BUAN 6312.001 – Applied Econometrics & Time Series Analysis

Group Project – Spring 2019

Team:

1. Abhishek Ohri
2. Ashish Mehra
3. Rohanjit Singh Virdi

Table of Contents

1.0 Introduction

Gun violence is one of the major issues ongoing in The United States. After each mass shooting, countless debates over the issue of gun control start to come up the surface and end without any significance changes in laws designed to combat the problem of gun control. One of the ways in which The United States government is trying to combat the issue of gun violence is by introducing the 'Shall – Issue' law.
A Shall-issue law is one that requires that governments issue concealed carry handgun permits to any applicant who meets the necessary criteria. The criteria are as follows:

- The applicant must be an adult.
- The applicant must not have a significant criminal record.
- The applicant must not have any history of mental illness.
- If required by law, the applicant must complete a course in firearms safety training.

If the above requirements are met, the granting authority has no discretion in the awarding of the licenses, and there is no requirement of the applicant to demonstrate "good cause".

1.1 Problem Description:

The problem we are trying to solve is to determine if the shall issue laws have any sort of affect in reducing the incidents of violent crimes. Our approach to solve this problem by analyzing historical data on crime in The United States.

2.0 Data Description

Our data is a balanced panel data consisting of 50 US states, plus the District of Columbia. The data for each state is from the year 1977 to 1999. There are a total of 51 states × 23 years, making the total number of observations to be 1,173.

The description of the variables are as follows:

1. *vio*: violent crime rate (incidents per 100,000 members of the population)
2. *rob:* robbery rate (incidents per 100,000)
3. *mur*: murder rate (incidents per 100,000)
4. *shall*: The value is '0' if the law is not in effect for that year, and '1' if the law is not in effect.
5. *incarc_rate*: incarceration rate in the state in the previous year (sentenced prisoners per 100,000 residents. Value for the previous year)
6. *density*: population per square mile of land area, divided by 1000
7. *avg_inc*: real per capita personal income in the state, in thousands of dollars.
8. *pop*: state population, in millions of people.
9. *pm1029: percent of state population that is male, ages 10 to 29*
10. *pw1064:* percent of state population that is white, ages 10 to 64
11. *pb1064:* percent of state population that is black, ages 10 to 64
12. *stateid:* ID number of states (Alabama = 1, Alaska = 2, etc.)
13. *year:* Year (1977-1999)

3.0 Exploratory Data Analysis:

After an initial look at the data and thinking upon how to get an answer to the question we are concerned with, it became clear that this is a regression problem with *vio* and *mur* being the potential dependent variables, and all the other remaining factors as the independent variables.

3.1 Descriptive Statistics

The minimum value, maximum value, 1st quadrant, 3rd quadrant, median, and mean values of each variable is given in the next page:

stateid	year	shall	vio	mur	rob	incarc_rate
Min. : 1.00	Min. :77	Min. :0.000	Min. : 47.0	Min. : 0.200	Min. : 6.4	Min. : 19.0
1st Qu.:16.00	1st Qu.:82	1st Qu.:0.000	1st Qu.: 283.1	1st Qu.: 3.700	1st Qu.: 71.1	1st Qu.: 114.0
Median :29.00	Median :88	Median :0.000	Median : 443.0	Median : 6.400	Median : 124.1	Median : 187.0
Mean :28.96	Mean :88	Mean :0.243	Mean : 503.1	Mean : 7.665	Mean : 161.8	Mean : 226.6
3rd Qu.:42.00	3rd Qu.:94	3rd Qu.:0.000	3rd Qu.: 650.9	3rd Qu.: 9.800	3rd Qu.: 192.7	3rd Qu.: 291.0
Max. :56.00	Max. :99	Max. :1.000	Max. :2921.8	Max. :80.600	Max. :1635.1	Max. :1913.0

Figure 1: Descriptive Statistics (1)

pb1064	pw1064	pm1029	pop	avginc	density
Min. : 0.2482	Min. :21.78	Min. :12.21	Min. : 0.4027	Min. : 8.555	Min. : 0.000707
1st Qu.: 2.2022	1st Qu.:59.94	1st Qu.:14.65	1st Qu.: 1.1877	1st Qu.:11.935	1st Qu.: 0.031911
Median : 4.0262	Median :65.06	Median :15.90	Median : 3.2713	Median :13.402	Median : 0.081569
Mean : 5.3362	Mean :62.95	Mean :16.08	Mean : 4.8163	Mean :13.725	Mean : 0.352038
3rd Qu.: 6.8507	3rd Qu.:69.20	3rd Qu.:17.53	3rd Qu.: 5.6856	3rd Qu.:15.271	3rd Qu.: 0.177718
Max. :26.9796	Max. :76.53	Max. :22.35	Max. :33.1451	Max. :23.647	Max. :11.102116

Figure 2: Descriptive Statistics (2)

From the initial look at the data:

1. 24.3% of the observations have initiated the 'shall issue' law.
2. The minimum and maximum violent crime rates are 47 per 100,000 population and 2921.8 per 100,000 population respectively, with the mean of 503.1 crimes per 100,000 population.
3. The minimum and maximum murder crime rate per 100,000 population is 0.2 and 80 respectively, with the mean of 7.665 murder crimes per 100,000 population.
4. The minimum and maximum robbery crime rate per 100,000 population is 6.4 and 1635.1 respectively, with the mean of 161.8 murder crimes per 100,000 population.
5. The mean incarceration rate 226.6 per 100,000 residents.
6. The mean percentage of state population that is black between ages 10 to 64 is 5.3362%.
7. The mean percentage of state population that is white between ages 10 to 64 is 62.95%.
8. The mean percentage of state population that is male between ages 10 to 29 is 16.08%.
9. The mean state population is 4.8 million people.
10. The mean real per capita personal income in the state is $13,725.

3.2 Assumptions of Linear Regression:

Before running a Ordinary Least Squares (OLS) algorithm, there are certain assumptions of linear regression that our model should satisfy in order to be the Best Linear Unbiased Estimator (BLUE). The 5 key assumptions of linear regression are:

1. **Linear Relationship:**
   The relationship between the dependent and the independent variables should be a linear relationship. This relationship can be verified with the help of scatter plots.

2. **Multivariate Normality:**
   Linear regression requires all variables to be distributed normally. This condition can be checked with the help of a histogram.

3. **Multicollinearity:**
   Multicollinearity issue refers to having a high correlation between the variables. Linear regression assumes little or no correlation between the independent variables. One way to check the issue of multicollinearity is with the help of a correlation matrix, which lists out the correlation between every variable in the dataset.

4. **Autocorrelation:**
   Linear regression requires that there is little or no serial autocorrelation of residuals in the data. Serial autocorrelation arises when the residuals are not independent of each other. Plotting the residuals of a model and/or conducting the Durbin Watson test is one way to detect the presence of serial autocorrelation of residuals.

5. **Heteroskedasticity:**
   Linear regression model assumes the error to be Homoscedastic. If the model residuals are heteroskedastic, they would exhibit some variance in the error term. Heteroskedasticity relaxes the assumption of the variance of the error term being constant.

Now, we will explore the data to see if the first 3 key assumptions of linear regression hold.

Linear Relationship:
The scatter plots depicting the relationship between the dependent variable *vio* and the relevant independent variables are given in the following pages

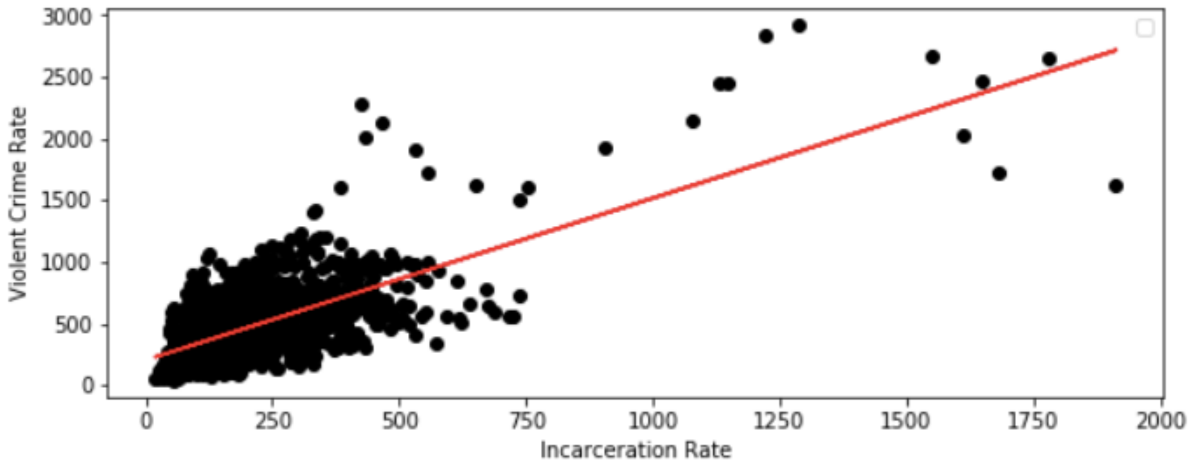

Figure 3: Relationship Between Incarceration Rate and Violent Crime Rate

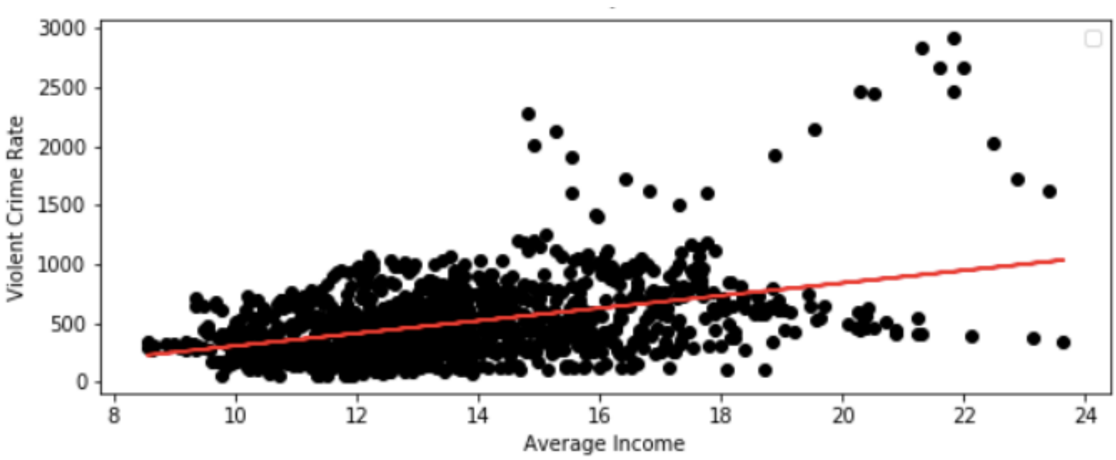

Figure 4: Relationship Between Average Income and Violent Crime Rate

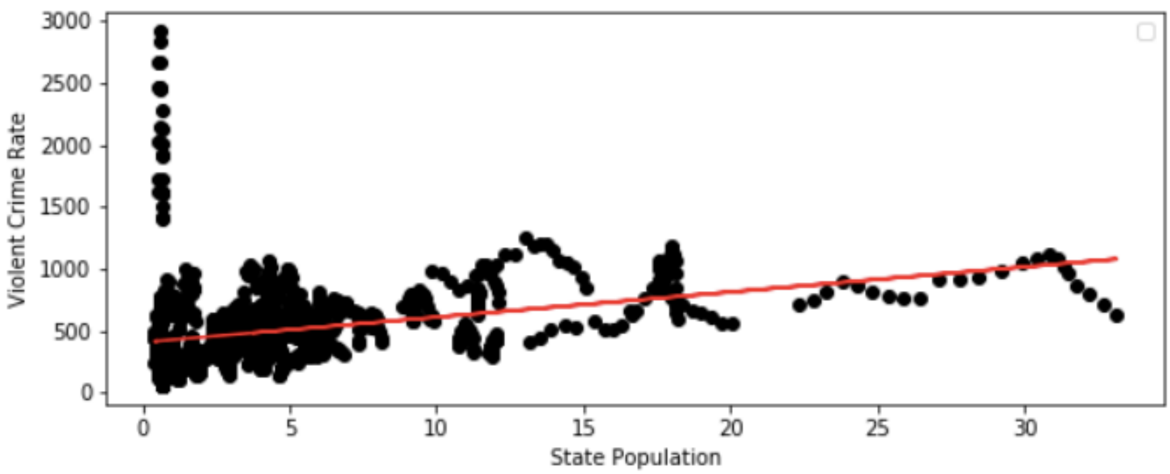

Figure 5: Relationship Between State Population and Violent Crime Rate

Figure 6: Relationship Between State Male Population and Violent Crime Rate

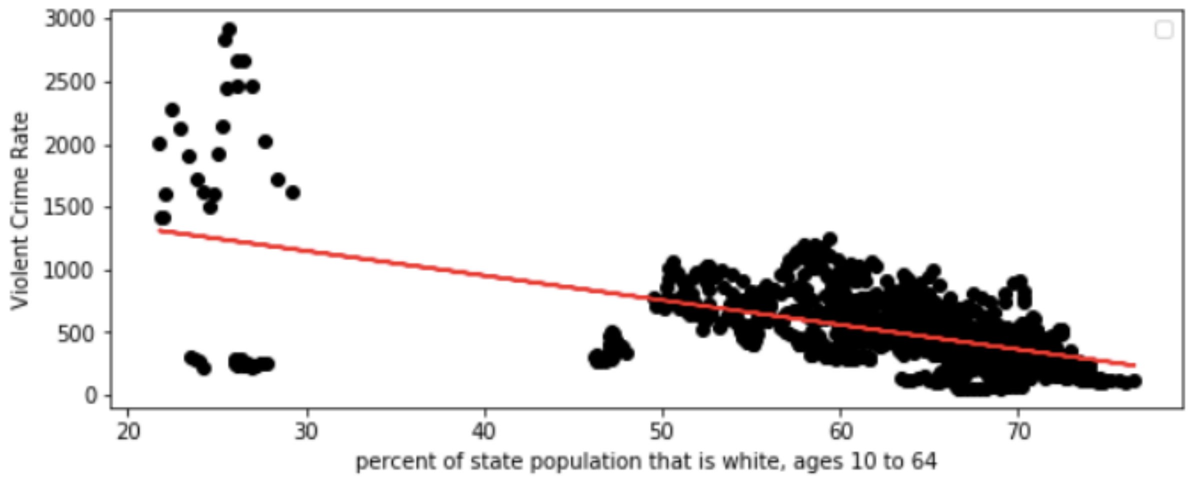

Figure 7: Relationship Between White Population and Violent Crime Rate

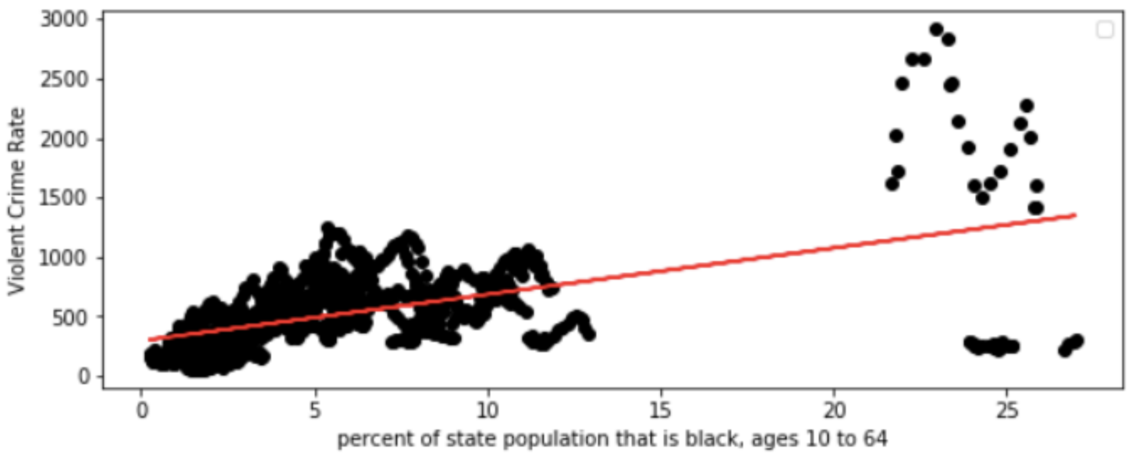

Figure 8: Relationship Between Black Population and Violent Crime Rate

From the above scatter plots, it can be inferred that the relationship between the dependent variable and the independent variables is fairly linear.

Multivariate Normality:

Now, the next step is to check how the variables are distributed with the help of a histogram.

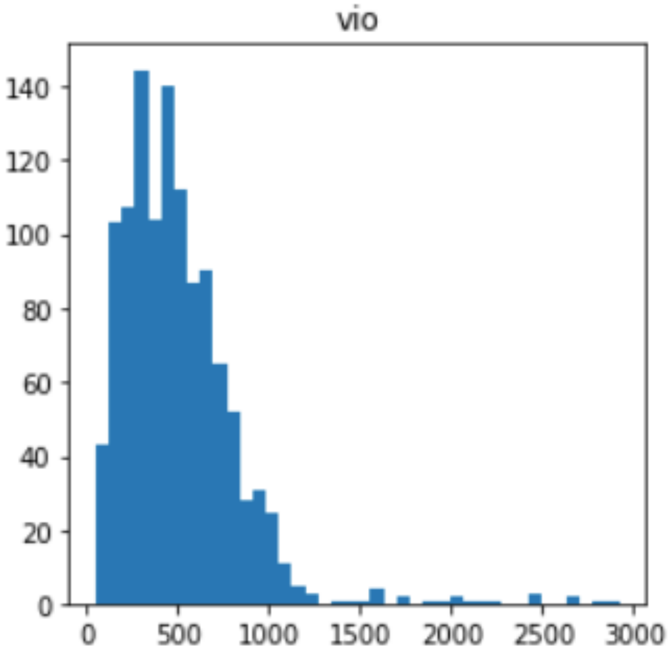

Figure 9: Distribution of Violent Crime Rate

The histogram distribution of the violent crime rates appears to be skewed to the right.

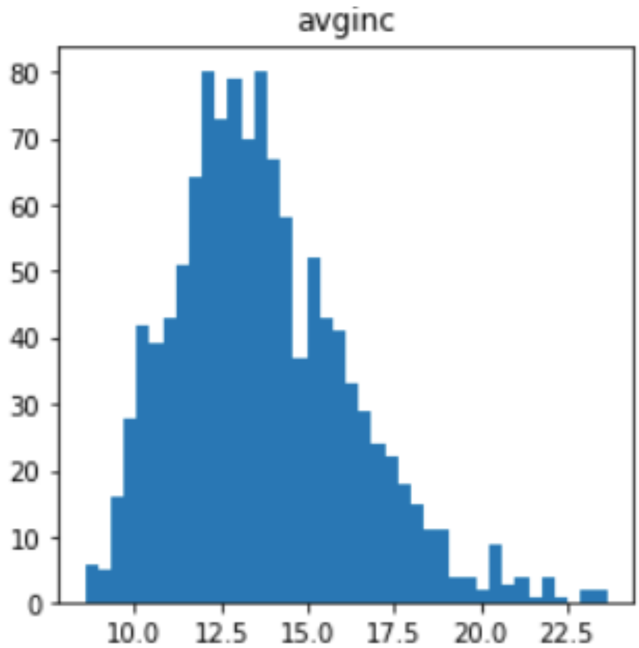

Figure 10: Distribution of Average Income

The histogram distribution of the average income appears to be skewed to the right and normally distributed.

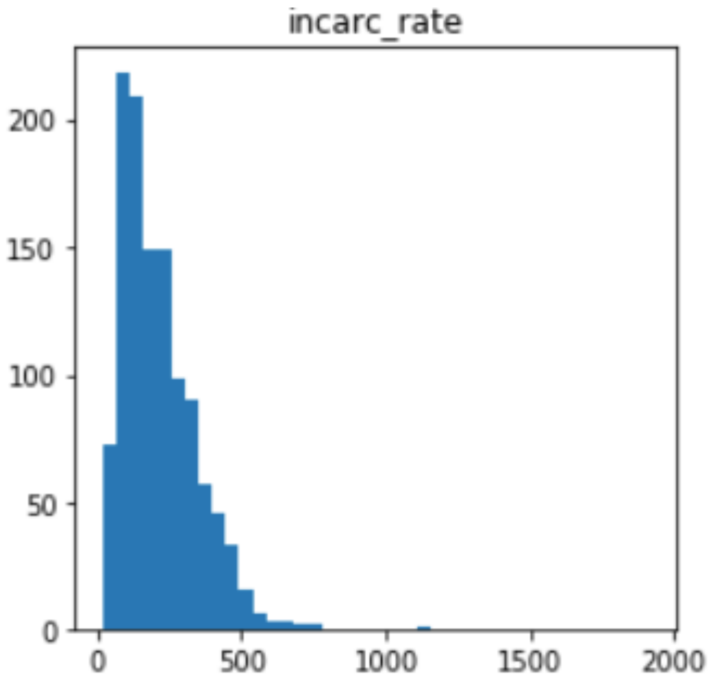

Figure 11: Distribution of Incarceration Rate

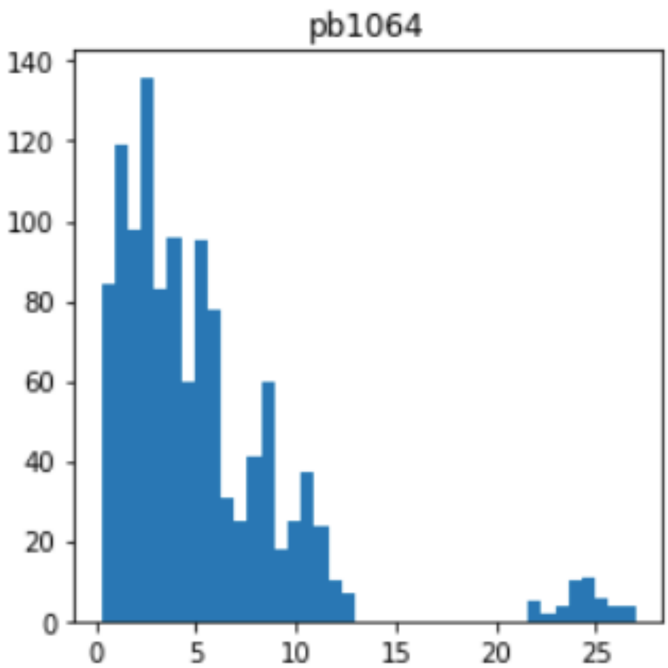

Figure 12: Distribution of Black Population

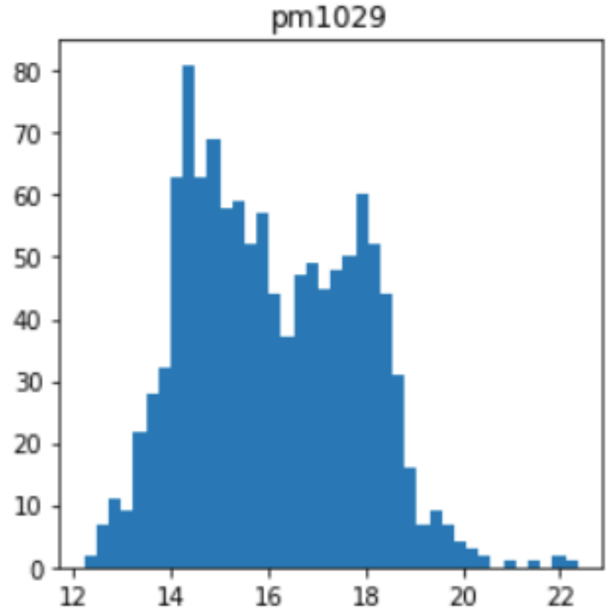

Figure 13: Distribution of Male Population

The histogram distribution of male population between ages 10 – 29 appears to be bimodal and normally distributed.

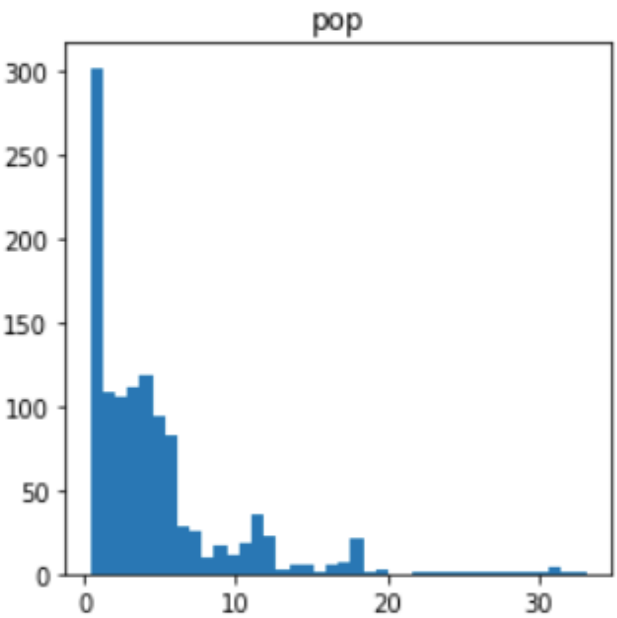

Figure 14: Distribution of State Population

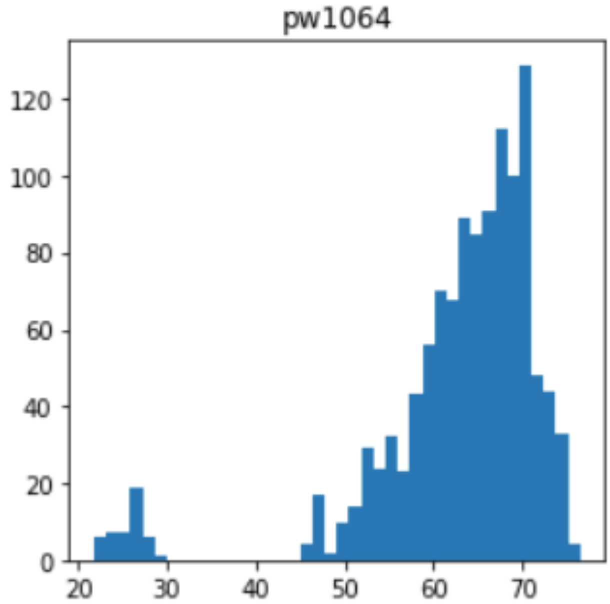

Figure 15: Distribution of White Population

The distribution of white population between ages 10 – 64 appears to be left skewed and normally distributed.
From the histogram distributions, it can be concluded that the distribution of the variables is fairly normal and skewed.

Multicollinearity:

Before running the regressions, it is vital to check the correlations between the different independent variables in order to not run into the problem of exact collinearity between the variables. In an ideal scenario, there will be little or no collinearity between the independent variables. Severe multicollinearity will lead to an increase in the variance of the coefficient estimates, which in turn, leads to the estimates being very sensitive to minor changes in the model. Correlations between the variables can be checked with the help of a correlation matrix given below:

	year	vio	mur	rob	incarc_rate	pb1064	pw1064	pm1029	pop	avginc	density	stateid	shall
year	1	0.12145	-0.0330127	-0.0141635	0.504058	0.0686072	-0.0334563	-0.865828	0.0593597	0.525232	-0.00395575	-7.92221e-18	0.379384
vio	0.12145	1	0.826509	0.907077	0.70266	0.569788	-0.573018	-0.169647	0.318966	0.407986	0.664726	-0.317035	-0.20689
mur	-0.0330127	0.826509	1	0.797606	0.709608	0.601833	-0.615368	0.0149792	0.0999223	0.220553	0.748592	-0.242765	-0.179406
rob	-0.0141635	0.907077	0.797606	1	0.56685	0.581202	-0.584192	-0.0860374	0.317193	0.414849	0.781834	-0.250699	-0.212471
incarc_rate	0.504058	0.70266	0.709608	0.56685	1	0.530776	-0.527107	-0.446318	0.0953411	0.461456	0.559313	-0.217131	0.0423743
pb1064	0.0686072	0.569788	0.601833	0.581202	0.530776	1	-0.981978	0.0161908	0.0580758	0.262694	0.543244	-0.310451	-0.183945
pw1064	-0.0334563	-0.573018	-0.615368	-0.584192	-0.527107	-0.981978	1	-0.0126023	-0.0654379	-0.191164	-0.555113	0.311228	0.212338
pm1029	-0.865828	-0.169647	0.0149792	-0.0860374	-0.446318	0.0161908	-0.0126023	1	-0.0975033	-0.527856	-0.0637151	0.00836074	-0.277211
pop	0.0593597	0.318966	0.0999223	0.317193	0.0953411	0.0580758	-0.0654379	-0.0975033	1	0.215202	-0.0780215	-0.0637422	-0.124368
avginc	0.525232	0.407986	0.220553	0.414849	0.461456	0.262694	-0.191164	-0.527856	0.215202	1	0.343284	-0.203481	-3.98932e-05
density	-0.00395575	0.664726	0.748592	0.781834	0.559313	0.543244	-0.555113	-0.0637151	-0.0780215	0.343284	1	-0.16397	-0.112609
stateid	-7.92221e-18	-0.317035	-0.242765	-0.250699	-0.217131	-0.310451	0.311228	0.00836074	-0.0637422	-0.203481	-0.16397	1	0.187303
shall	0.379384	-0.20689	-0.179406	-0.212471	0.0423743	-0.183945	0.212338	-0.277211	-0.124368	-3.98932e-05	-0.112609	0.187303	1

Figure 16: Correlation Matrix

The correlation matrix shows the correlation between each and every variable in the dataset. The independent variables, which will be used to study the effects of shall issue law are the dummy variable shall issue law, incarceration rate, percentage of black population, percentage of male population, state population, average income, and population density. The dependent variable will be the number of violent crimes per 100,000 population.
Percentage of white population will not be used in the pooled OLS model because it has a very high correlation (- 0.98) with percentage of black population. Using these 2 variables in the same model will cause the issue of multicollinearity.

Average Violent Crime Rate Over Time
The average violent crime rates for each state over the years is given below:

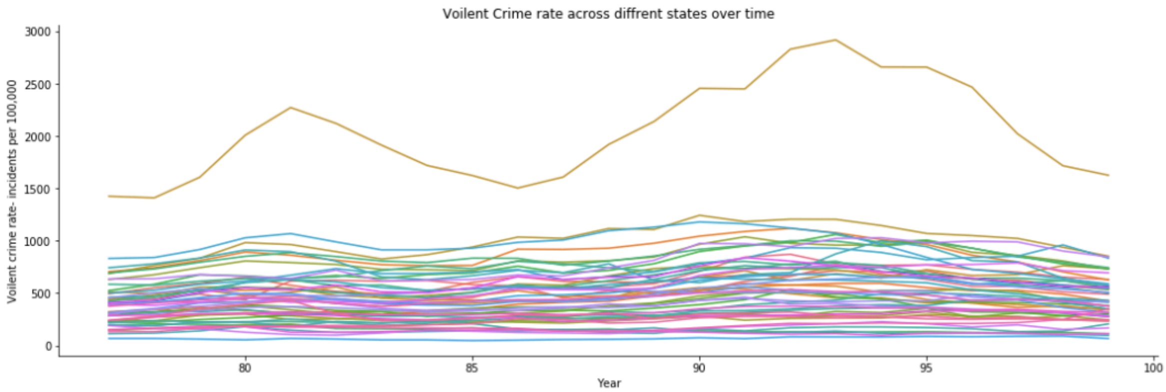

Figure 17: Average Violent Crime Rate Over Time

Average Robbery Rate Over Time
The average robbery rates for each state over the years is given below:

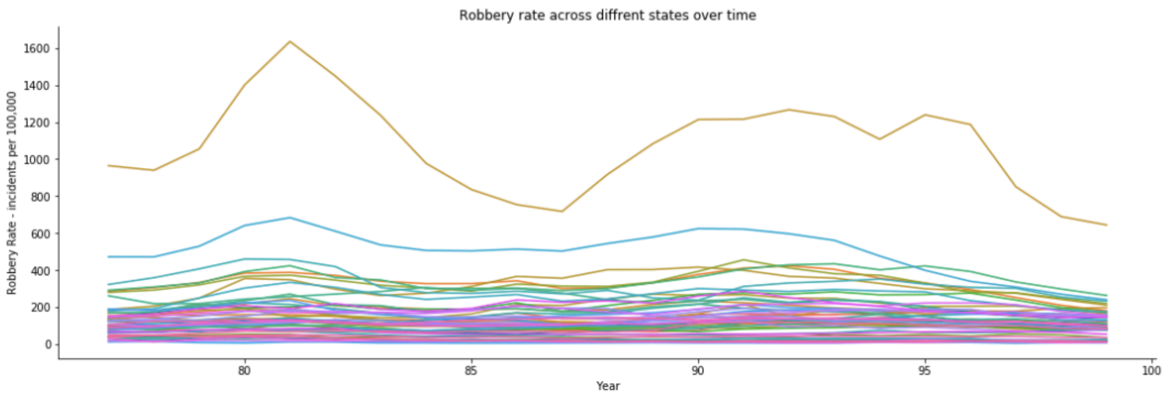

Figure 18: Average Robbery Rates Over Time

Average Murder Rate Over Time

The average murder rates for each state over the years is given below:

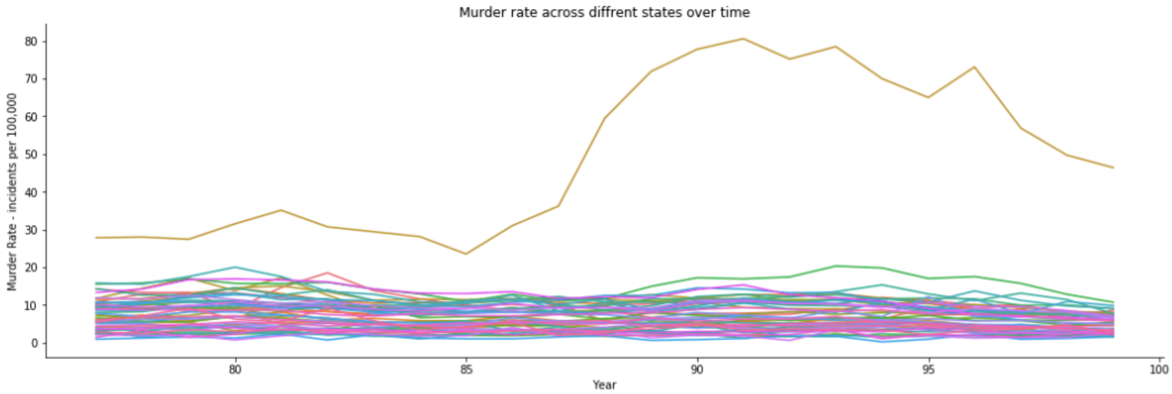

Figure 18: Average Murder Rates Over Time

The average crime, robbery, and murder rate for each state, except state 11 appears to not be fluctuating very much over the years.

4.0 Regression Models

When working with Panel data, it is possible that there be serial correlation hiding in the error term. Serial correlation of the residuals is one of the major violations of the Least Square Estimates. Pooled OLS also relaxes the assumption of the errors being constant. The consequences of proceeding with these relaxed assumptions are as follows:

- The least square estimate is still unbiased and consistent, but it is no longer the best (efficient).
- The standard errors computed by the least square estimates are biased and inconsistent i.e. incorrect. This will lead to incorrect hypotheses and confidence intervals.

In order to overcome this problem of serial correlation and heteroskedasticity, we can use cluster robust standard errors for the least square estimator. In the cluster robust standard errors, the standard errors are correct, but the estimator is still not efficient.

4.1 Regression 1 Pooled OLS

Regression 1 is the pooled OLS model. We have excluded *pw1064* in order to avoid the problem of multicollinearity.

```
model_2 <- lm(log(vio) ~ shall + incarc_rate + pb1064 + pm1029 + pop + avginc + density, data = data)
```

Figure 19: Pooled OLS

The output of the model is:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -18.94618   91.95087  -0.206   0.8368    
shall       -89.99315   12.96295  -6.942 6.39e-12 ***
incarc_rate   0.81564    0.04417  18.467  < 2e-16 ***
pb1064        6.03393    1.40196   4.304 1.82e-05 ***
pm1029       10.53630    4.15814   2.534   0.0114 *  
pop          18.35589    1.03717  17.698  < 2e-16 ***
avginc        2.63427    2.66191   0.990   0.3226    
density      93.40365    5.18119  18.027  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 176.3 on 1165 degrees of freedom
Multiple R-squared:  0.7236,	Adjusted R-squared:  0.7219 
F-statistic: 435.6 on 7 and 1165 DF,  p-value: < 2.2e-16
```

Figure 20: Pooled OLS Output

Estimation of the Model:

1. **Shall Issue Law:**
   After running the pooled OLS and controlling for other relevant factors, the shall issue law is a significant variable at the 1% significance level.
   The estimate indicates that for states in which the shall issue law is in effect, the incidents of violent crimes is approximately 90 less than the states in which the shall issue law is not in effect.

2. **Incarceration Rate:**
   After running the pooled OLS model and controlling for all other factors, the incarceration is a significant variable in predicting the number of violent crimes.
   The estimate suggests that a one unit increase in incarceration rate is going to increase the number of violent crimes increases by approximately 0.81 incidents per 100,000 population.

3. **Percentage of Population that is Black Between Ages 10 - 64:**
   After running the pooled OLS model and controlling for all other factors, pb1064 is a significant variable in predicting the number of violent crimes.

The estimate suggests that for a 1% increase in the black population between ages 10-64, the number of violent crimes increases by approximately 6.03 incidents per 100,000 population.

4. **Percentage of Population that is a Male Between Ages 10 - 29:**
   After running the pooled OLS model and controlling for all other factors, pm1029 is a significant variable in predicting the number of violent crimes.
   The estimate suggests that for a 1% increase in the male population between ages 10-29, the number of violent crimes increases by approximately 10.53 incidents per 100,000 population.

5. **Population of the State:**
   After running the pooled OLS model and controlling for all other factors, the population of a state is a significant variable in predicting the number of violent crimes.
   The estimate suggests that a one million increase in population of a state is going to increase the number of violent crimes by approximately 18 incidents per 100,000 population.

6. **Average Income:**
   After running the pooled OLS model and controlling for all other factors, the average income is not a significant variable in predicting the number of violent crimes.
   The estimate suggests that a one unit increase in the average income is going to increase the number of violent crimes by approximately 2.61 incidents per 100,000 population.

7. **Population Density:**
   After running the pooled OLS model and controlling for all other factors, the population density is a significant variable in predicting the number of violent crimes.
   The estimate suggests that a one unit increase in population density is going to increase the number of violent crimes by approximately 93 incidents per 100,000 population.

95% Confidence Interval:

	2.5 %	97.5 %
(Intercept)	-199.3540077	161.4616540
shall	-115.4264878	-64.5598117
incarc_rate	0.7289856	0.9022954
pb1064	3.2832774	8.7845915
pm1029	2.3780212	18.6945722
pop	16.3209547	20.3908292
avginc	-2.5884078	7.8569471
density	83.2381408	103.5691546

Figure 21: 95% Confidence Interval

Checking for Serial Correlation and Heteroskedasticity:

Serial correlation can be tested by plotting the residuals and conducting the Durbin-Watson test.

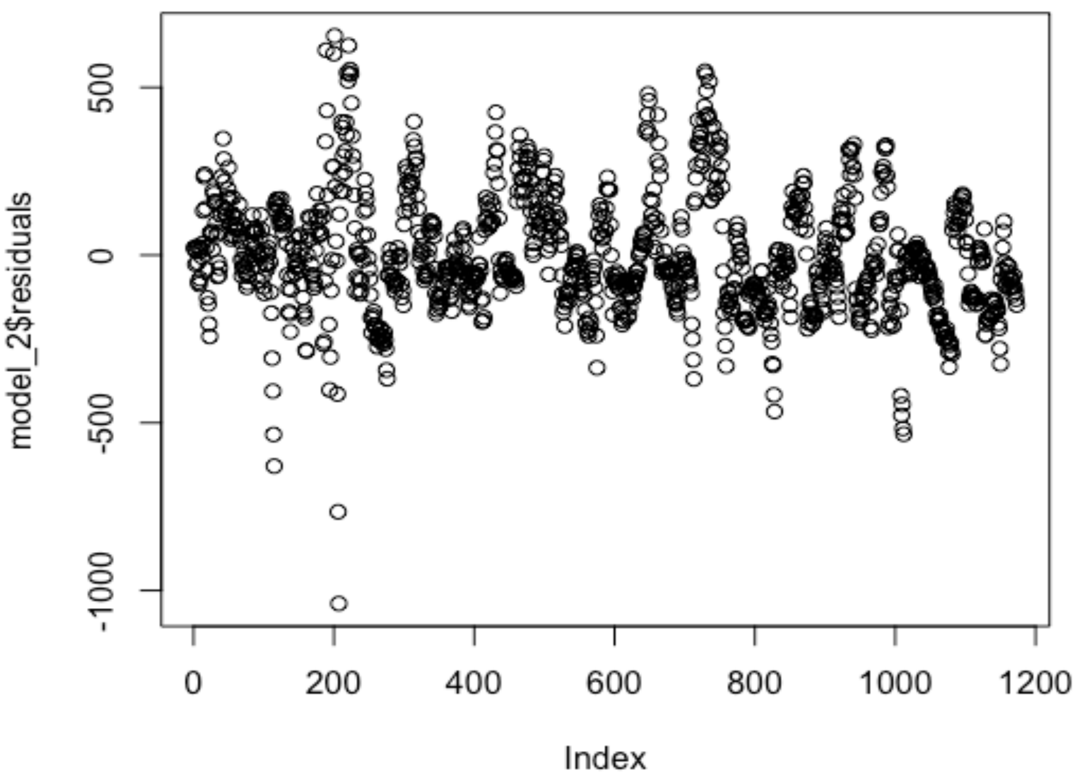

Figure 22: Residuals Plot Pooled OLS

After looking at the residuals plot, there appears to be a serial correlation between the errors, however, in order to be completely sure, we are going to conduct a Durbin-Watson test.
The **null hypotheses** of the Durbin Watson test indicates there is no serial correlation present. The **alternate hypotheses** indicates there is serial correlation present. The output of the Durbin Watson test is as follows:

```
	Durbin-Watson test

data:  model_2
DW = 0.23228, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

Figure 23: Durbin Watson Test

Since the p-value is less than 0.05, we are going to reject the null hypotheses at 5% significance level. Therefore, serial correlation does exist in our data.

Heteroskedasticity can be checked by plotting the model residuals against the fitted values and by conducting Breusch Pagan test.

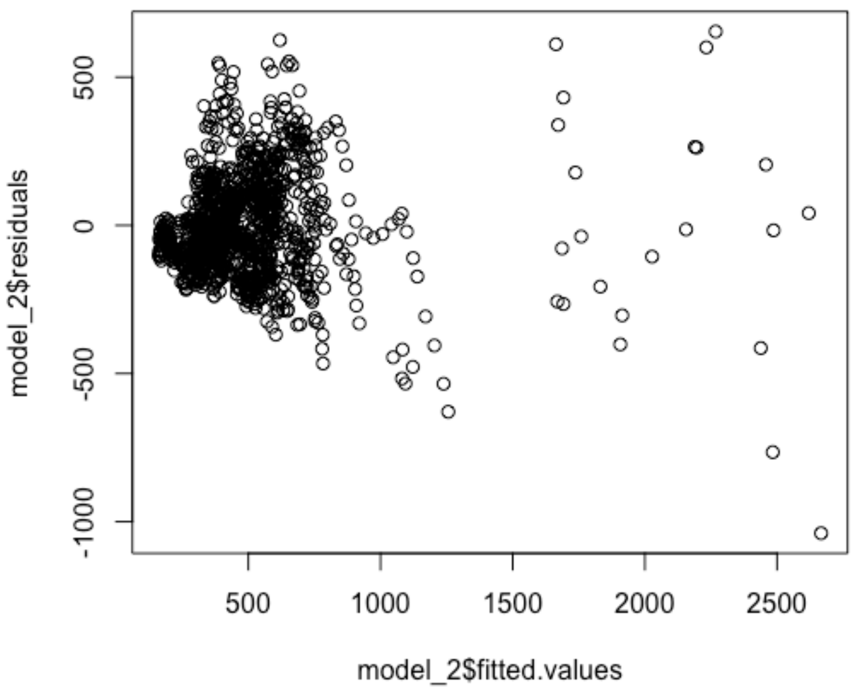

Figure 24: Fitted Values vs. Residuals Plot

Looking at the residuals, it can be concluded that the residuals do exhibit some variation and there might be heteroskedasticity present. However, in order to be certain, we are going to conduct the Breusch Pagan test for heteroskedasticity.

The **null hypotheses** of the Breusch Pagan test indicates there is no heteroskedasticity present. The **alternate hypotheses** indicates there is heteroskedasticity present. The output of the Breusch Pagan test is as follows:

```
        studentized Breusch-Pagan test

data:  model_2
BP = 265.77, df = 7, p-value < 2.2e-16
```

Figure 25: Breusch Pagan Test

Since the p-value is less than 0.05, we are going to reject the null hypotheses at 5% significance level. Therefore, heteroskedasticity does exist in our data.

Correcting Serial Correlation and Heteroskedasticity:

Using HAC estimator to correct for heteroskedasticity and serial correlation, we get the following output:

```
Coefficients:
             Estimate Std. Error t-value  Pr(>|t|)
(Intercept) -18.94618  275.39103 -0.0688  0.945163
shall       -89.99315   29.60345 -3.0400  0.002419 **
incarc_rate   0.81564    0.17158  4.7537 2.247e-06 ***
density      93.40365   14.21379  6.5713 7.503e-11 ***
avginc        2.63427    8.46956  0.3110  0.755835
pop          18.35589    4.46375  4.1122 4.194e-05 ***
pb1064        6.03393    7.62665  0.7912  0.429009
pm1029       10.53630   11.73751  0.8977  0.369552
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    130960000
Residual Sum of Squares: 36203000
R-Squared:      0.72356
Adj. R-Squared: 0.72189
F-statistic: 647.449 on 7 and 50 DF, p-value: < 2.22e-16
```

Figure 26: HAC Output

After correcting for Heteroskedasticity and serial correlation, it can be observed that variables *pb1064* and *pm1029* are not significant variables anymore.

While the standard errors are now corrected for, our estimator is still not the best.

Gourieroux, Holly, and Monfort (1982) Test

This test implements Lagrange multiplier tests of individual and/or time effects based on the results of the pooling model. It is used to test the presence of individual and time effects model

The **null hypotheses** of the Gourieroux, Holly, and Monfort test indicates there is no presence of individual and time fixed effects.

The **alternate hypotheses** indicates there is a presence of individual and time fixed effects. The output of the test is as follows:

```
        Lagrange Multiplier Test - two-ways effects (Gourieroux, Holly and
        Monfort) for balanced panels

data:  vio ~ shall + incarc_rate + density + avginc + pop + pb1064 +  ...
chibarsq = 4891.3, df0 = 0.00, df1 = 1.00, df2 = 2.00, w0 = 0.25, w1 =
0.50, w2 = 0.25, p-value < 2.2e-16
alternative hypothesis: significant effects

>
```

Figure: Lagrange Multiplier Test

After consulting the p-values, it can be concluded that we should reject the null hypotheses and go for the fixed effects model.

4.2 Fixed Effects Model

There might be some unobserved factors that are correlated with the independent variables in the model. Therefore, the independent variables capture the effect of the unobserved heterogeneity.

If heterogeneity is due to unobserved characteristics, and the observed factors are correlated with the explanatory variable, then we will have endogeneity problem. To fix this endogeneity, we will use the fixed effects model. Fixed effects controls for unobserved heterogeneity. In fixed effects, the intercepts are different for different individuals, but slope coefficients are assumed to be constant.

Individual intercepts control for individual heterogeneity – both observed and unobserved. The intercepts are called fixed effects.

4.2.1 Entity Fixed Effects

After running the entity fixed effects model, we get the following results:

```
Coefficients:
              Estimate Std. Error t-value  Pr(>|t|)
shall         -4.15670   10.76436 -0.3862  0.699456
incarc_rate    0.11319    0.05741  1.9716  0.048907 *
density     -143.32063   52.21144 -2.7450  0.006148 **
avginc        -2.71741    3.61435 -0.7518  0.452305
pop            8.86199    5.27229  1.6809  0.093070 .
pb1064        -4.62096    9.57482 -0.4826  0.629464
pm1029       -21.69618    3.88739 -5.5812 2.999e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 27: Entity Fixed Effects Model

- After running our Fixed Effects model and controlling for all other factors, the variable ‘Shall’ is not a significant variable in predicting the number of violent crimes.
  The estimate for ‘Shall’ suggests that a -4.15 decrease in violent crime rate where gun laws have been implemented, this variable is insignificant at 5% levels of significance.
- The estimate for incarc_rate is 0.11319 suggesting that with increase in incarceration rate by 1 unit, there is an increase in violent crimes by 0.11319. This variable is significant at 5 % levels.
- The estimate for density is -143.3206 suggesting with increase in population per square mile of land area by 1 unit there is a decrease in violent crimes rates by -143.32 units. This variable is significant at 5 % levels.
- The estimate for pm1029 is -21.69 suggesting with increase in percent state population of males by 1 unit there is a decrease in violent crimes rates by -21.69 units. This variable is significant at 5 % levels of significance.

- The variables avginc, pop , pb1064 are all insignificant at 5% levels of significance.

After running the entity fixed effects model with cluster robust standard errors, we get the following results:

```
Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
shall         -4.15670   17.76511 -0.2340  0.81504
incarc_rate    0.11319    0.14373  0.7875  0.43116
density     -143.32063  107.80057 -1.3295  0.18396
avginc        -2.71742    5.99519 -0.4533  0.65045
pop            8.86199    8.82306  1.0044  0.31540
pb1064        -4.62095   23.25244 -0.1987  0.84251
pm1029       -21.69618    8.58898 -2.5260  0.01167 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    13576000
```

Figure 28: Cluster Robust Standard Errors

It can be observed that only percentage of male population is a significant variable in predicting violent crimes.

4.2.2 Entity and Time Fixed Effects

In the entity and time fixed, we use time dummy variables. We are controlling for unobserved heterogeneity in the entities. The output of the model is given in the next page:

```
Coefficients:
                      Estimate Std. Error t-value  Pr(>|t|)    
shall                -2.758150  10.389593 -0.2655 0.7906956    
incarc_rate           0.224799   0.052893  4.2500 2.320e-05 ***
density             -85.435998  46.144483 -1.8515 0.0643688 .  
avginc                9.078834   3.790309  2.3953 0.0167754 *  
pop                   5.446395   4.544636  1.1984 0.2310123    
pb1064              -11.381194   8.577706 -1.3268 0.1848407    
pm1029               31.990687   6.784226  4.7155 2.723e-06 ***
as.factor(year)78    18.990197  16.991285  1.1176 0.2639649    
as.factor(year)79    65.479630  17.098943  3.8295 0.0001357 ***
as.factor(year)80   102.819078  17.259513  5.9572 3.455e-09 ***
as.factor(year)81   110.484681  17.520163  6.3061 4.145e-10 ***
as.factor(year)82    97.487449  18.097637  5.3868 8.782e-08 ***
as.factor(year)83    70.402262  18.980325  3.7092 0.0002183 ***
as.factor(year)84    72.225008  20.399201  3.5406 0.0004161 ***
as.factor(year)85    89.585246  21.923558  4.0863 4.704e-05 ***
as.factor(year)86   123.571671  23.786525  5.1950 2.441e-07 ***
as.factor(year)87   118.265575  25.648555  4.6110 4.480e-06 ***
as.factor(year)88   149.481010  27.659983  5.4042 7.987e-08 ***
as.factor(year)89   173.377423  29.550943  5.8671 5.874e-09 ***
as.factor(year)90   232.932248  31.244057  7.4552 1.821e-13 ***
as.factor(year)91   262.623377  32.663680  8.0402 2.315e-15 ***
as.factor(year)92   274.502391  34.396131  7.9806 3.659e-15 ***
as.factor(year)93   284.176477  35.561958  7.9910 3.378e-15 ***
as.factor(year)94   263.408870  36.985569  7.1219 1.926e-12 ***
as.factor(year)95   247.157593  38.386891  6.4386 1.802e-10 ***
as.factor(year)96   207.681292  39.752912  5.2243 2.092e-07 ***
as.factor(year)97   182.112002  41.047723  4.4366 1.006e-05 ***
as.factor(year)98   141.361038  42.495865  3.3265 0.0009088 ***
as.factor(year)99   104.080776  43.615493  2.3863 0.0171871 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 29: Entity and Time Fixed Effects

After running the Fixed Effects model with both Entity and Time Effects we can see there are considerable changes in our model.

- The estimate for SHALL is still insignificant at 5% levels and the effects has reduced by almost 50%, from -4.15 to -2.75.
- Incarceration rate is still significant at 5% levels of significance with the estimate at 0.2247, suggesting increase in Violent crime rate by 0.2247 with increase in Incarceration rate by 1 unit.
- The variable density which was earlier significant in our entity effects model is now insignificant at 5% levels of significance with an estimate of -85.43.
- There is a significant change in the estimate for variable pm1029, it has changed from -21.69 to 31.99 suggesting a positive relationship between the percent state

populations of males and violent crime rate. This variable is significant at 5% levels of significance.

- Avginc which was insignificant earlier is now significant at 5% levels of significance. The estimate of 9.07 suggesting increase in violent crime rate by 9.07 with increase in real per capita personal income in the state by 1000 dollars.
- Taking times effects in the model we can clearly see that all time variables are highly significant at 5% levels of significance and have a positive coefficient with an increasing trend over time with respect to our base year.

Running cluster robust standard error (HAC) for entity and time fixed model:

```
Coefficients:
                      Estimate Std. Error t-value  Pr(>|t|)
shall                 -2.75815   19.77720 -0.1395 0.8891115
incarc_rate            0.22480    0.12457  1.8046 0.0714074 .
density              -85.43600   82.70294 -1.0330 0.3018106
avginc                 9.07883    9.07202  1.0008 0.3171689
pop                    5.44639    7.87712  0.6914 0.4894487
pb1064               -11.38119   19.69618 -0.5778 0.5634927
pm1029                31.99069   16.09093  1.9881 0.0470476 *
as.factor(year)78     18.99020    5.53669  3.4299 0.0006264 ***
as.factor(year)79     65.47963    8.12107  8.0629 1.943e-15 ***
as.factor(year)80    102.81908   16.97888  6.0557 1.920e-09 ***
as.factor(year)81    110.48468   22.17462  4.9825 7.293e-07 ***
as.factor(year)82     97.48745   21.82427  4.4669 8.760e-06 ***
as.factor(year)83     70.40226   22.95934  3.0664 0.0022198 **
as.factor(year)84     72.22501   26.86649  2.6883 0.0072910 **
as.factor(year)85     89.58525   31.84332  2.8133 0.0049912 **
as.factor(year)86    123.57167   38.91752  3.1752 0.0015391 **
as.factor(year)87    118.26558   44.49342  2.6580 0.0079742 **
as.factor(year)88    149.48101   50.10666  2.9833 0.0029153 **
as.factor(year)89    173.37742   55.54004  3.1217 0.0018455 **
as.factor(year)90    232.93225   61.09900  3.8124 0.0001453 ***
as.factor(year)91    262.62338   64.69558  4.0594 5.271e-05 ***
as.factor(year)92    274.50239   70.16541  3.9122 9.710e-05 ***
as.factor(year)93    284.17648   71.79113  3.9584 8.033e-05 ***
as.factor(year)94    263.40887   72.48139  3.6342 0.0002919 ***
as.factor(year)95    247.15759   72.27871  3.4195 0.0006505 ***
as.factor(year)96    207.68129   73.86551  2.8116 0.0050175 **
as.factor(year)97    182.11200   75.38346  2.4158 0.0158635 *
as.factor(year)98    141.36104   79.76697  1.7722 0.0766440 .
as.factor(year)99    104.08078   83.16762  1.2515 0.2110353
---
```

Figure 30: Cluster Robust Standard Errors

Average income and density are not significant variables anymore.

4.3 Random Effects

- We can also choose a model where the individual differences between states are treated and captured in a random fashion. This model is known as the Random Effects model.
- However, we saw that the data is not collected using random sampling. Therefore, we should not implement the Random Effects model.
- This leaves us with Fixed effects model with entity and time fixed effects as our best model for understanding impact of shall law on violent crime rate.

However, we are going to run the random effects model in order to prove there is endogeneity by running the Hausman test.
The **null hypotheses** of the Hausman test indicates there is no endogeneity present. The **alternate hypotheses** indicates there is endogeneity present. The output of the Hausman test is as follows:

```
> phtest(model41,model5)

	Hausman Test

data:  vio ~ shall + incarc_rate + density + avginc + pop + pb1064 +  ...
chisq = 1490.2, df = 7, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

Figure 31: Hausman Test

From the output of the Hausman test, it can be concluded that there is indeed endogeneity present, and hence, we will go with the fixed effects model.

Limitations of Using Entity and Time Fixed Effects Model:

- Important variables that vary across states and over time might be omitted, and that can cause omitted variable bias.
- Inclusion of incarceration rate as one of the explanatory variables could cause a major simultaneous causality bias. Increased incarceration rate will cause fear in people, and that could reduce the crime rate. If the crime rate were to increase due to some unknown factors, it will force the government to enforce stricter laws, which could lead to an increased incarceration rate.

5.0 Conclusion

Qualities that vary over entities or time are not accounted in the pooled OLS models. Coefficients of 'shall' in Pooled OLS model are very biased and seem to be an overestimation of true parameters. This could be caused by omitted variable bias.
We know that there is unobserved heterogeneity being introduced due to variables so Time and Entity Fixed model is the most appropriate model for our balanced panel data.

Looking at the output of all the results, the most appropriate model appears to be entity and time fixed estimator because it accounts for unobserved heterogeneity and time effects. Also, introduction of shall-issue law has not had a major impact on the crime rates over the years. The United States government should probably look into alternative methods to reduce the crime rate.