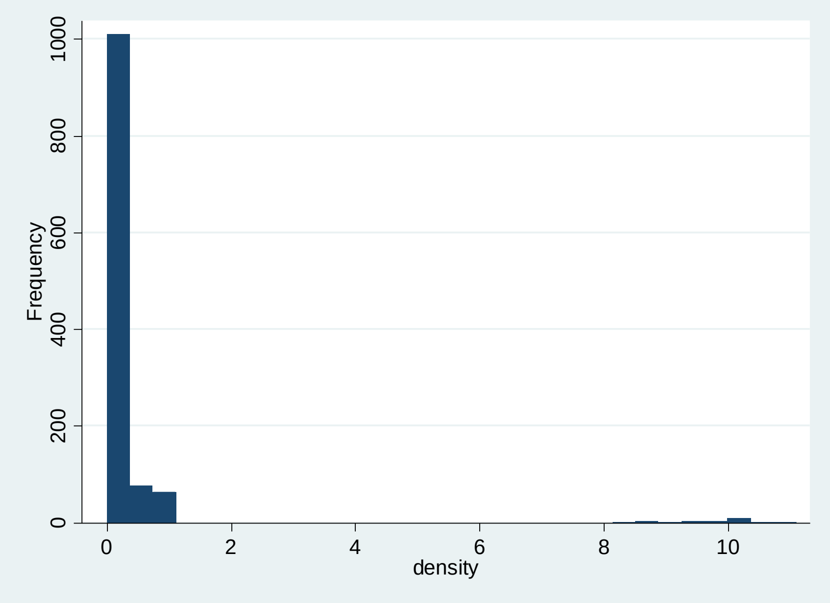
Summary statistic

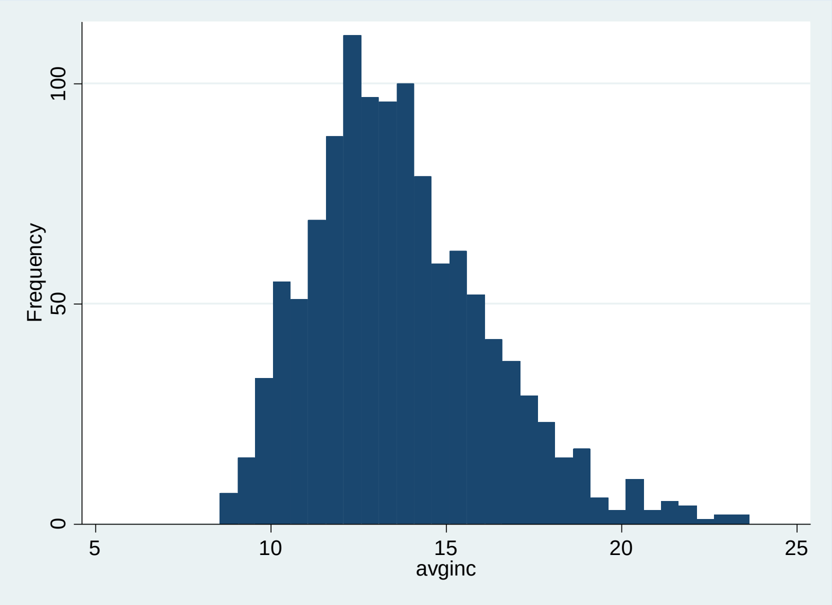
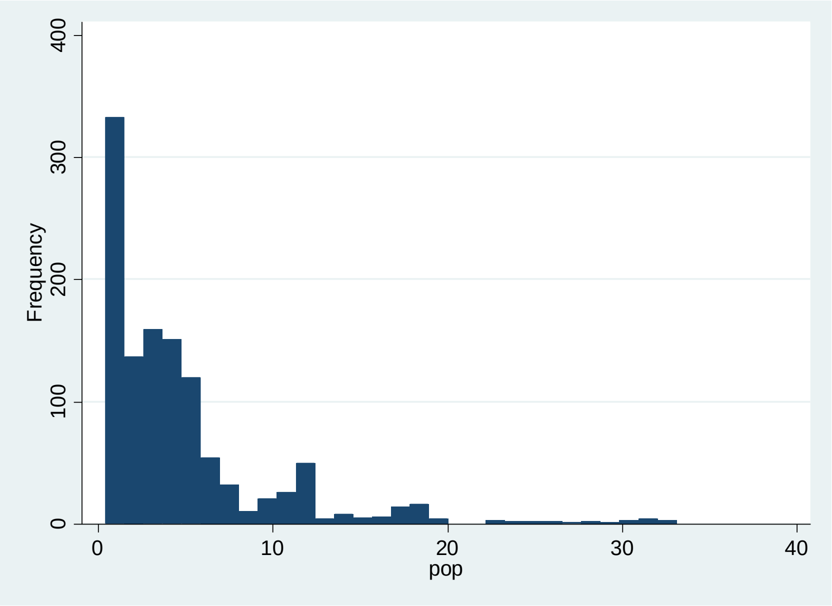
Variable	Mean	Median	Std	Min	Max
vio	503.07	439.45	334.28	47.00	2,921.80
rob	161.82	107.45	170.51	6.40	1,635.10
mur	7.67	11.25	7.52	0.20	80.60
incarc_rate	226.58	166.00	178.89	19.00	1,913
density	0.35	0.08	1.36	0.00	11.10
avginc	13.72	9.83	2.56	8.55	23.65
pop	4.82	3.92	5.25	0.40	33.15
pm1029	16.08	17.59	1.73	12.21	22.35
pw1064	62.95	54.90	9.76	21.78	76.53
pb1064	5.34	8.50	4.89	0.25	26.98

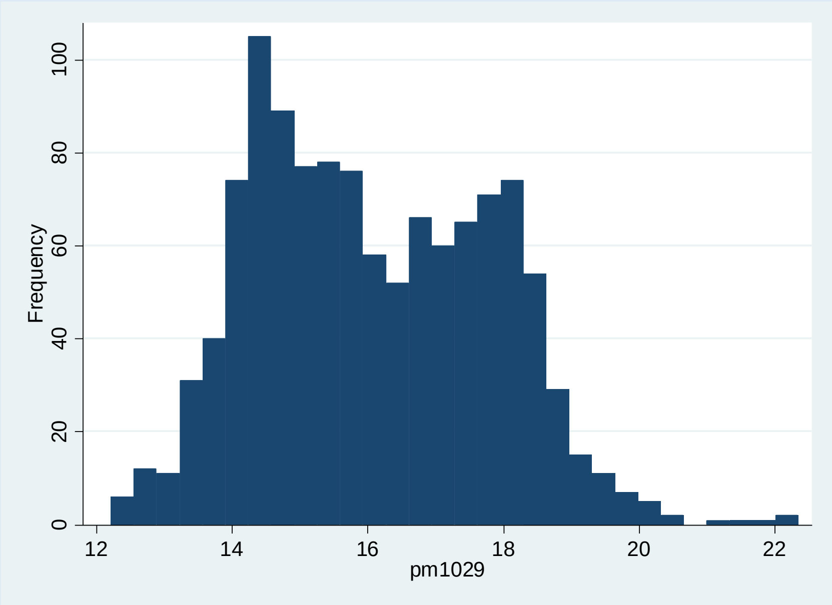
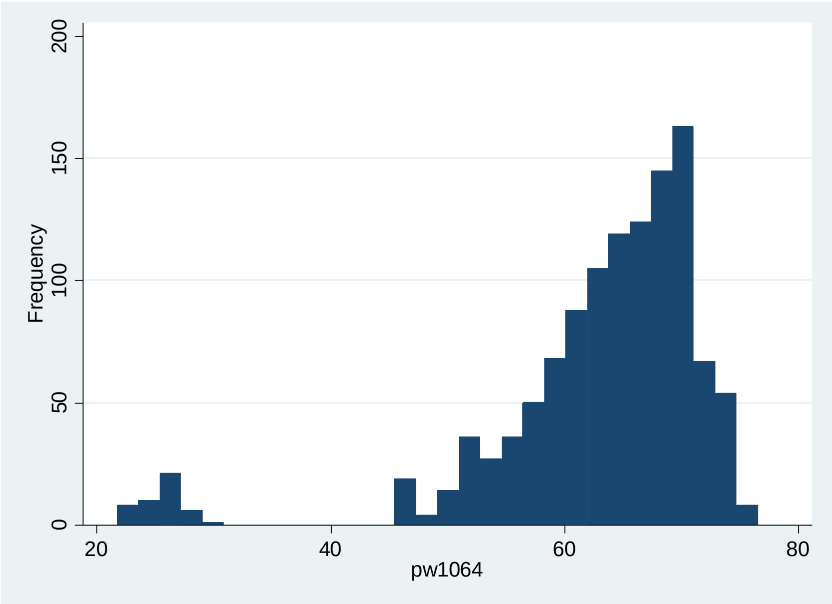
Distributions

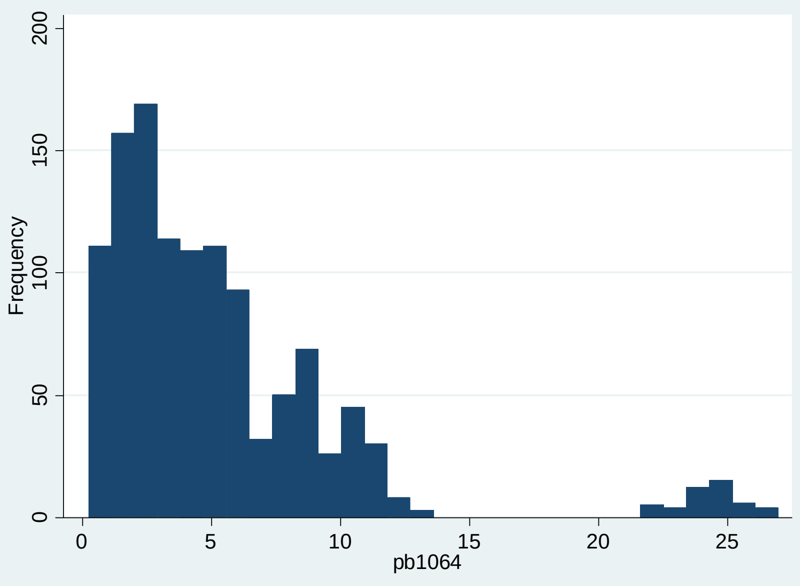
Variable	Distribution
vio	
rob	

mur	400 300 200 100 0 Frequency 0 20 40 60 80 Murder Crime Rate per 100,000 population (BJS)
shall	0 1 0 200 400 600 800 1,000 frequency

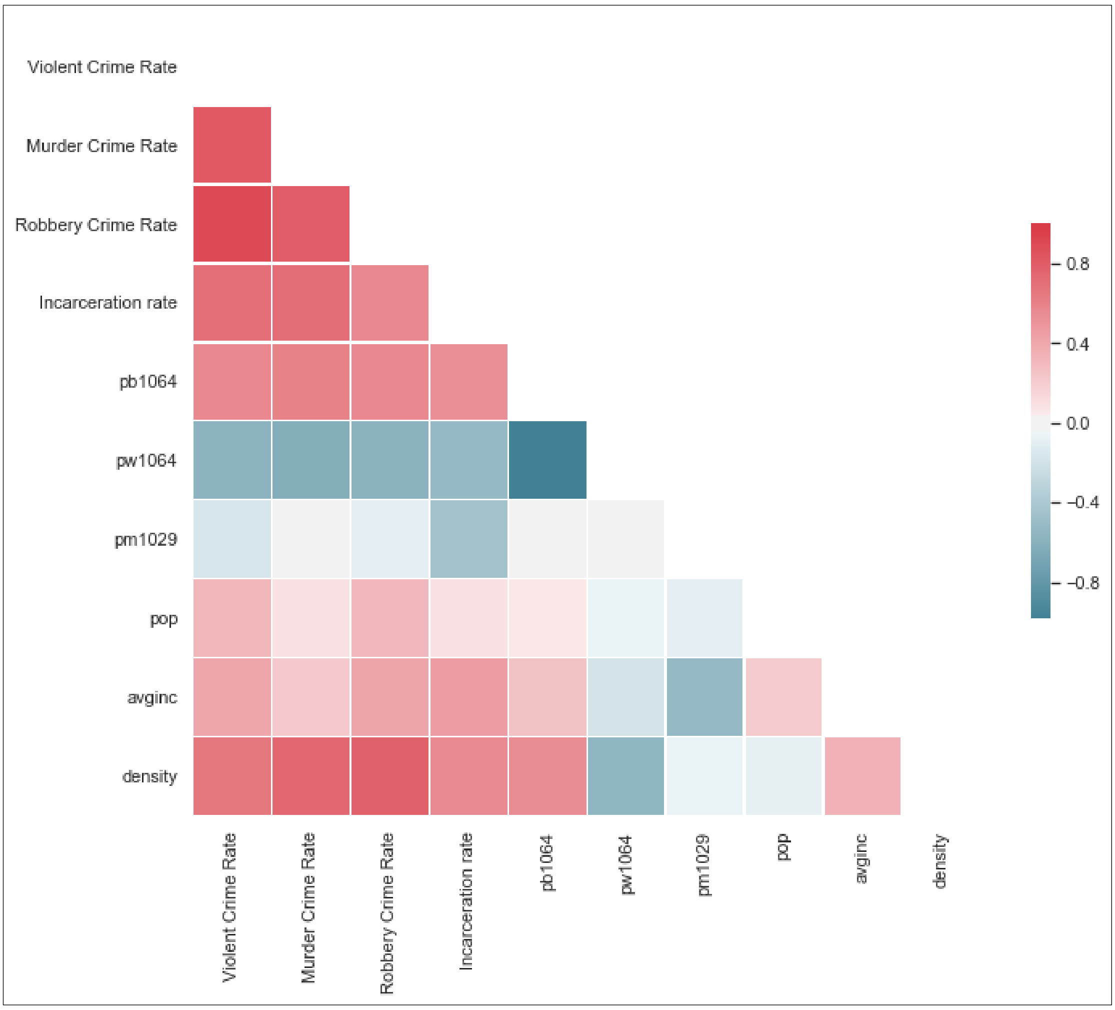
incarc_rate	
density	

avginc	Frequency 0 50 100 5 10 15 20 25 avginc
pop	Frequency 0 100 200 300 400 0 10 20 30 40 pop

pm1029	
pw1064	

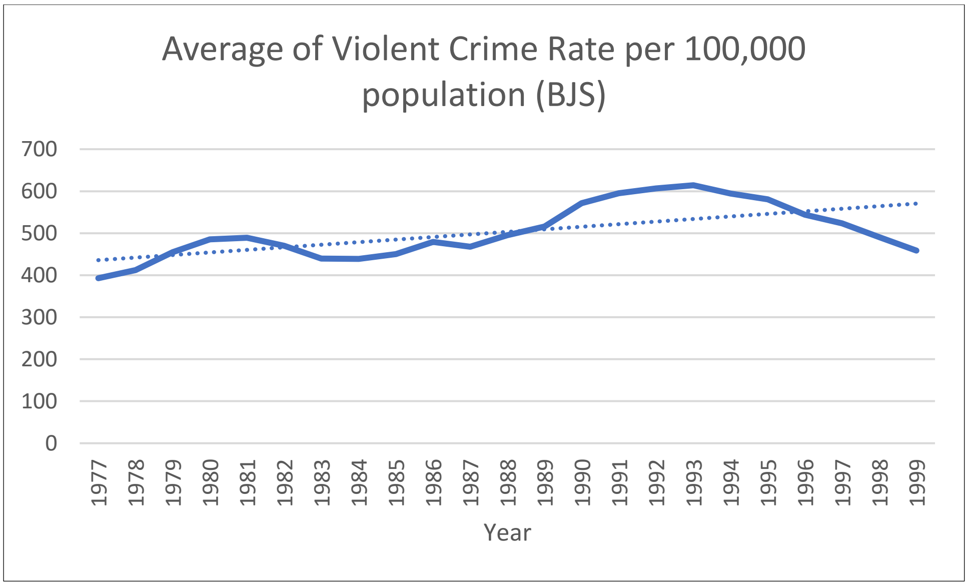
pb1064	Frequency 0 50 100 150 200 0 5 10 15 20 25 pb1064

4. Correlation matrix

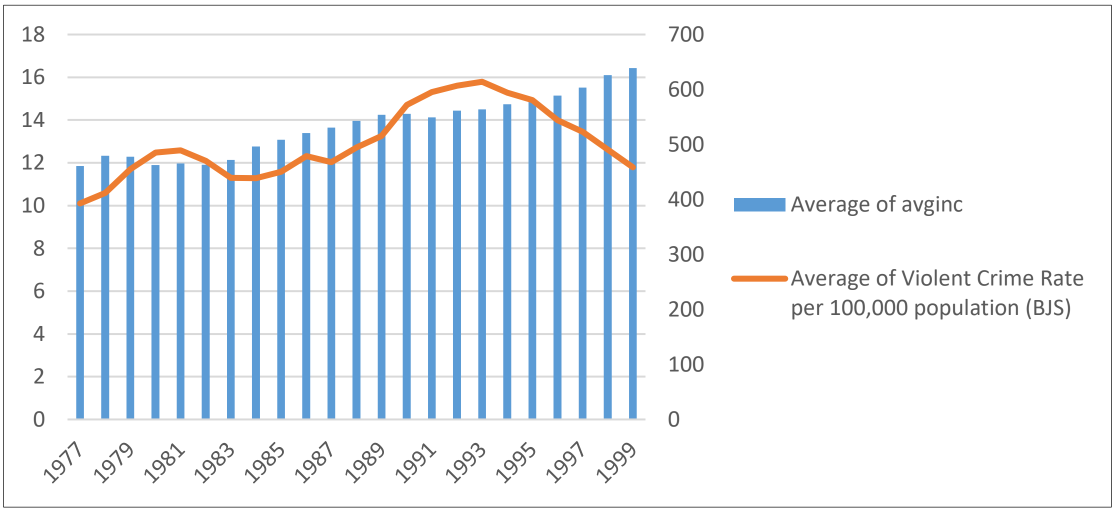


- Higher percentage of white populations tends to have lower crime rate
- Higher percentage of black populations tends to have higher crime rate
- Higher percentage of male populations tends to have lower crime rate
- Higher percentage of white populations might like to live in lower density of population area
- Higher density of population area tends to have higher crime rate

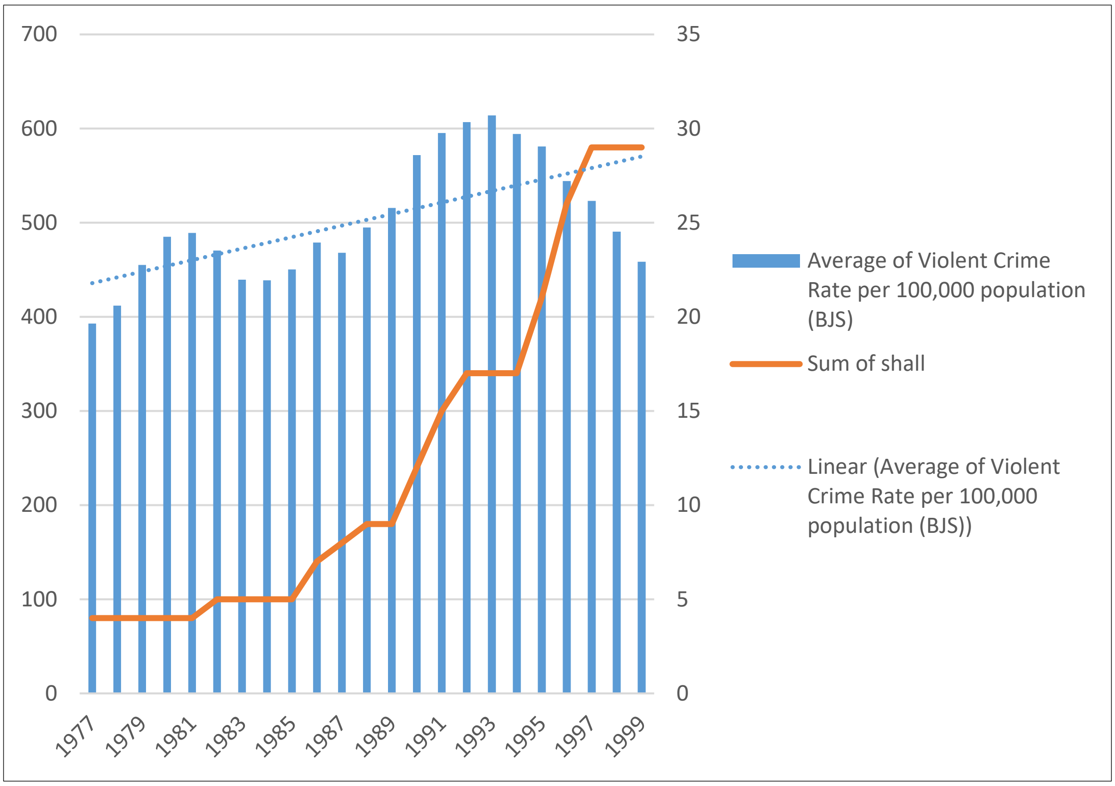
Violent crime rate across different years



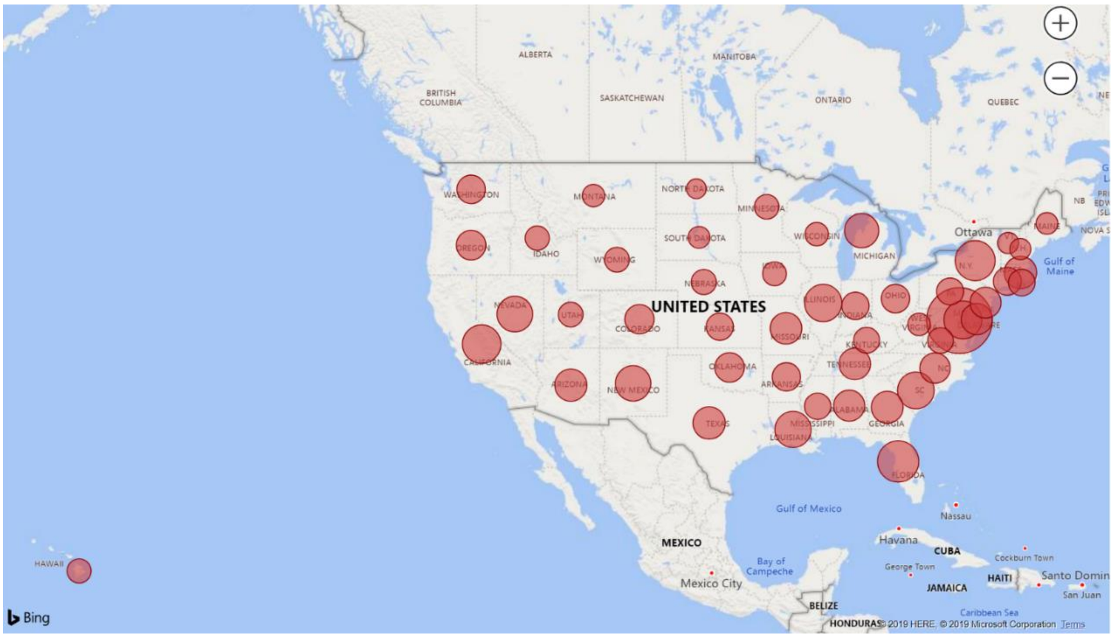
Average avginc V.S average of vio across different years



Shall-carry law V.S average of vio across different years



Violent crime rate across different states



5. Expectation

- Rob: We expect that higher robbery rate because higher violent crime rate, so we expect rob is positive.
- Mur: We expect that higher murder rate because higher violent crime rate, so we expect mur is positive.
- Shall: The argument claims that allow citizens to carry concealed guns would decrease violent crime rate, so we expect that shall is negative.
- incarc_rate: The law system would affect human behavior, which means strict laws would decrease crime rate. Therefore, we expect that this variable to be negative.
- Density: Higher population density tends to have higher crime rate.
- Avginc: People who have lower income may involve more illegal activities to get money, so we expect that avginc is negative.
- Pop: We do not have any expectation whether it is positive or negative on this variable, because when population increased, people who commit crime might also increase.
- Pm1029: Male populations are expected to involve more illegal activities, thus the crime rate increased.
- Pw1064: Contrary to black populations.
- Pb1064: Black populations are more likely to involve illegal activities, so we expected it is positive.
- Although there are more and more state with shall-carry law in effect, the overall violent crime rate still slightly increase from 1977 to 1999

6. Approach

1. Conduct exploratory data analysis and correlation matrix
2. Check Heteroscedasticity by Breusch-Pagan test
3. Use the Pooled ordinary least squares without cluster robust error
4. Use the Pooled ordinary least squares with cluster robust error
5. Check Endogenous problem by Hausman test
6. Run the entity fixed effect model with cluster robust error
7. Run the entity and time fixed effect model with cluster robust error

7. Models

i. Checking for Heteroskedasticity:

We use white test for testing heteroskedasticity. The White test fit an OLS on the residuals and all the Xs, then calculates the White statistics which follows a chi-square distribution.

The null hypothesis is that no heteroskedasticity exists and all the coefficients for the OLS is 0.

The alternative hypothesis is that there is heteroskedasticity and at least one of the coefficients for the OLS is not 0.

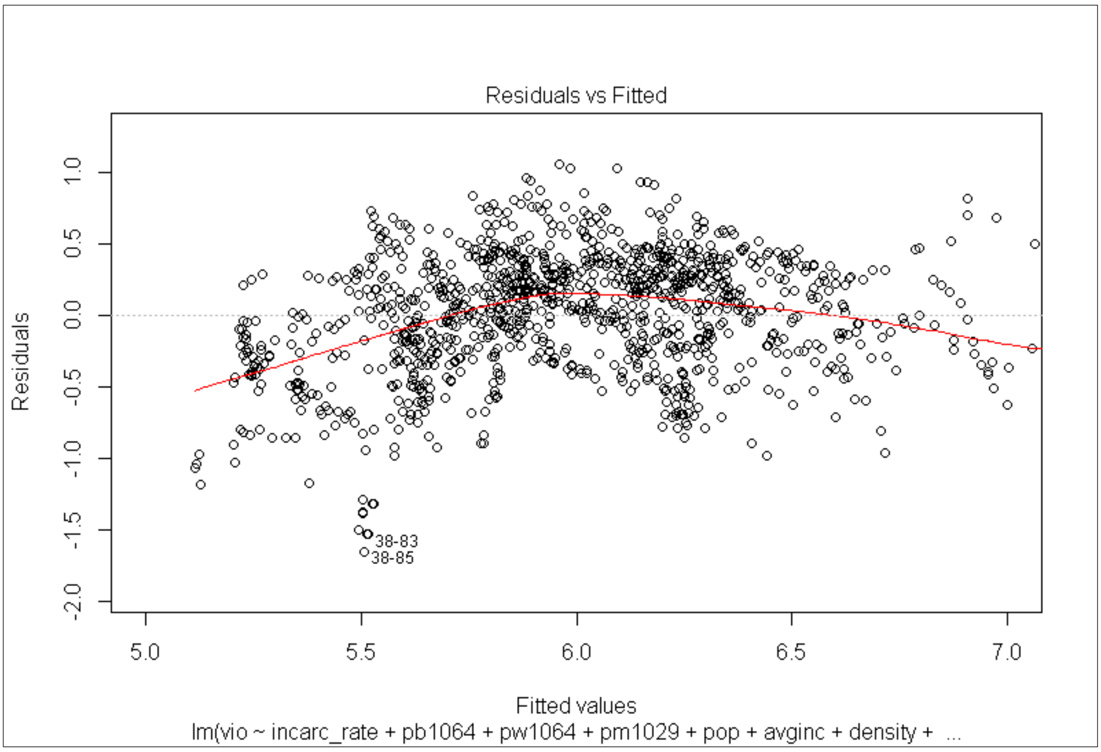
We used R for this task using the bptest function. BPtest only tests linear heteroskedasticity but it would suffice.

The BP statistics calculated is 39.952, and it is significant. We have enough evidence to reject the null hypothesis and conclude there is heteroskedasticity in this dataset.

```
       studentized Breusch-Pagan test

data:  vio ~ incarc_rate + pb1064 + pw1064 + pm1029 + pop +
 avginc +     density + shall
BP = 39.952, df = 8, p-value = 3.271e-06
```

We further confirm our findings on the scale-location plot which shows standardized residuals on fitted values which is a linear combination of all the Xs. We can clearly see on the plot that the variance of residuals changes over Y_predicted, and this shows heteroskedasticity exists. Since var(e) is not constant.



In the later part, we also fitted 2 pooled OLS without and with cluster robust errors, and the later model has significantly higher standard errors, this also is a sign of heteroskedasticity, and confirms our theory.

ii. Model 1- Pooled OLS model (without cluster robust errors)

```
. regress log_vio shall incarc_rate pb1064 pw1064 p
> m1029 pop avginc 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

------------------------------------------------------------------------------
     log_vio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       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
      pw1064 |   .0312005   .0083776     3.72   0.000     .0147636    .0476374
      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
     density |   .0266885    .013168     2.03   0.043     .0008527    .0525242
       _cons |   2.981738   .5433938     5.49   0.000     1.915598    4.047879
------------------------------------------------------------------------------
```

Key Observations:

1. The coefficient equals -0.368, which suggests that shall-issue laws reduce the violent crime rate by 36%. This is a huge effect and highly significant as well.
2. As expected all the other variables are positively contributing to crime including pop, avginc, density etc.
3. We will now run a regression to correct for heteroskedasticity.

iii. Model 2- Pooled OLS model (adjusted for cluster robust errors)

```
.
. regress log_vio shall incarc_rate pb1064 pw1064 pm1029 pop avginc density, vc
> e(cluster stateid)

Linear regression                               Number of obs     =      1,173
                                                F(8, 50)          =      62.13
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5643
                                                Root MSE          =     .42769

                                (Std. Err. adjusted for 51 clusters in stateid)
```

log_vio	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
shall	-.3683869	.113937	-3.23	0.002	-.5972361	-.1395378
incarc_rate	.0016126	.0005999	2.69	0.010	.0004076	.0028177
pb1064	.0808526	.0713875	1.13	0.263	-.0625334	.2242386
pw1064	.0312005	.03409	0.92	0.364	-.0372713	.0996723
pm1029	.0088709	.0340964	0.26	0.796	-.0596137	.0773554
pop	.0427098	.011729	3.64	0.001	.0191515	.0662681
avginc	.0012051	.0240808	0.05	0.960	-.0471626	.0495728
density	.0266885	.0414909	0.64	0.523	-.0566485	.1100255
_cons	2.981738	2.166513	1.38	0.175	-1.369831	7.333307

Key Observations:

1. We can see that the standard errors have increased to .114 from .0325 for Shall variable. This indicates the presence of heteroskedasticity in the data.
2. Other variables such as pop avginc and density also show significant increase in SE.
3. We can now suspect that there is endogeniety because there could be some qualitative factors that got added in the error term. One example could be the attitude of people towards driving in each state.
4. We will proceed with fixed effects model as that will confirm the existence of endogeniety, however, we show endogeniety using hausman test as well. We reject the null hypothesis of no endogeniety at 6% level.

iv. Hausman test for Endogeneity

shall	-.0279935	-.034406	.0064124	.
incarc_rate	.000076	.0003103	-.0002343	.0000625
pb1064	.0291862	.0788527	-.0496665	.0134343
pw1064	.0092501	.0275948	-.0183447	.0026951
pm1029	.0733254	.0411452	.0321802	.004474
pop	-.0047544	.0156487	-.0204031	.0050154
avginc	.0009587	.0090237	-.008065	.0015127
year				
78	.0585261	.0467262	.0117999	.
79	.1639486	.1446496	.0192989	.
80	.2170759	.1948856	.0221903	.
81	.2172551	.1866042	.030651	.
82	.1946328	.1540466	.0405862	.
83	.158645	.105391	.0532539	.0039485
84	.1929883	.1243402	.0686481	.007715
85	.2444764	.1631641	.0813123	.0101575
86	.3240904	.2297914	.094299	.0125222
87	.324365	.2168778	.1074873	.0148186
88	.3867412	.2647078	.1220334	.0173247
89	.4422143	.3063517	.1358626	.0196668
90	.5430478	.3594444	.1836034	.0276469
91	.5959456	.4022777	.1936679	.0291873
92	.6275171	.4204634	.2070537	.0314318
93	.6497414	.4324066	.2173348	.0330578
94	.6354187	.404894	.2305247	.0352098
95	.6276831	.384303	.2433801	.0374058
96	.5713423	.3170672	.2542751	.0393173
97	.5501153	.2845673	.265548	.0412996
98	.4932904	.2131114	.280179	.0438737
99	.4328776	.1389619	.2939157	.046243
density	-.091555	.0697575	-.1613125	.067037

```
                     b = consistent under Ho and Ha; obtained from xtreg
      B = inconsistent under Ha, efficient under Ho; obtained from xtreg

  Test:  Ho:  difference in coefficients not systematic

             chi2(29) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =       41.63
            Prob>chi2 =      0.0606
            (V_b-V_B is not positive definite)
```

We can reject the null hypothesis of no Endogeneity at ~6% significance level.

v. Model 3- Fixed Effects Model – Entity Fixed (adjusted for cluster robust errors)

```
. xtreg log_vio shall incarc_rate pb1064 pw1064 pm1029 pop avginc density, fe vce

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
     log_vio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -.0461415   .0417616    -1.10   0.275    -.1300223    .0377392
 incarc_rate |   -.000071   .0002504    -0.28   0.778    -.0005739    .0004318
      pb1064 |   .1042804   .0326849     3.19   0.002     .0386308    .1699301
      pw1064 |   .0408611   .0134585     3.04   0.004     .0138289    .0678932
      pm1029 |  -.0502725   .0206949    -2.43   0.019    -.0918394   -.0087057
         pop |   .0115247    .014224     0.81   0.422    -.0170452    .0400945
      avginc |  -.0092037   .0129649    -0.71   0.481    -.0352445     .016837
     density |  -.1722901   .1376129    -1.25   0.216    -.4486936    .1041135
       _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)
```

Key Observations

- The results change when we run the fixed effects model with states fixed.
- The absolute effect of Shall decreases to 4.6%. The effect of shall issue laws on the violent crime rate is no longer statistically significantly different from zero due to significant p values.
- It would seem that there are important omitted variable bias or unobserved Endogeneity in the specification without fixed effects.

- The regression model with fixed effects is more credible because this controls for unobserved characteristics that vary between states but that are constant over time

vi. Model 4- Fixed Effects Model – Entity and Time Fixed

```
. xtreg log_vio shall incarc_rate pb1064 pw1064 pm1029 pop avginc density 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
     log_vio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -.0279935   .0407168    -0.69   0.495    -.1097757    .0537886
 incarc_rate |    .000076   .0002079     0.37   0.716    -.0003416    .0004935
      pb1064 |   .0291862   .0495407     0.59   0.558    -.0703192    .1286916
      pw1064 |   .0092501   .0237564     0.39   0.699    -.0384659    .0569662
      pm1029 |   .0733254   .0524733     1.40   0.168    -.0320704    .1787211
         pop |  -.0047544   .0152294    -0.31   0.756    -.0353436    .0258347
      avginc |   .0009587   .0164931     0.06   0.954    -.0321688    .0340861
     density |   -.091555   .1238622    -0.74   0.463    -.3403396    .1572296
```

```
             |
        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)
------------------------------------------------------------------------------
```

Key Observations

- The results change when we run the fixed effects model with both states and time fixed.
- Effect of Shall is now 2.78% instead of 4.6% in the only state fixed model
- Coefficient is still not significantly different from 0 as observed in p values
- As per the below test, time variables are statistically significant. So we can conclude that this model is a better fit than the only state fixed model.

vii. F-Test for significance of time variables

```
. testparm i.year

 ( 1)  78.year = 0
 ( 2)  79.year = 0
 ( 3)  80.year = 0
 ( 4)  81.year = 0
 ( 5)  82.year = 0
 ( 6)  83.year = 0
 ( 7)  84.year = 0
 ( 8)  85.year = 0
 ( 9)  86.year = 0
 (10)  87.year = 0
 (11)  88.year = 0
 (12)  89.year = 0
 (13)  90.year = 0
 (14)  91.year = 0
 (15)  92.year = 0
 (16)  93.year = 0
 (17)  94.year = 0
 (18)  95.year = 0
 (19)  96.year = 0
 (20)  97.year = 0
 (21)  98.year = 0
 (22)  99.year = 0

       F( 22,    50) =   21.62
            Prob > F =    0.0000
```

Since F-value is very high, time variables are statistically significant.

8. Conclusion

- There is a large estimated effect of concealed weapons laws in pooled OLS models.
- This effect is however due to omitted variable bias and unobserved characteristics because the effect disappears when state and time effects are added.
- So the model with both time and state fixed effects is the best model.
- We can conclude that there is no significant effect of concealed weapon laws on the violent crime rate.

9. Limitations of analysis

- Even the FE estimate could have following bias:

  - It captures only within variation
  - It captures only variables which are constant

- FE estimator is still biased if the unobserved heterogeneity changing over time, and correlated with the regressors.
- This means that if some variable such as attitude of people towards driving changes over time then it will not be captured here.
- We will need more data to check for that effect but it is really difficult to get such data