BUAN 6312.002 – Applied Econometrics and Time Series Analysis

Do more guns lead to less crime?

Neelraj Khawale (NVK210000)

Rahul Mukherjee (RXM210065)

Table of Contents

1. Introduction

1.1 Background

The ongoing discourse surrounding gun control policies, particularly the implementation of shall-issue laws, has been a subject of substantial debate within the United States. Shall-issue laws, which grant citizens the right to carry concealed handguns, have ignited discussions on public safety, crime prevention, and individual rights. This study delves into the historical crime data of the United States to critically examine the potential impact of shall-issue laws on crime rates.

1.2 Objectives of the Study

The primary aim of this research is to investigate the relationship between the presence of shall-issue laws and crime rates. Through a rigorous analysis of historical data spanning from 1977 to 1999, we seek to understand the nuances of this relationship, exploring temporal trends, demographic variations, and statistical associations.

1.3 Research Question

The central question guiding this study is: "Do shall-issue laws reduce crime rates?" This inquiry prompts an exploration into the complex dynamics between firearm regulations and various crime indicators, including violent crime, robbery, and murder rates.

1.4 Significance of the Study

Understanding the implications of shall-issue laws on crime rates holds significant relevance for policymakers, law enforcement agencies, and the general public. The findings of this study aim to contribute valuable insights to the ongoing national conversation on gun control, informing evidence-based decision-making and fostering a deeper understanding of the societal impacts of firearm policies.

2. Economic Theory

There are various economic concepts at play at the heart of this complex topic:

Rational Choice Theory

Shall issues laws can potentially alter the perceived cost benefit analysis for individuals who are plotting to commit a violent crime.

Proponents of the shall-issues law often cite defensive gun use. They argue that armed citizens can defend themselves in dangerous situations potentially preventing violent crimes.

Public Choice Theory

One can also argue that policies such as the shall-issues law are pushed forward by interest groups like the NRA to serve their own interests, potentially impacting policy decisions irrespective of their impact on the rate of violent crimes.

We can hypothesize how certain variables in the given dataset might behave based on a few economic models:

1. Deterrence Hypothesis: Shall-issues laws might act as a deterrent to crime. States with these laws might experience lower violent crime rates, robbery rates, and murder rates due to the perceived risk of encountering an armed victim.
2. Socioeconomic Factors: Economic factors and demographics are also important factors that influence crime rate. Higher per capita income might be associated with lower crime rates. Similarly high population densities and population might be associated with a high rate of crime.
3. Demographic Factors: Certain demographic populations such as the percentage of young males, black or white people within a certain age group also might have an impact on crime rates.

2.1 Economic Model:

Assumptions:

- Shall-carry laws potentially increase the likelihood of encountering an armed victim for criminals.
- Criminals weigh the costs and benefits of committing crimes, considering potential deterrent effects.

Variables and Relationships:

Dependent Variables (Crime Rates):

- Violent Crime Rate (vio)
- Robbery Rate (rob)
- Murder Rate (mur)

Independent Variable:

- Shall-carry law presence (shall)

Control Variables:

- Incarceration Rate (incarc_rate)
- Population Density (density)
- Real Per Capita Income (avginc)
- Population Demographics (pop, pm1029, pw1064, pb1064)

2.2 Proposed Model:

$$vio = b0 + b1\ shall + b2\ incarc\_rate + b3\ density + b4\ avginc + b5\ pop + b6\ pm1029 + b7\ pw1064 + b8\ pb1064 + e$$

Where:

- b1 represents the impact of shall-carry laws on the violent crime rate.
- e is the error term capturing unobserved factors influencing crime rates.

2.3 Econometric Model:

To estimate the coefficients in the economic model, an econometric model like a panel data regression can be employed. Given the panel nature of the data (51 states over 23 years), a fixed-effects or random-effects panel regression can be applied.

$$Vio_{it} = B0 + B1\ shall_{it} + B2\ incarc\_rate_{it} + B3\ density_{it} + B4\ avginc_{it} + B5\ pop_{it} + B6\ pm1029_{it} + B7\ pw1064_{it} + B8\ pb1064_{it} + e_{it}$$

Where:

- Vio_{it} represents the violent crime rate for state i in year t.
- e_{it} captures the error term for each observation.

Applying appropriate econometric techniques, such as panel data regression, can help estimate the impact of shall-carry laws while controlling for other factors influencing crime rates. This estimation will allow us to evaluate the significance and direction of the relationship between shall-carry laws and crime rates, considering other relevant variables.

3. Exploratory Data Analysis

3.1 Data Overview

Our dataset encompasses 51 states, including the District of Columbia, and spans 23 years from 1977 to 1999. With 1173 observations, this dataset provides a comprehensive canvas for exploring the intricate relationships between crime rates and shall-issue laws.

3.1.1 Data Source

The data utilized in this analysis is a balanced panel, capturing essential variables such as violent crime rate, robbery rate, murder rate, shall-issue law status, incarceration rate, population density, average income, demographic percentages, state identifiers, and the temporal aspect of the data.

3.1.2 Data Preparation

Ensuring data integrity, the dataset underwent meticulous preparation using the haven package in R. Initial checks were conducted for the structure and potential missing values to guarantee the reliability of subsequent analyses.

3.2 Descriptive Statistics - Overview of Variables

A comprehensive exploration of the dataset's variables was initiated through descriptive statistics. This section provides an insightful overview of the central tendencies and dispersions within the dataset.

```
> summary(guns)
      year            vio              mur              rob          incarc_rate        pb1064          pw1064          pm1029     
 Min.   :77   Min.   :  47.0   Min.   : 0.200   Min.   :   6.4   Min.   :  19.0   Min.   : 0.2482   Min.   :21.78   Min.   :12.21  
 1st Qu.:82   1st Qu.: 283.1   1st Qu.: 3.700   1st Qu.:  71.1   1st Qu.: 114.0   1st Qu.: 2.2022   1st Qu.:59.94   1st Qu.:14.65  
 Median :88   Median : 443.0   Median : 6.400   Median : 124.1   Median : 187.0   Median : 4.0262   Median :65.06   Median :15.90  
 Mean   :88   Mean   : 503.1   Mean   : 7.665   Mean   : 161.8   Mean   : 226.6   Mean   : 5.3362   Mean   :62.95   Mean   :16.08  
 3rd Qu.:94   3rd Qu.: 650.9   3rd Qu.: 9.800   3rd Qu.: 192.7   3rd Qu.: 291.0   3rd Qu.: 6.8507   3rd Qu.:69.20   3rd Qu.:17.53  
 Max.   :99   Max.   :2921.8   Max.   :80.600   Max.   :1635.1   Max.   :1913.0   Max.   :26.9796   Max.   :76.53   Max.   :22.35  
      pop              avginc          density             stateid          shall       
 Min.   : 0.4027   Min.   : 8.555   Min.   : 0.000707   Min.   : 1.00   Min.   :0.000  
 1st Qu.: 1.1877   1st Qu.:11.935   1st Qu.: 0.031911   1st Qu.:16.00   1st Qu.:0.000  
 Median : 3.2713   Median :13.402   Median : 0.081569   Median :29.00   Median :0.000  
 Mean   : 4.8163   Mean   :13.725   Mean   : 0.352038   Mean   :28.96   Mean   :0.243  
 3rd Qu.: 5.6856   3rd Qu.:15.271   3rd Qu.: 0.177718   3rd Qu.:42.00   3rd Qu.:0.000  
 Max.   :33.1451   Max.   :23.647   Max.   :11.102116   Max.   :56.00   Max.   :1.000  
```

Let's explore the findings:

- **Violent Crime Rate (vio):**
  - The least we've seen violent crime is 47 incidents for every 100,000 people, while the most is 2921.8.
  - Half of the observations fall below 443, which is the middle point.
  - On average, the violent crime rate is around 503.1.
- **Murder Rate (mur):**
  - Murder rates go from as low as 0.2 to as high as 80.6 for every 100,000 people.
  - The middle number is 6.4, and on average, it's about 7.665.
- **Robbery Rate (rob):**
  - Robbery rates go from 6.4 to a whopping 1635.1 for every 100,000 people.
  - The middle point is 124.1, and on average, it's about 161.8.
- **Demographic Percentages (pb1064, pw1064, pm1029):**
  - These percentages show how many people in the population are black, white, and male aged 10 to 29.

    - For example, the middle percentage of the population that is black is 4.0262%, and for white, it's 65.06%.
- **Average Income (avginc):**
    - In terms of money, people's average income goes from $8,555 to $23,647.
    - The middle income is $13,402, and on average, it's about $13,725.
- **Population Density (density):**
    - How crowded a place is varies a lot, from 0.000707 to 11.102116.
    - This means there's a big difference in how spread-out people are.
- **Shall-Issue Law (shall):**
    - Some states have this law (1) and some don't (0).
    - On average, about 24.3% of states have this law.
- **State Population (pop):**
    - How many people live in a state is different everywhere, from 0.4027 million to 33.1451 million.

3.2.1 Central Tendencies

Measures of central tendency, including mean, median, and mode, were computed to offer a robust understanding of the typical values within the dataset.

For most variables, the mean is slightly higher than the median, indicating a slight right skewness in the distributions.

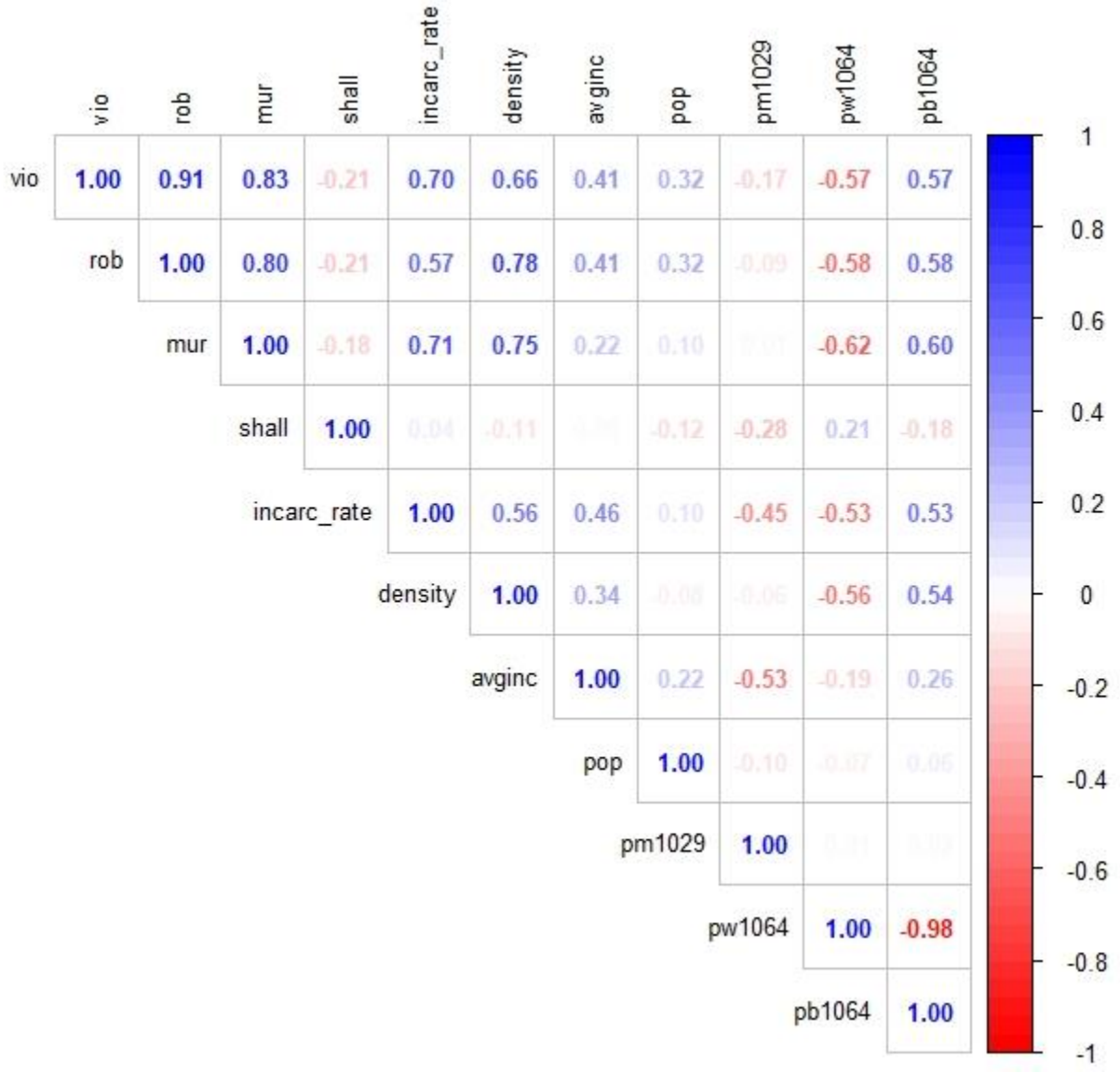
3.2.2 Dispersion Measures

Dispersion measures, such as variance and standard deviation, were calculated to gauge the extent of variability present in the dataset.

The considerable range in crime rates, demographic percentages, and socio-economic variables suggests notable heterogeneity among states.

3.2.3 Correlation Analysis - Correlation Matrix

A detailed correlation matrix was generated to unravel the intricate relationships between key variables. This matrix was subsequently visualized using the corrplot package to enhance interpretability.



Interpretation:

IMPORTANT: In the correlation matrix, we observe a strong correlation among the variables for violent crime (vio), murder (mur), and robbery (rob). We believe

that murder and robbery are subsets of violent crime and given this strong correlation, we opt to focus our analysis specifically on the violent crime rate (vio) as our target variable. This allows us to concentrate on a comprehensive understanding of violent crime without redundancy from subsets of closely related variables.

Crime Rates and Shall-Issue Law

- **Violent Crime (vio) and Shall-Issue Law (shall):**
    - There is a negative correlation of approximately -0.21 between the presence of shall-issue laws and the violent crime rate. This suggests that states with shall-issue laws tend to have slightly lower violent crime rates.
- **Robbery (rob) and Shall-Issue Law (shall):**
    - Similarly, there is a negative correlation of around -0.21 between the presence of shall-issue laws and the robbery rate. This implies a tendency for states with shall-issue laws to exhibit lower robbery rates.
- **Murder (mur) and Shall-Issue Law (shall):**
    - The negative correlation of about -0.18 suggests a weak tendency for states with shall-issue laws to have slightly lower murder rates.

Crime Rates and Other Factors

- Strong positive correlations are observed between violent crime and incarceration rate (0.70) and population density (0.66). This suggests that higher violent crime rates tend to align with higher incarceration rates and population density.
- Robbery rates also exhibit positive correlations with incarceration rate (0.57) and population density (0.78).
- Murder rates are positively correlated with incarceration rate (0.71) and population density (0.75).

Shall-Issue Law and Demographic Factors

- **Shall-Issue Law (shall) and Demographic Factors:**
    - Shall-issue laws show weak negative correlations with demographic factors such as the percentage of the population that is male ages 10 to 29 (pm1029), white ages 10 to 64 (pw1064), and black ages 10 to 64 (pb1064).
- **Demographic Factors:**
    - There is a strong negative correlation of approximately -0.98 between the percentage of the population that is white ages 10 to 64 (pw1064) and black ages 10 to 64 (pb1064). This indicates that it is the percentage of the same population. As a result, there is a high degree of inverse relationship between these demographic categories.

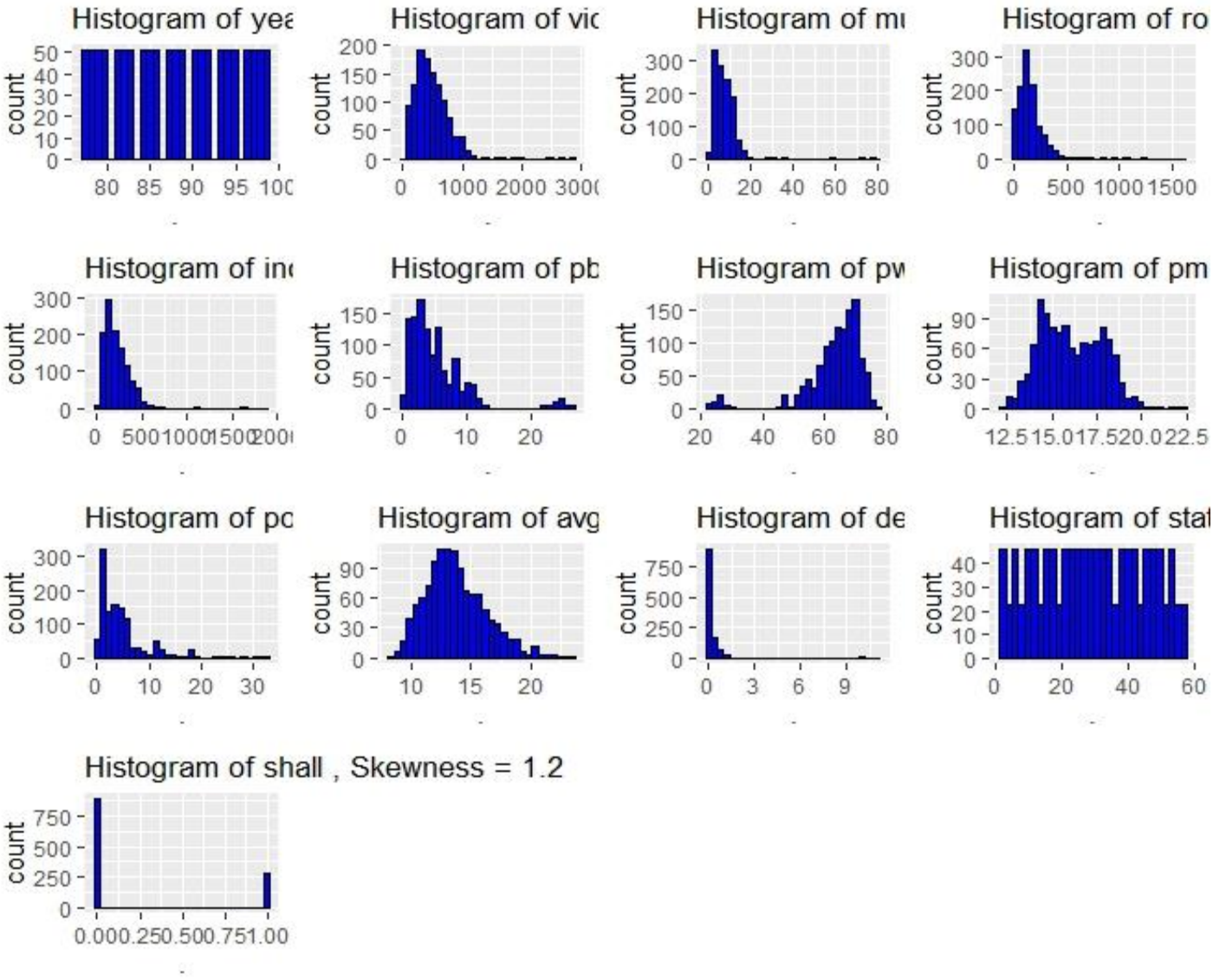
  NOTE: Because of this, we choose to use only one of these variables (pb1064) in our analysis to address the issue of multicollinearity. We aim to mitigate the impact of multicollinearity and adhere to key econometric assumptions, ensuring the robustness of our analysis.

Other Correlations

- **Population (pop) and Socio-Economic Factors:**
    - The positive correlation between population and socio-economic factors such as incarceration rate and density indicates that more populous states tend to have higher incarceration rates and population density.
- **Income (avginc) and Demographic Factors:**
    - Income (avginc) shows positive correlations with demographic factors, indicating that higher-income states tend to have higher percentages of the population that is white (pw1064) and lower percentages that are black (pb1064).

3.2.4 Skewness Values

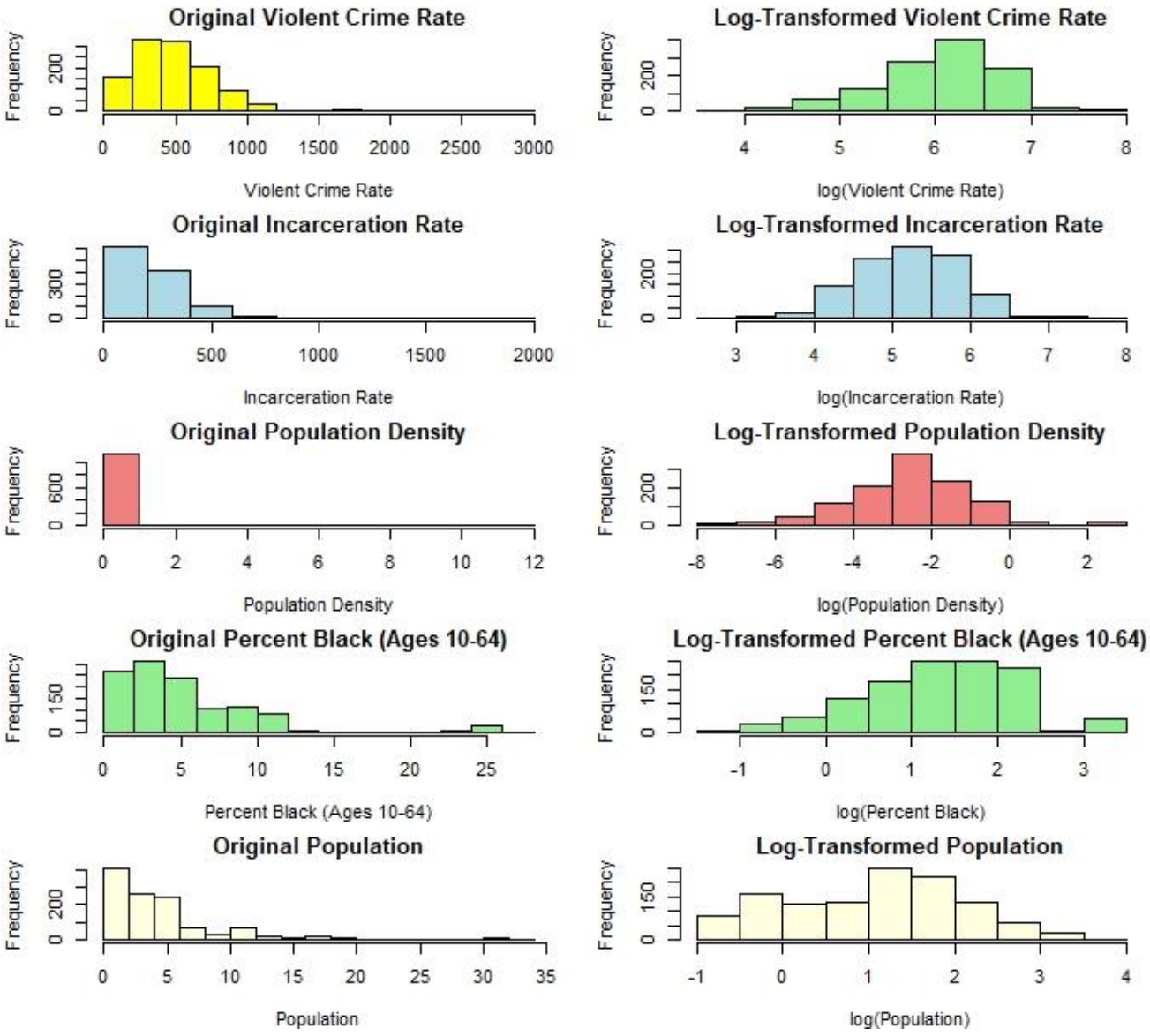
Skewness values were computed for each variable, providing insights into the asymmetry of the distribution and the presence of potential outliers.



Positive skewness in some variables like violence rate, incarceration rate, density, pb1064, and population, indicates a rightward tail in the distribution, suggesting the presence of states with higher-than-average values.

3.2.5 Handling Skewness:

We will use log transformations on violence rate, incarceration rate, density, pb1064, population.



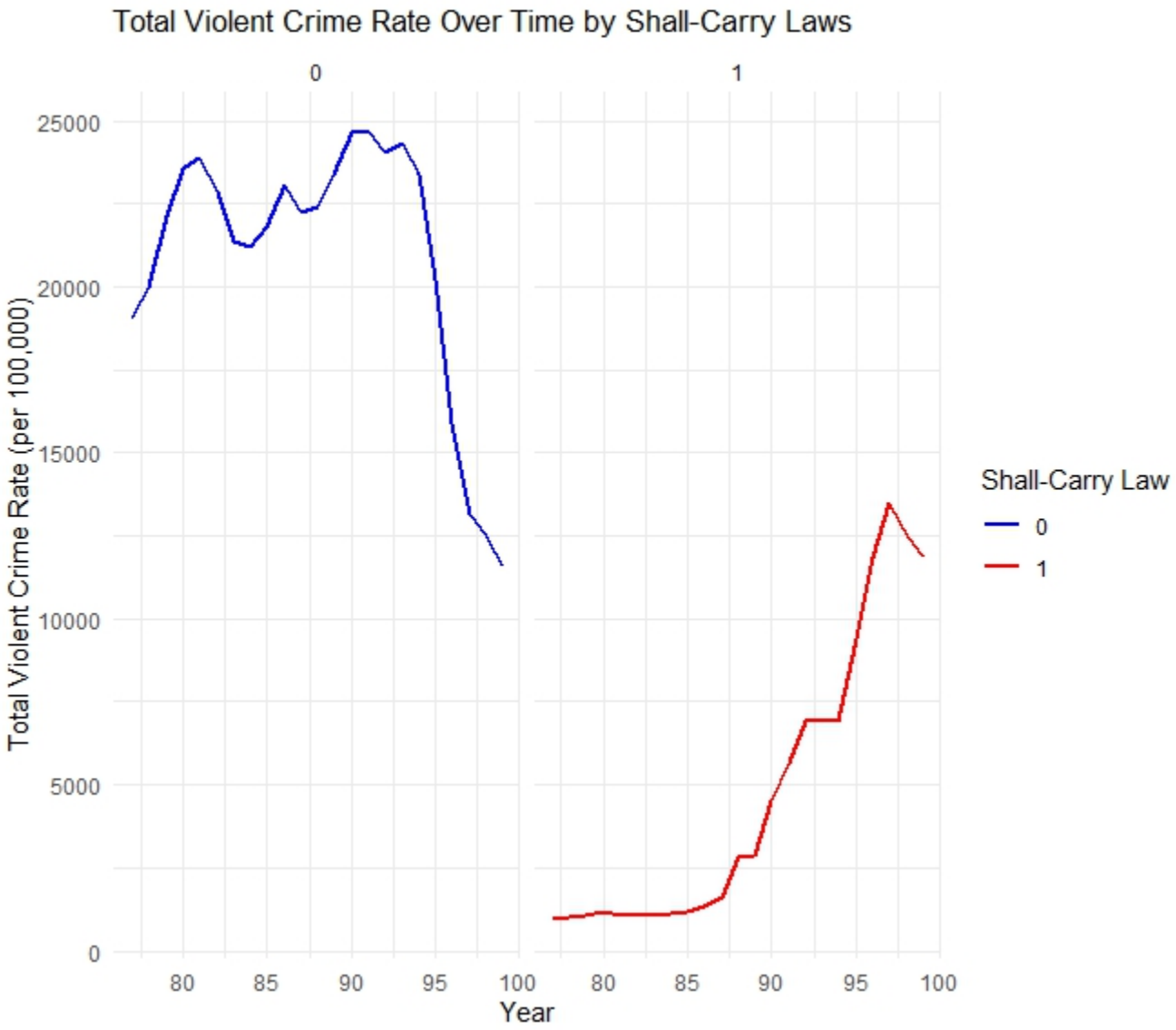
We now have normally distributed data for the variables which will help us generate a robust model further.

3.3 Closer Look at the Variables

3.3.1 Crime Rates Over Time - Total Violent Crime Rates by Shall-Issue Laws

Temporal trends in total violent crime rates were scrutinized, aiming to identify overarching patterns and potential divergences based on the implementation of shall-issue laws.

A closer examination of trends specifically categorized by shall-issue law status provided a granular perspective on the potential impact of such laws on crime rates.



Interpretation: Looking at the above figure, we notice that both groups—states with the law and those without—show pretty much the same pattern in their violent crime rates over time.

But, In earlier years, having shall-carry laws seemed to help bring down violent crime until around 1989. After that, though, things kind of leveled out. In the last decade, the impact of these laws on violent crime rates seems to be more consistent and similar to places without such laws.

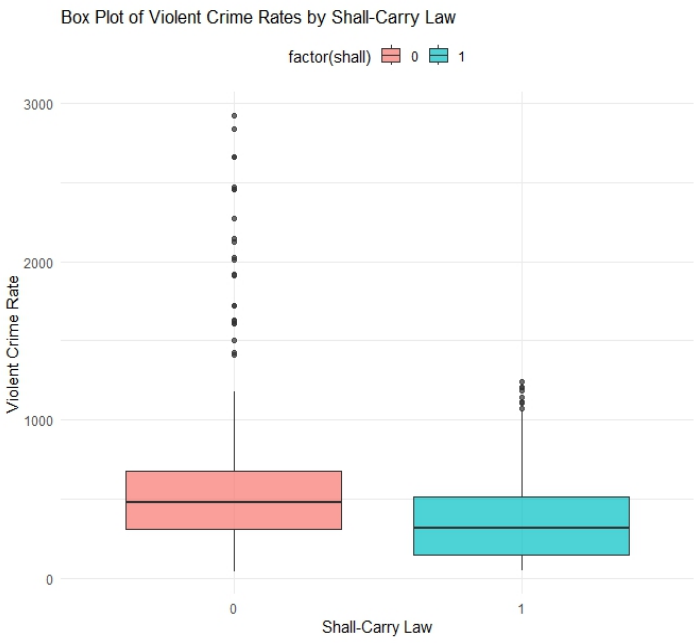
3.3.2 Comparative Analysis

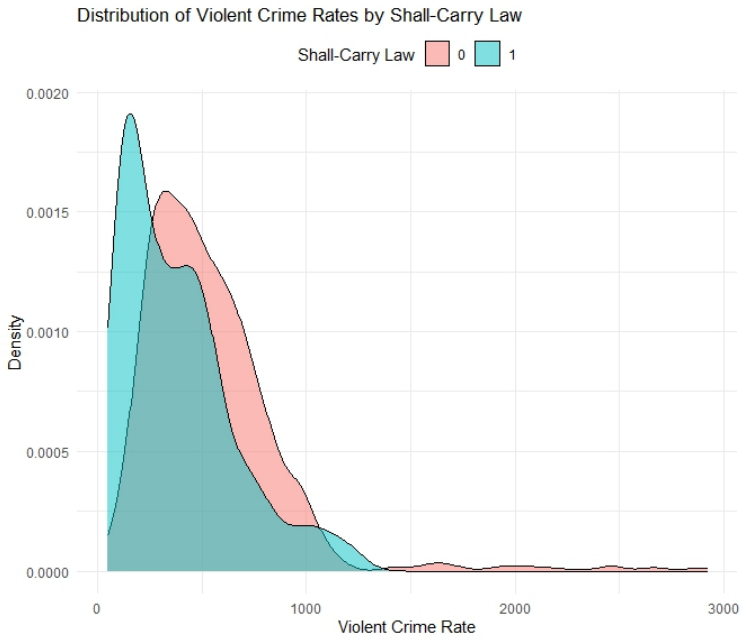
Distributions of Crime Rates

Crime rate distributions were compared between states with and without shall-issue laws, employing density plots and box plots to highlight potential disparities.

Box Plots and Density Plots

Detailed box plots and density plots are made to study the effect of the implementation of the Shall-Carry Law.



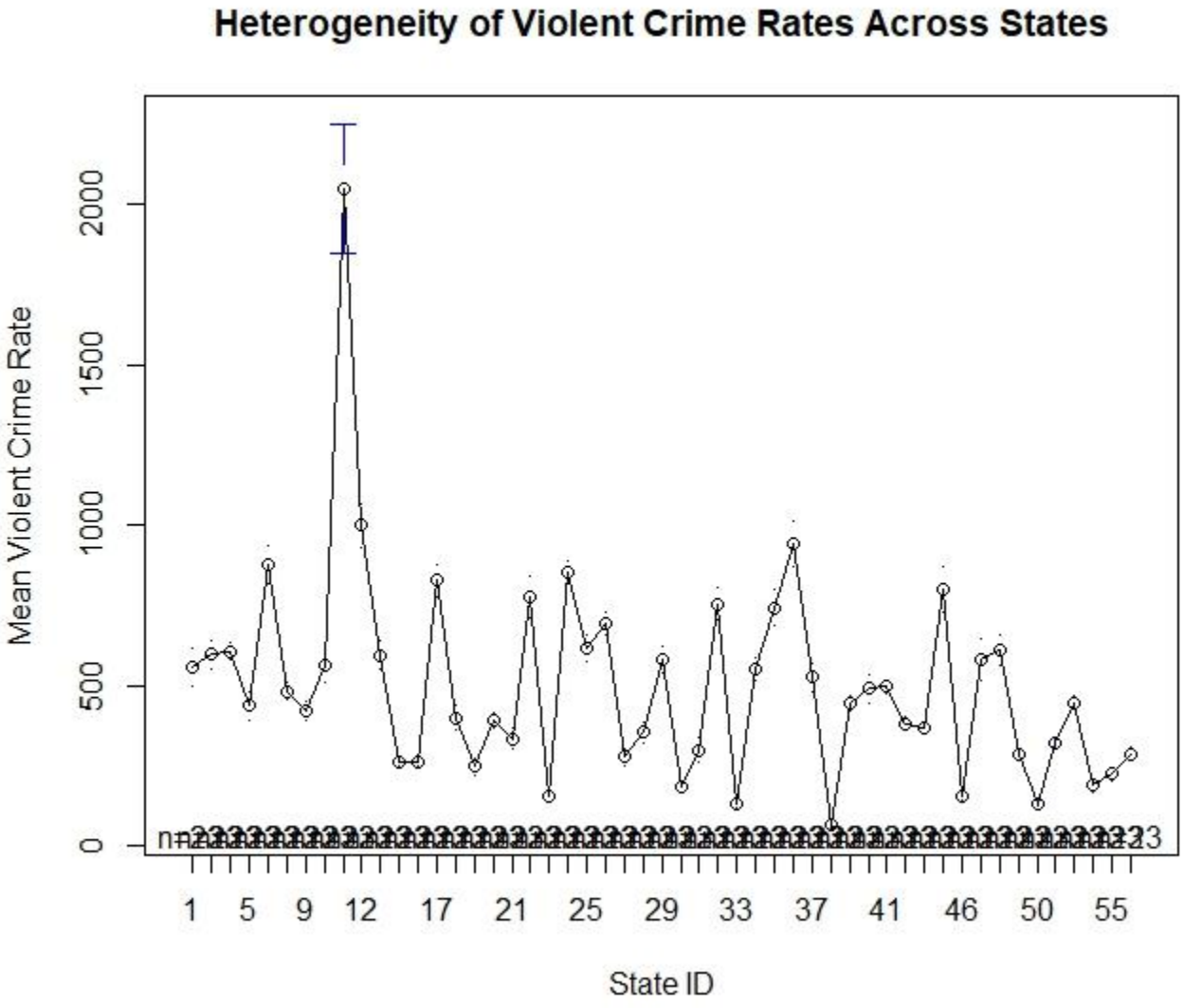


Interpretation: We can observe that the overall crime rate is lower in states that have these laws. So, it looks like having Shall-Carry Laws does make a difference in bringing down the overall crime rates.

3.3.3 Geographic and Temporal Variation

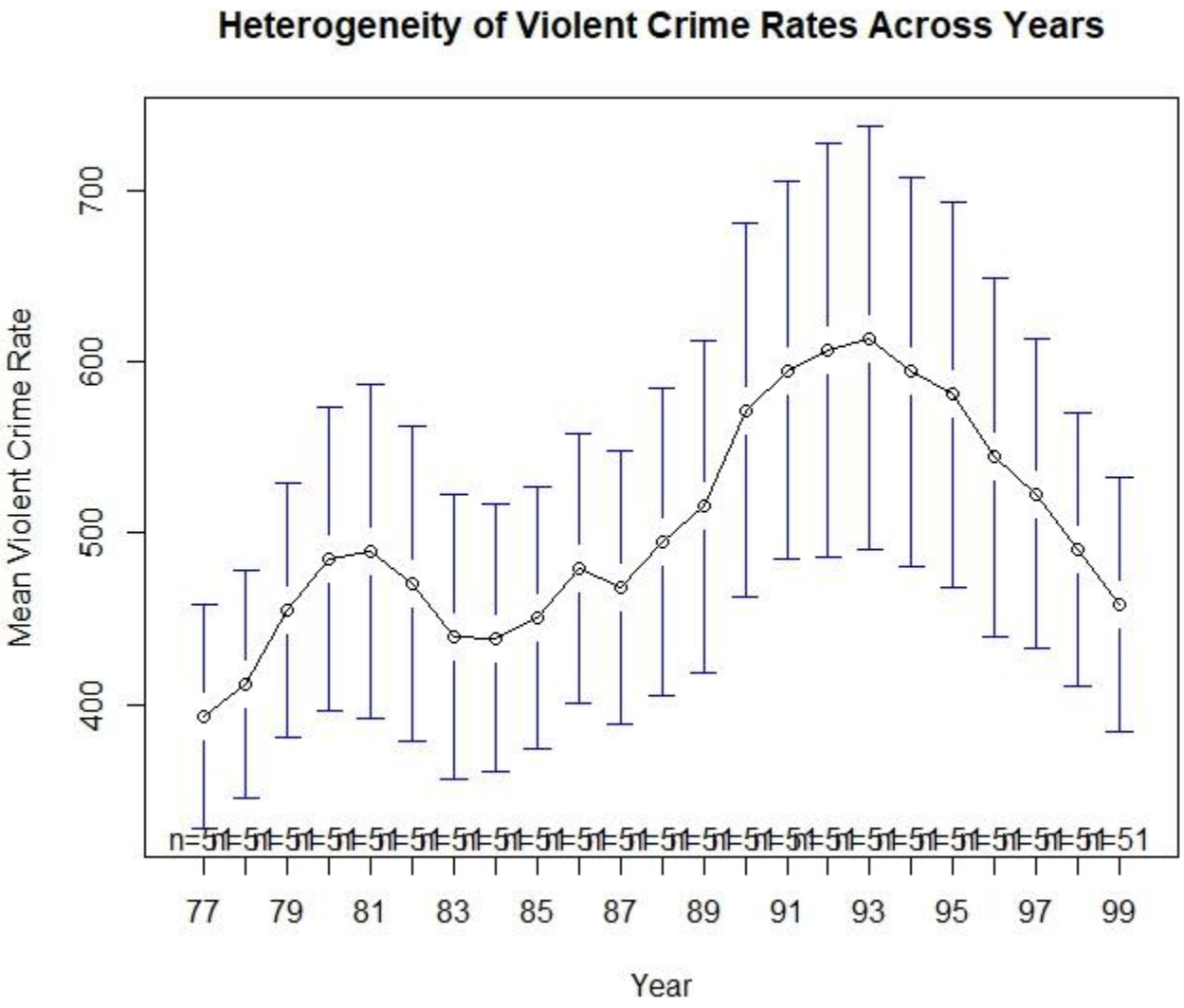
Variation Across States

Heterogeneity in crime rates across states was explored to identify potential geographic patterns and variations contributing to our understanding of crime dynamics.



Variation Across Years

Temporal variations in crime rates were examined to identify potential trends and fluctuations over the 23-year period.



We observed heterogeneity in both cases.

4. Models

4.1 Pooled OLS Model

We begin with the simplest model possible for panel data. In the Pooled Ordinary Least Squares Model, the data for all entities are pooled together and the regression is estimated using the least squares method.

We estimate the following model:

$$\ln(vio) = \beta_1 + \beta_2 * shall_{it} + \beta_3 * \ln(incarc\_rate)_{it} + \beta_4 * \ln(density)_{it} + \beta_5 * avginc_{it} + \beta_6 * \ln(pop)_{it} + \beta_7 * pm1029_{it} + \beta_8 * \ln(pb1064)_{it} + e_{it}$$

term	estimate	std.error	statistic	p.value
(Intercept)	0.6356357	0.2706187	2.348824	0.0189992
factor_shall1	-0.2294789	0.0272536	-8.420139	0.0000000
log_incarc	0.5894483	0.0277798	21.218572	0.0000000
log_density	0.0581206	0.0083101	6.993994	0.0000000
avginc	0.0361363	0.0052711	6.855556	0.0000000
log_pop	0.1556188	0.0117547	13.238829	0.0000000
log_pb1064	0.0784065	0.0193357	4.055007	0.0000535
pm1029	0.1095482	0.0093655	11.696982	0.0000000

Interpretation:

We see that according to this model, that if the concealed carry law is in place, all kinds of violent crime drop by around 23%. We can also see that this is significant at all significance levels.

A reduction in 23% in violent crime just by the introduction of the shall law is too large in the real-world sense. The drop in violent crime is too large in magnitude.

According to the model, as incarceration rates go up, the rate of violent crime also goes up which in contrary to what we expected. The model states that a 1% increase in incarceration rate would result in a 0.59% increase in violent crimes. We suspect the relationship between both variables to be bidirectional. That is, the model assumes that change in incarceration rates cause an increase in violent crime, but it could also be that an increase in violent crime might cause an

increase in incarceration rate. There could be endogeneity or a few socioeconomic factors that might have been omitted.

The model also suggests that a 1% increase in population density might result in around a 0.06% rise in violent crimes. This is also significant at all confidence levels.

According to the model, if the average income increases by a $1000, the rate of violent crime increases by 3.6%. This is significant at all confidence levels. This is opposite to a logical assumption that an increase in the average income results in a drop in violent crime rate. This suggests the presence of bias or omitted socioeconomic ort cultural variables.

The model rightly suggests that if the population increases by 1%, the rate of violent crime will rise by 0.15%. This is also significant at all confidence levels.

The model estimates that a 1% increase in the percentage of black males aged from 10 to 64 will result in an increase in the rate of violent crimes by 0.078%. This is also significant at all confidence levels.

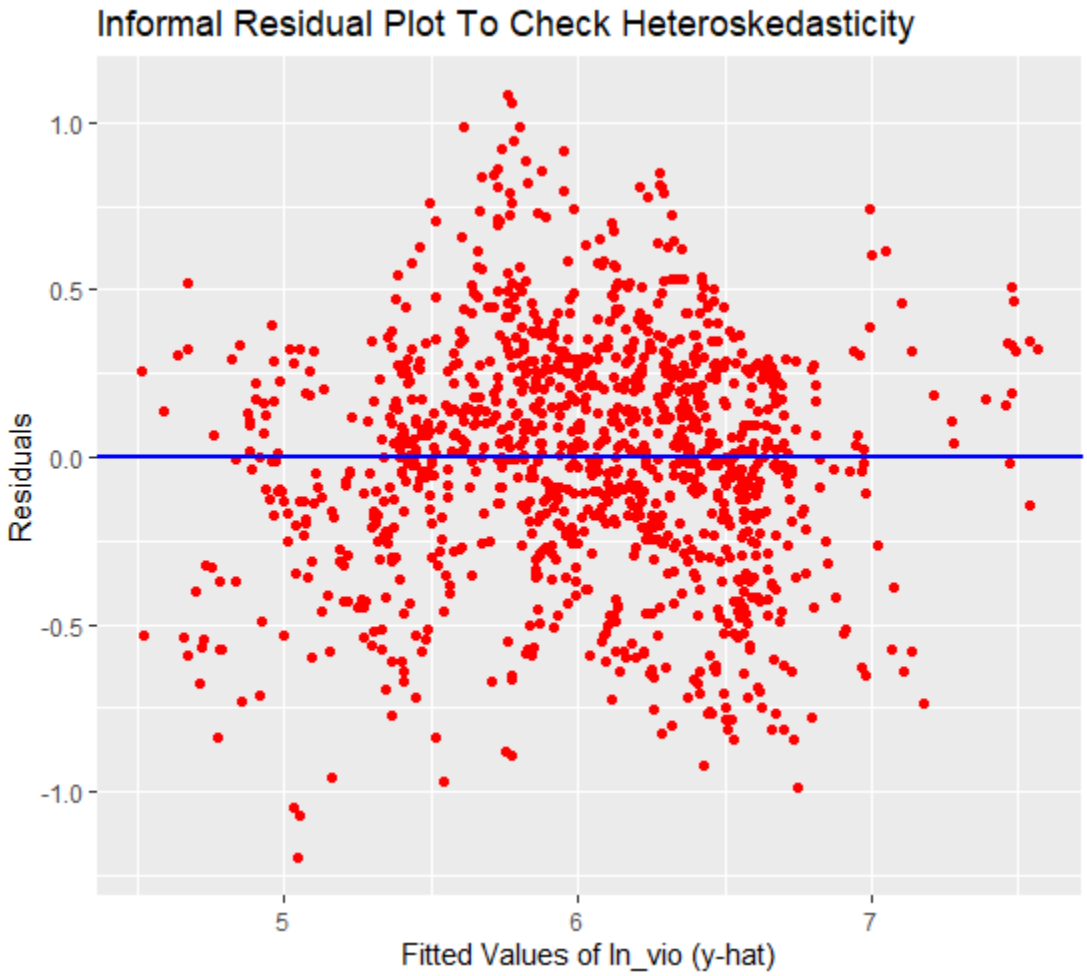
The effect of an increase in the percentage of young males of the age group 10 to 29 is consistent with our guess. However, the model suggests that a 1% increase results in a 11% increase in violent crime. We believe the true effect to be much lower than what has been estimated by the model. This could also be due to unobserved heterogeneity and endogeneity introducing bias in our model. This estimate by the model is also significant at all confidence levels.

The results from the Pooled OLS Model strongly indicate the presence of heterogeneity in the data, and it is not wise to pool all individual states together assuming that they behave in the same way. It is also incorrect to assume that a particular state also behaves the same way throughout 23 years.

It is also not sensible to assume that the same state's errors will not be correlated with each other. All unobservable characteristics that are included in the error term are likely to be correlated for the same state. This also violates the assumption that the data is not heteroskedastic and that the error terms are not serially correlated.

4.1.1 Heteroskedasticity Check:

To confirm our suspicions of heteroskedasticity, we can perform two tests one informal and one formal. We can informally test it by plotting the residuals against it's fitted values. The plot is shown below:



We can see that there is a slight trend in the values of the residuals when plotted against the predicted values from the OLS Model. We can see that an inverted parabola passes through the error telling us that there is a heteroskedasticity present in the data. However, we will need to confirm this using a formal test.

4.1.2 White's Formal test:

For the general regression model, we know that $E(y_i) = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK}$

A general form for the variance function related is: $var(y_i) = \sigma^2 = E(e^2_i) = h(\alpha_1 + \alpha_2 z_{i2} + \ldots + \alpha_s z_{is})$ where z's are the combinations and functions of x's

H_0: $\alpha_2 = \alpha_3 = \ldots = \alpha_s = 0$

H_1: not all the α_i in H_0 are 0

On performing the Breusch Pagan test for heteroskedasticity, we found the following results:

```
        studentized Breusch-Pagan test

data:  ols_model
BP = 55.465, df = 8, p-value = 3.586e-09
```

We can see that the BP statistic is extremely high with a value of 55.5 and the p-value is very close to 0. Therefore, we reject the null hypothesis and conclude that there is heteroskedasticity present in the Pooled OLS model.

Since the model is heteroskedastic, the estimators are no longer the best and the standard errors of the usual least squares are incorrect.

4.2 Pooled OLS Model with Cluster Robust Standard Errors

We can still tolerate the estimator not being the best, but we cannot have incorrect standard errors. We can apply Cluster Robust Standard Errors to correct the overestimation of the precision of the estimates.

We calculate the new standard errors and run the Pooled OLS with them. These are the results:

term	estimate	std.error	statistic	p.value
(Intercept)	0.6356357	1.0268894	0.6189914	0.5360431
factor_shall1	-0.2294789	0.0660592	-3.4738381	0.0005319
log_incarc	0.5894483	0.1206983	4.8836503	0.0000012
log_density	0.0581206	0.0317121	1.8327551	0.0670941
avginc	0.0361363	0.0155746	2.3202126	0.0205015
log_pop	0.1556188	0.0451652	3.4455486	0.0005902
log_pb1064	0.0784065	0.0870347	0.9008650	0.3678463
pm1029	0.1095482	0.0305805	3.5822893	0.0003546

We notice that the estimates are still the same, confirming that the estimators are not the best, but the standard errors are now correct and more than that of the simple Pooled OLS model.

Since there is still unobserved heterogeneity and various forms of bias, we move on to the next model.

4.3 Entity Fixed Model

In the Pooled OLS model, we found that the estimates were unreliable due to unobserved heterogeneity. These unobserved factors could also be correlated with the explanatory variables, which could result in an endogeneity problem.

Panel data allows us more methods to control for unobserved heterogeneity. We will now be estimating an Entity Fixed Model. This model will account for the unobserved heterogeneity across all states with time being invariant.

We estimate the following model:

$$\ln(vio) = \beta_1 + \beta_2 * shall_{it} + \beta_3 * \ln(incarc\_rate)_{it} + \beta_4 * \ln(density)_{it} + \beta_5 * avginc_{it} + \beta_6 * \ln(pop)_{it} + \beta_7 * pm1029_{it} + \beta_8 * \ln(pb1064)_{it} + e_{it}$$

term	estimate	std.error	statistic	p.value
factor_shall1	0.0201490	0.0180757	1.1146966	0.2652207
log_incarc	0.0200823	0.0280648	0.7155699	0.4744068
log_density	-1.2322403	1.5171777	-0.8121925	0.4168545
avginc	0.0051527	0.0060346	0.8538643	0.3933636
log_pop	1.0814423	1.5020356	0.7199845	0.4716854
log_pb1064	-0.2056104	0.0562488	-3.6553735	0.0002688
pm1029	-0.0556181	0.0083375	-6.6708426	0.0000000

Interpretation:

We see that according to this model, most of the factors: shall carry in effect, incarceration rate, population density, average income, population are insignificant. We could say that theses variables are not changing drastically in the state over the years.

Also, the only two significant factors are percent of black males between the ages of 10 and 64 and the percentage of males between the ages of 10 and 29.

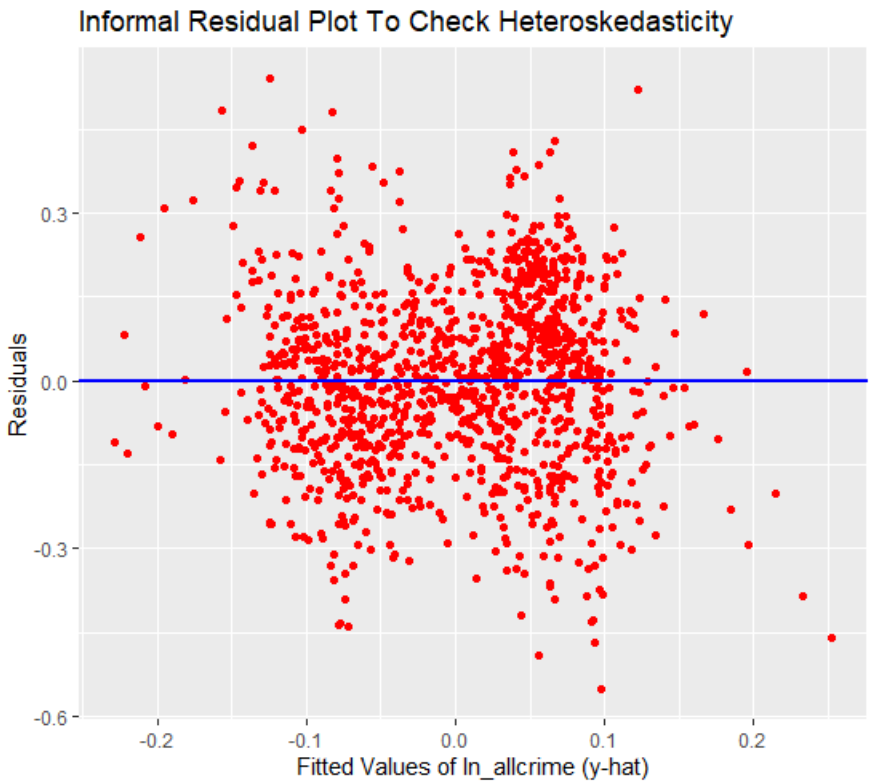
According to the model, 1% increase in the percentage of black male population between the ages of 10 and 64 results in a 0.2% decrease in violent crime. This is against traditional expectations, and this might indicate that unobserved heterogeneity may still be present in the model. There could also be several socio-economic factors that vary with time that might give incorrect results.

Surprisingly, the model also suggests that if the male population between the ages of 10 and 29 increases by 1% that results in a 5% decrease in violent crime rates. This also contrary to what we expected. This also indicates the presence of unobserved heterogeneity.

We notice that the unobserved heterogeneity is still present in the model, and there also might be endogeneity in the model. There is a possibility that various factors like economic conditions, laws in the federal level, social conditions are constant within states but vary over time. These factors are still in the error term that the Entity Fixed model failed to explain.

4.3.1 Heteroskedasticity Check:

Also, if we plot the residuals against the fitted values in the model, we get the following plot:



We can still see that there is some pattern in the way the residuals are plotted and that indicates the presence of heteroskedasticity.

4.4 Time and Entity Fixed Model

As discussed earlier, we still have reason to believe that our model is not free from unobserved heterogeneity. Factors that vary over time such as federal laws and socioeconomic factors were not explained using the entity-fixed model.

So, we have decided to use the Time and Entity Fixed model. This model includes time effects by including indicator variables for the years. Ideally this should get rid of any unobserved heterogeneity that is constant across states but vary with time.

We estimate the following model:

$$\ln(vio) = \beta_0 + \beta_1 * shall_{it} + \beta_2 * \ln(incarc\_rate)_{it} + \beta_3 * \ln(density)_{it} + \beta_4 * avginc_{it} + \beta_5 * \ln(pop)_{it} + \beta_6 * pm1029_{it} + \beta_7 * \ln(pb1064)_{it} + \delta_2 * year78 + \delta_3 * year79 + \delta_4 * year80 + \delta_5 * year81 + \delta_6 * year82 + \delta_7 * year83 + \delta_8 * year84 + \delta_9 * year85 + \delta_{10} * year86 + \delta_{11} * year87 + \delta_{12} * year88 + \delta_{13} * year89 + \delta_{14} * year90 + \delta_{15} * year91 + \delta_{16} * year92 + \delta_{17} * year93 + \delta_{18} * year94 + \delta_{19} * year95 + \delta_{20} * year96 + \delta_{21} * year97 + \delta_{22} * year98 + \delta_{23} * year99 + e_{it}$$

term	estimate	std.error	statistic	p.value
factor_shall1	-0.0437320	0.0167690	-2.607913	0.0092338
log_incarc	-0.1018248	0.0269613	-3.776697	0.0001675
log_density	-2.1273620	1.2781834	-1.664364	0.0963264
avginc	0.0092357	0.0060169	1.534965	0.1250819
log_pop	1.8933166	1.2669645	1.494372	0.1353670
log_pb1064	-0.4307926	0.0483676	-8.906637	0.0000000
pm1029	0.0841088	0.0107193	7.846483	0.0000000
y78	0.0722705	0.0268694	2.689689	0.0072608
y79	0.1990381	0.0272362	7.307854	0.0000000
y80	0.2774745	0.0276091	10.050111	0.0000000
y81	0.2946342	0.0283246	10.402052	0.0000000
y82	0.2973850	0.0299848	9.917854	0.0000000
y83	0.2825431	0.0322712	8.755273	0.0000000
y84	0.3287487	0.0349359	9.410051	0.0000000
y85	0.3941642	0.0377008	10.455074	0.0000000
y86	0.4892547	0.0411170	11.899074	0.0000000
y87	0.5056618	0.0444524	11.375358	0.0000000
y88	0.5847119	0.0479773	12.187263	0.0000000
y89	0.6571426	0.0513407	12.799630	0.0000000
y90	0.8050387	0.0545221	14.765361	0.0000000
y91	0.8779596	0.0572618	15.332393	0.0000000
y92	0.9268929	0.0603834	15.350133	0.0000000
y93	0.9658707	0.0625862	15.432644	0.0000000
y94	0.9663054	0.0650853	14.846747	0.0000000
y95	0.9782117	0.0678254	14.422499	0.0000000
y96	0.9401469	0.0705324	13.329288	0.0000000
y97	0.9334669	0.0729611	12.794031	0.0000000
y98	0.8895973	0.0755739	11.771231	0.0000000
y99	0.8438378	0.0775642	10.879221	0.0000000

Interpretation:

According to the Time and Entity Fixed model, if the shall carry law is in effect, the rate of violent crime is reduced by 4.3%. This is significant at the 1% significance level. This suggests that the states that have shall carry law in effect has only 4.3% decrease violent crimes.

The model also estimates that a 1% increase in incarceration rate results in a 0.1% decrease in the rate of violent crime. It is also found to be significant at all significance levels. The model's estimate is also in line with our assumption, and we can say that there is no simultaneous causality bias in the model anymore.

The model also estimates that the population density, average income, and population of the states are not significant indicators of the rate of violent crime. The coefficient of population density is also negative which suggests that more dense areas are less likely to have incidents of violence which is not according to our expectations. However, the model suggests that these variables are not statistically any different from 0.

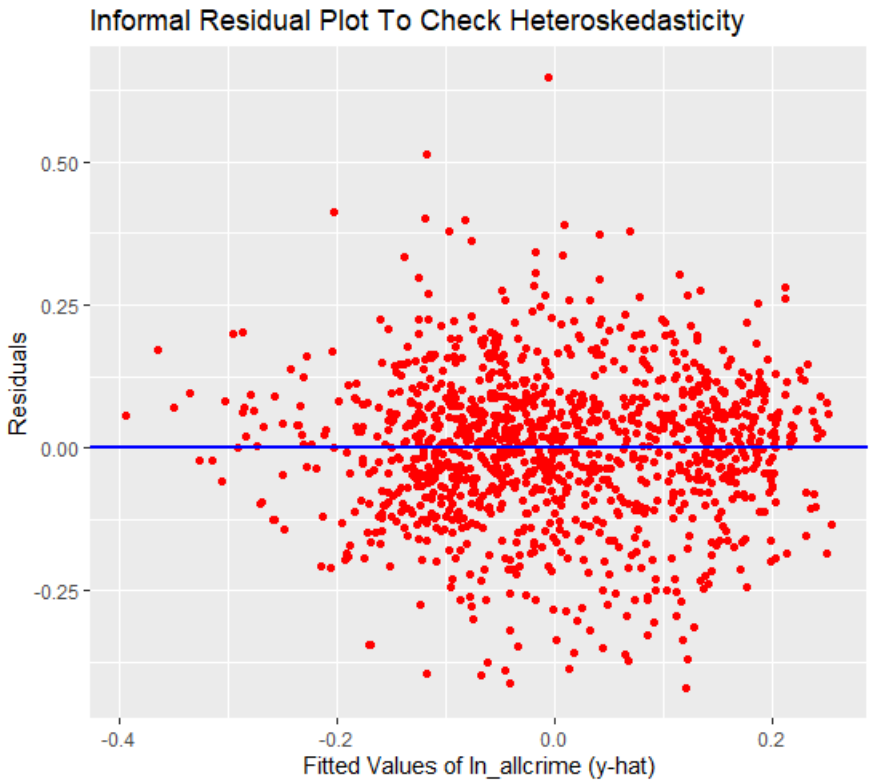
According to the model, a 1% increase in the percentage of black individuals between the ages of 10 and 64 in the state results in a drop in the rate of violent crime by 0.43%. This is significant at all levels of confidence. This is contrary to traditional expectations, suggesting that there might be various complex socio-economic factors at play.

According to our expectations, the model estimates that a 1% increase in the young male population in the state results in an 8.4% increase in the rate of violent crime in that state. This is significant at all levels. We can confirm that all unobserved heterogeneity has been removed and the model is accurate.

We also see a steady rise in the rate of violence in the states from 1978 to 1995 where it peaks and then we notice a gradual decline to the year 1999.

4.4.1 Heteroskedasticity Check:

Also, if we plot the residuals against the fitted values in the model, we get the following plot:



We see that there is no pattern present, and we can confirm that this model has no heteroskedasticity.

4.5 Comparing the models

Entity Fixed Model - Time and Entity Fixed Model:

We first compare the Entity Fixed Model with the Time and Entity Fixed Model.

To make this comparison we conduct the following hypothesis test:

H_0: $\delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = \delta_7 = \delta_8 = \delta_9 = \delta_{10} = \delta_{11} = \delta_{12} = \delta_{13} = \delta_{14} = \delta_{15} = \delta_{16} = \delta_{17} = \delta_{18} = \delta_{19} = \delta_{20} = \delta_{21} = \delta_{22} = \delta_{23} = 0$

H_1: At least one of the coefficients of the time effects is not 0

We perform the F-test using R and get the following result:

```
        F test for individual effects

data:  log_vio ~ factor_shall + log_incarc + log_density + avginc +  ...
F = 26.155, df1 = 22, df2 = 1093, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Since the p-value < 0.05 we can reject the null hypothesis.

This means that at least one of the coefficients of the time effects of the Time and Entity Fixed model is significantly different from 0.

Hence, between the Entity Fixed model and the Time and Entity Fixed model, we choose the Time and Entity Fixed model.

Time and Entity Fixed Model – Pooled Ordinary Least Squares Model:

Now we compare the Time and Entity Fixed Model and the Pooled Ordinary Least Squares Model.

We construct the following hypothesis test to check whether all the coefficients of the states are the same or at least one of the coefficients are different. If even one of the coefficients is statistically different, we can reject the null hypothesis, and say that there is a fixed effect for the states.

H_0: $\beta_{1,1} = \beta_{1,2} = \beta_{1,3} = \ldots = \beta_{1,i}$

H_1: At least one of the coefficients are different.

We conducted the above hypothesis test in R and here is the output:

```
        F test for individual effects

data:  log_vio ~ factor_shall + log_incarc + log_density + avginc +  ...
F = 100.15, df1 = 72, df2 = 1093, p-value < 2.2e-16
alternative hypothesis: significant effects
```

As we can see, the p-value < 0.05, so we can reject the null hypothesis.

We can therefore confirm the presence of fixed effects and can thereby choose the Time and Entity Fixed model as the most appropriate model for this scenario.

4.6 Random Effects Model:

The use of the random effects model is not suitable for this data set. Using Random Effects makes more sense for a sample that is randomly selected from the population.

Since we know that we are dealing with 50 states of America and the district of Columbia using the random effects model is not suitable.

5. Conclusion

We initially explored the dataset and found that many of the provided variables to be strongly correlated with each other.

We saw that the rate of violent crimes, rate of murders, and rate of robberies were highly correlated, and we assumed the rate of murders and robberies to be subsets of the rate of violent crimes.

We also noticed that the percentage of black people in the state and the percentage of white people in the state were highly inversely correlated. We therefore chose to only use a percentage of black people in our models to avoid multicollinearity.

The results from the Pooled OLS Model strongly indicate the presence of heterogeneity in the data, and it is not wise to pool all individual states together assuming that they behave in the same way. It is also incorrect to assume that a particular state also behaves the same way throughout 23 years.

We notice that the unobserved heterogeneity is still present in the entity-fixed model, and there also might be endogeneity in the model. There is a possibility that various factors like economic conditions, laws at the federal level, and social conditions are constant within states but vary over time. These factors are still in the error term that the Entity Fixed model failed to explain
We faced a lot of issues such as unobserved heterogeneity, serially correlated errors, endogeneity, heteroskedasticity, etc.

On comparing all three models we found that the Time and Entity Fixed Model performs best in this scenario. We received various insights from the model that we have listed below:

1. When the shall carry law is in effect the rate of violent crime is reduced by 4.3%. This is significant at the 1% significance level. However, this is a modest amount and doesn't decrease the rate of violent crime by a lot.

2. A 1% increase in incarceration rate results in a 0.1% decrease in the rate of violent crime. It is also found to be significant at all significance levels. The model's estimate is also in line with our assumption, and we can say that

there is no simultaneous causality bias in the model anymore.

3. The population density, average income and population of the states are not significant indicators of the rate of violent crime.

4. A 1% increase in the percentage of black individuals in between the ages of 10 and 64 in the state result in a drop in the rate of violent crime by 0.43%. This is significant at all levels of confidence. This is in contrary to traditional expectations, suggesting that there might be various complex socio-economic factors at play.

5. A 1% increase in the young male population in the state results in an 8.4% increase in the rate of violent crime in that state. This is significant at all levels.

6. We also see a steady rise in the rate of violence in the states from 1978 to 1995 where it peaks and then we notice a gradual decline to the year 1999.