**Do shall-issues law reduce crime-or not?**

**BUAN 6312.002 Applied Econometrics and Time Series Analysis**

**University Of Texas, Dallas**

**By**

**Group Members**

Anurag Gupta- AXG210034

Krina Mehul Shah - KMS220004

Moksh Mehool Mehta - MXM220009

Rohit Patil -RXP220030

# 1. Introduction

The influence of firearms on crime in America has sparked extensive public discourse, with a prevalent perspective asserting that state laws allowing citizens to carry concealed handguns contribute to a reduction in crime. Advocates of this viewpoint argue that gun control measures strip law-abiding citizens of their firearms, leaving potential victims defenseless while criminals disregard such regulations. Organizations like the National Rifle Association (NRA) and numerous politicians champion the idea of expanding the freedom to carry guns.

Consequently, several states in the United States have enacted right-to-carry laws, also known as shall-issue laws. A shall-issue law mandates that governments issue concealed carry handgun permits to any applicant meeting specific criteria. These criteria typically include being an adult, possessing no significant criminal record, having no history of mental illness, and successfully completing a firearms safety training course if required by law. Importantly, once these conditions are met, the granting authority is obligated to issue the license, without the need for the applicant to demonstrate a "good cause."

Our research aims to investigate whether the implementation of shall-issue laws has a direct impact on the increase in crime rates and violence across states in the United States based on the 23 years historical data for 51 states on demographics, shall laws, violenes and incarceration rate.

# 2.Dataset Description

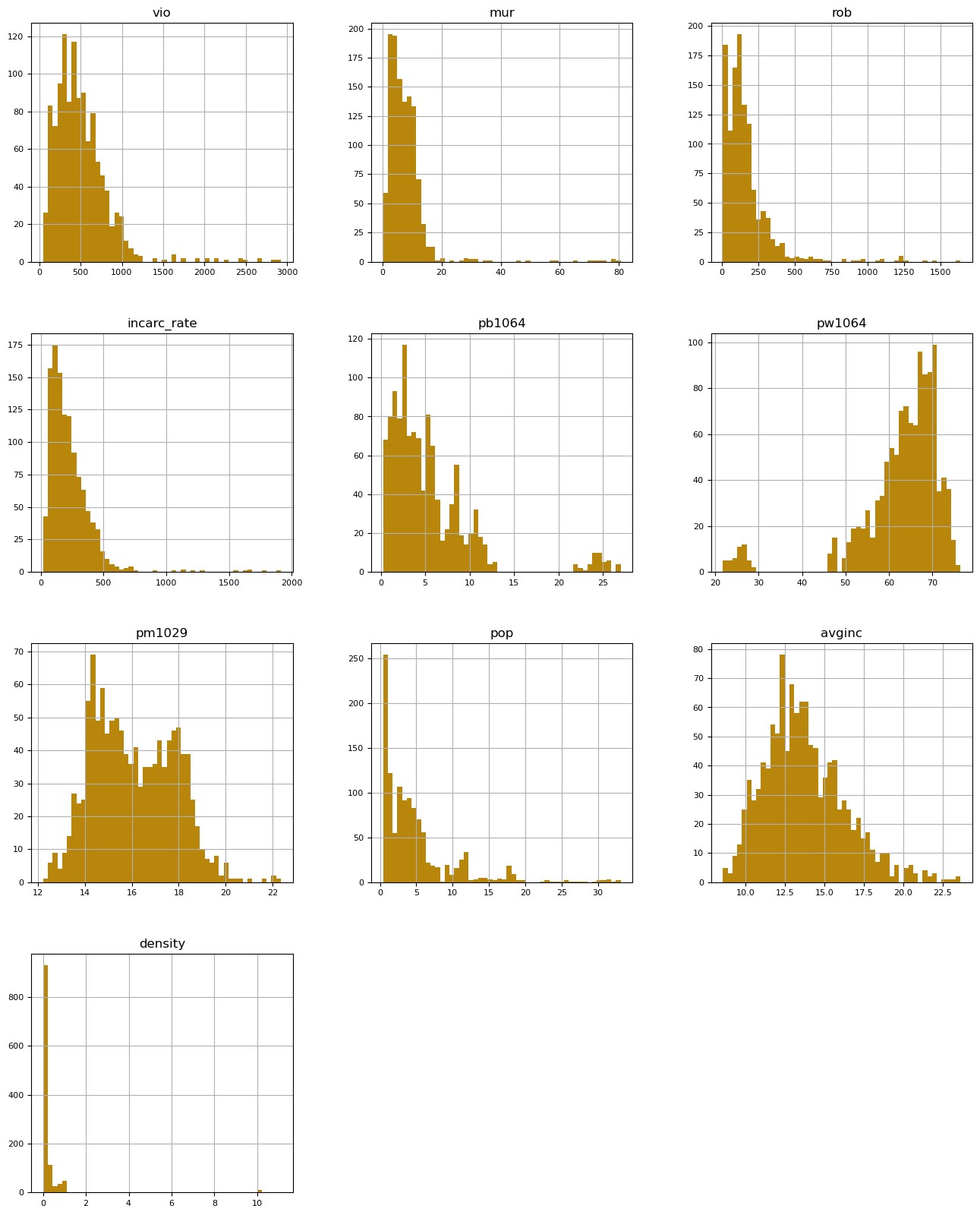
The given dataset contains data for 51 states including 50 US states and the District of Columbia. It is a panel of data that contains yearly observations for 23 years for these 51 states beginning from year 1977 to 1999. In total there are 1173 observations containing data for the following variables:

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

# 3.Exploratory Data Analysis

Now we performed some Data Exploration Analysis we delved into the intricacies of the dataset to glean insights and patterns. The examination revealed notable trends and variations across the 51 states over the 23-year timeframe (1977-1999). These initial explorations lay the groundwork for a more in-depth analysis, offering a preliminary understanding of the dataset's dynamics and providing a basis for subsequent statistical modeling testing.

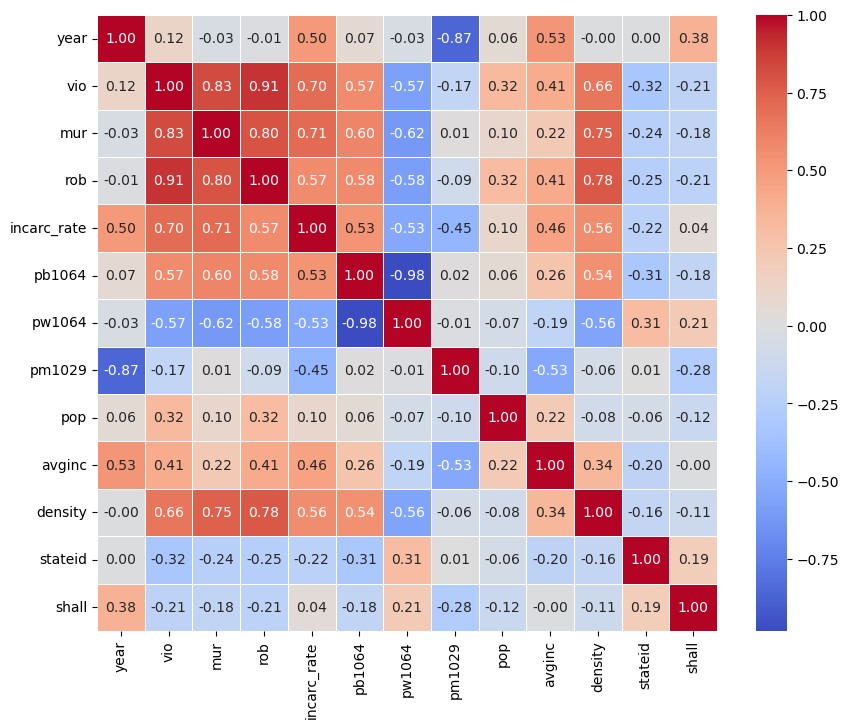
**3.1 Data Distribution**



The histograms above reveal a notable skewness in the dataset. Notably, when examining the data for dependent variables such as violence, murder, and robbery, it becomes evident that the distribution is skewed to the left. Additionally, the data for explanatory variables exhibits a high degree of skewness. Addressing this skewness can be achieved by applying a log transformation to both the dependent and independent variables.

The log transformation is a common technique used to mitigate the impact of skewness, promoting a more symmetric distribution and facilitating a more robust analysis of the data.

## 3.2 Correlation Matrix



The correlation heatmap analysis yields several noteworthy observations regarding the relationship between variables. In the context of investigating the impact of shall-issue laws on crime rates, it is discerned that the correlation of shall-issue laws with all other variables is notably low.

The incarceration rate exhibits a strong positive correlation and is particularly associated with a positive correlation to the black population while displaying a negative correlation with the white population. Additionally, average income and population density demonstrate a positive correlation with the incarceration rate.

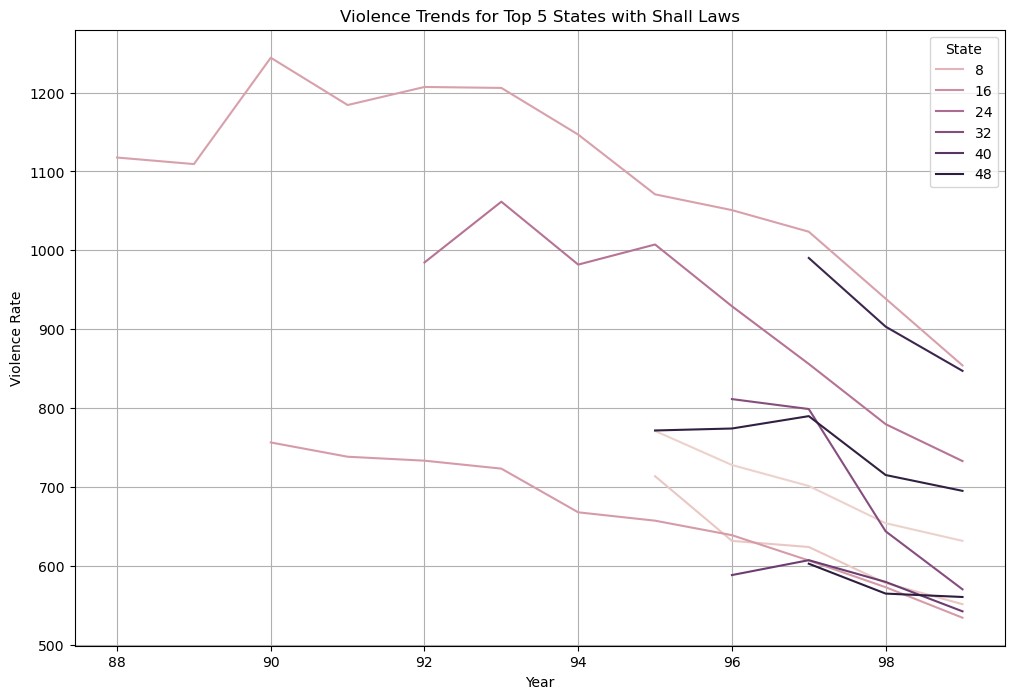
Population density emerges as a significant factor, displaying strong positive correlations with variables such as murder, robbery, and overall crime rates. Moreover, all types of crime exhibit direct and positive associations with average income.

In conclusion, the findings suggest a positive correlation between demographic and economic variables with crime rates. This prompts further investigation into how these demographic and economic factors may be influencing crime rates in states implementing shall-issue laws.

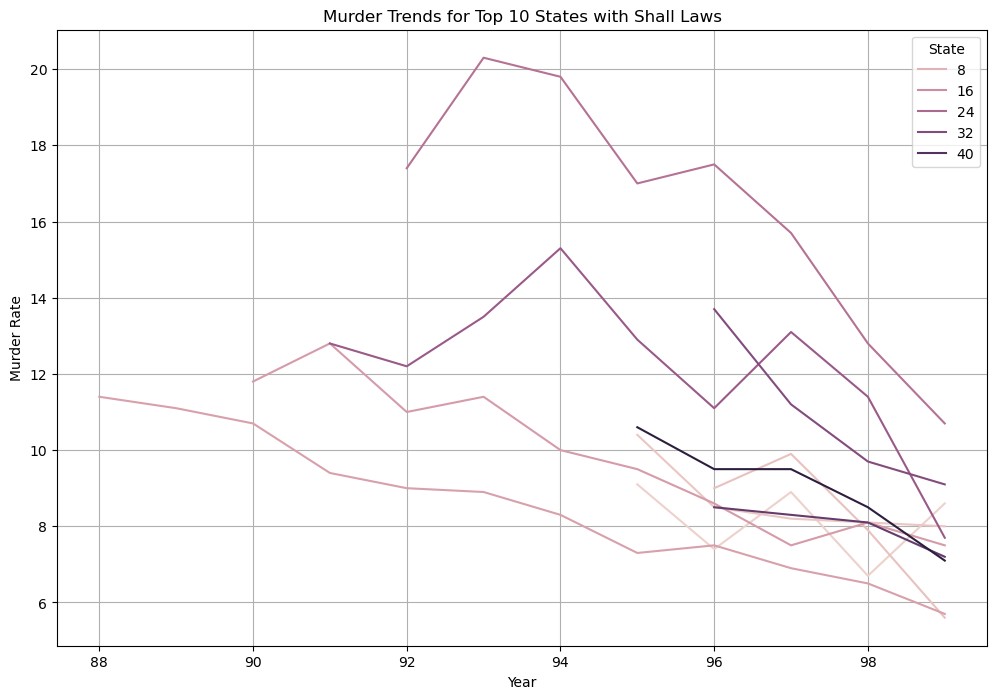
## 3.3 Analyzing Criminal Trends in States with Shall Law

As our main objective focuses on the analysis of effectived of Shall issue laws for every state over crime. We will plot the crime rate for top 10 states against the year when the shall laws was issued and trends furtheward to understand if the shall issue laws helping in reducing crimes?

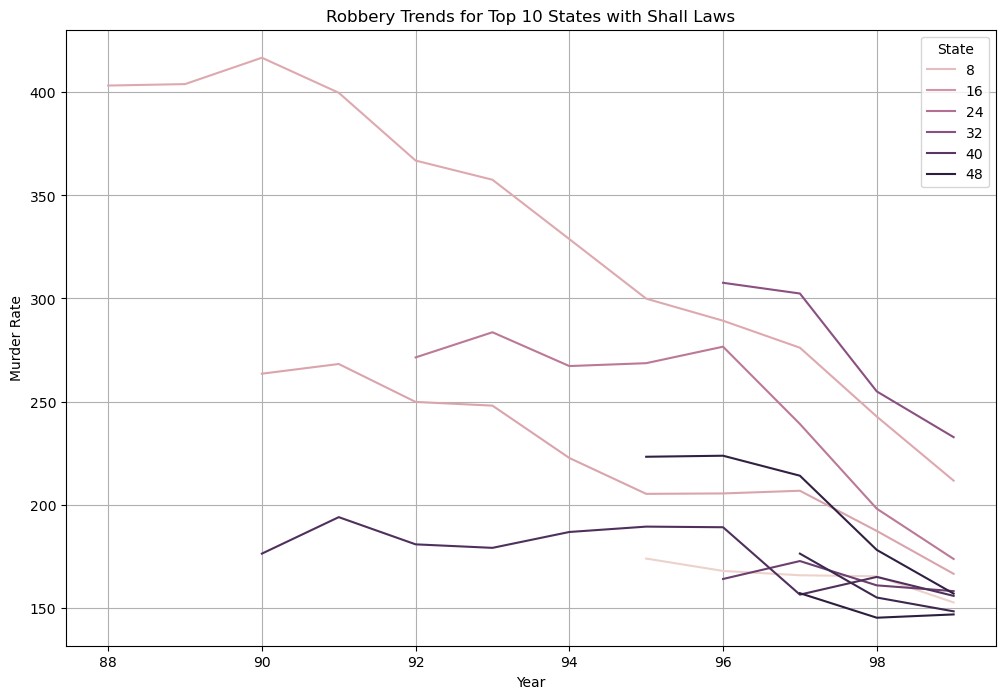
1. **Violence Trends in Top 10 States with Shall Laws**



1. **Murder Trends in Top 10 States with Shall Laws**



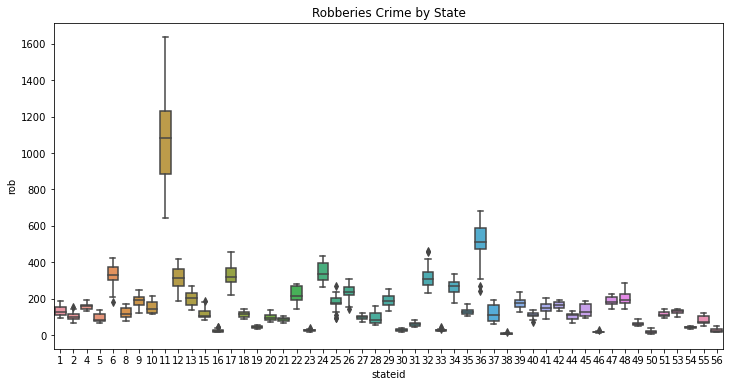
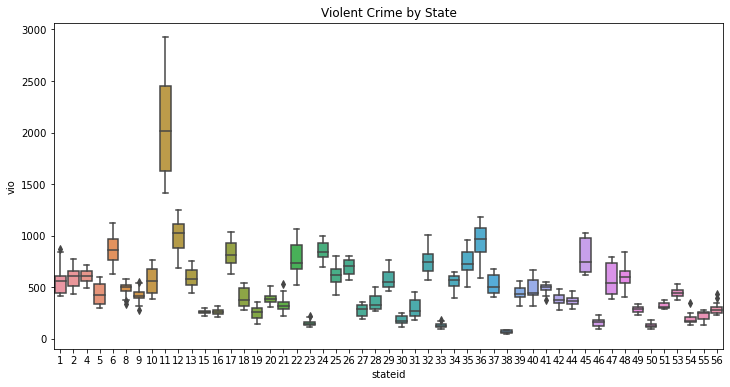
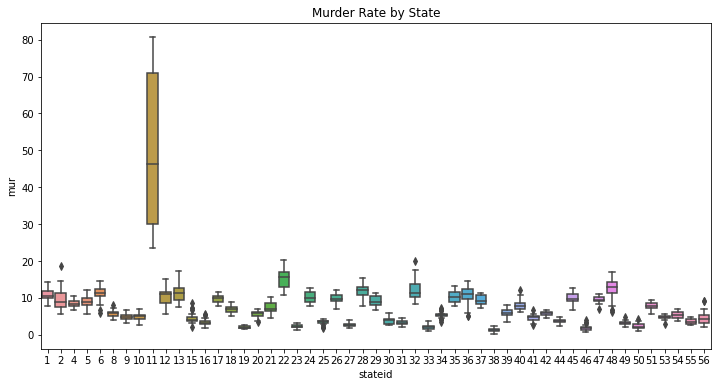
1. **Robbery Trends in Top 10 States with Shall Laws**



Based on the observed trends in various crime types, it appears that the implementation of shall-issue laws has led to a reduction in crime rates up to a certain extent. While the correlation matrix did not indicate a strong correlation between crime trends and shall-issue laws, the line graphs offer a distinct visual representation, revealing a discernible decrease in crime rates in states where shall-issue laws are applicable.

To establish more definitive and rigorous statements, further analysis involving the testing of specific hypotheses through regression models is warranted. This would involve designing regression models with different crime types as dependent variables and shall-issue laws as independent variables. The line graphs provide an initial visual insight, and conducting formal statistical tests would contribute to a more comprehensive understanding of the relationship between shall-issue laws and crime rates.

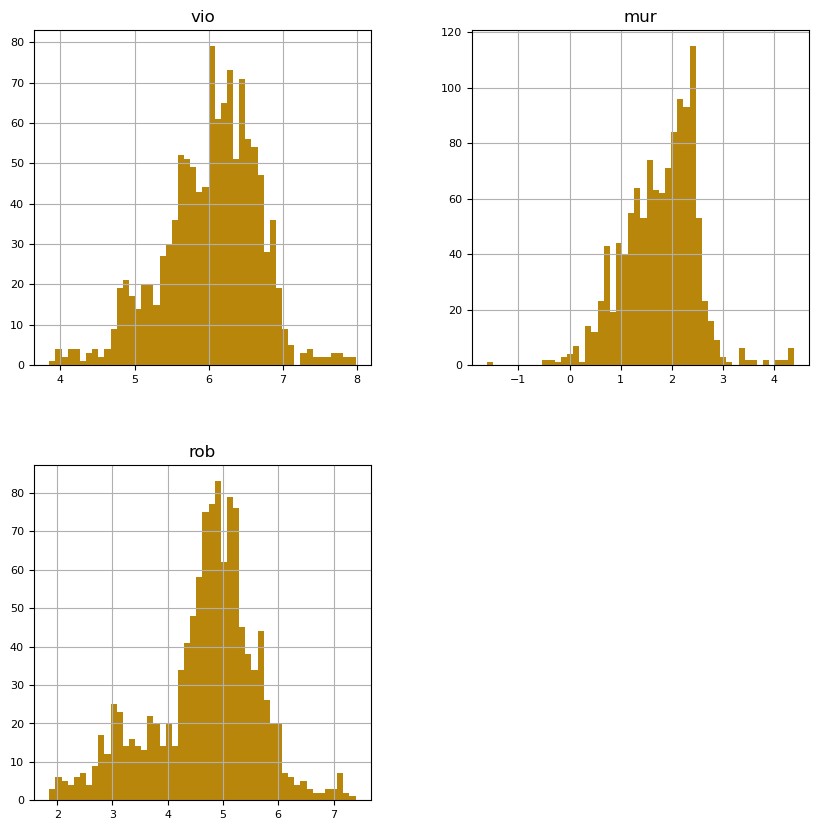
**3.4 Detecting Outliers in the Statewise crime data using Boxplots**



In our examination of crime rates across the 51 states, one state, denoted as State 11, was identified as an outlier due to its crime rate significantly deviating from the overall trend. Using robust statistical methods, we made the considered decision to exclude this outlier from our dataset. This strategic removal was undertaken to safeguard the reliability of our analysis, preventing the potential distortion of our findings.

To ensure transparency and assess the impact of State 11 on our overall model, we plan to conduct additional analyses. Specifically, we will run our statistical model both with and without the data point from State 11. If the inclusion or exclusion of State 11 results in significant changes to our model's outcomes, we will thoroughly document these effects and, if warranted, make an informed decision on whether to retain or exclude the outlier in our final analysis. This iterative approach underscores our commitment to meticulous analysis and the robustness of our conclusions.

## 3.5 Log Transformation of the Dependent Variables



In response to the skewed distributions observed in our dependent variables, a log transformation was applied as a corrective measure. The purpose of this transformation was to address the asymmetry in the data distributions and promote a more balanced and normalized spread. Skewness in the dependent variables can potentially impact the reliability and interpretability of statistical analyses, and log transformations serve as a common approach to mitigate such effects.

The log transformation was chosen due to its ability to compress larger values, thereby reducing the impact of extreme observations and bringing the distribution closer to symmetry. This aligns with the assumption of normality often required by statistical models, facilitating more robust and accurate analyses.

# 4. Regression Modelling

## 4.1 Regression with Violence as a Dependent Variable

In this phase, we initiated the construction of a simple linear regression model that encompassed all variables within the dataset. This model was designed to comprehensively examine the overall impact on the violent crime rate (denoted as 'vio'), serving as a representative indicator of the general crime rate. Subsequently, a systematic approach, akin to backward selection modeling, was implemented. This involved leveraging observations from statistical significance tests and assessing collinearity to identify and eliminate potentially problematic variables.

The refinement process aimed to streamline the model by iteratively excluding variables that did not contribute significantly or introduced collinearity issues. Simultaneously, additional variables, such as interaction terms and quadratic terms, were considered for inclusion to capture nuanced relationships. This iterative approach reflects a dynamic modeling strategy that adapts to the evolving understanding of the data.

The finalized regression model for the violent crime rate serves as a foundational framework for delving into more specific sub-crime rates, such as robbery ('rob') and murder ('mur'). By extending insights gained from the initial model, the subsequent models aim to elucidate the factors influencing these specific crime categories. This hierarchical approach facilitates a nuanced exploration of the multifaceted dynamics underlying different forms of criminal activity.

**(a)** ln(vio) = b0 + b1 incarc\_rate + b2 pb1064 +b3 pw1064 + b4 pm1029 + b5 pop + b6 avginc + b7 density + b8 stateid + b9 shall +b10year



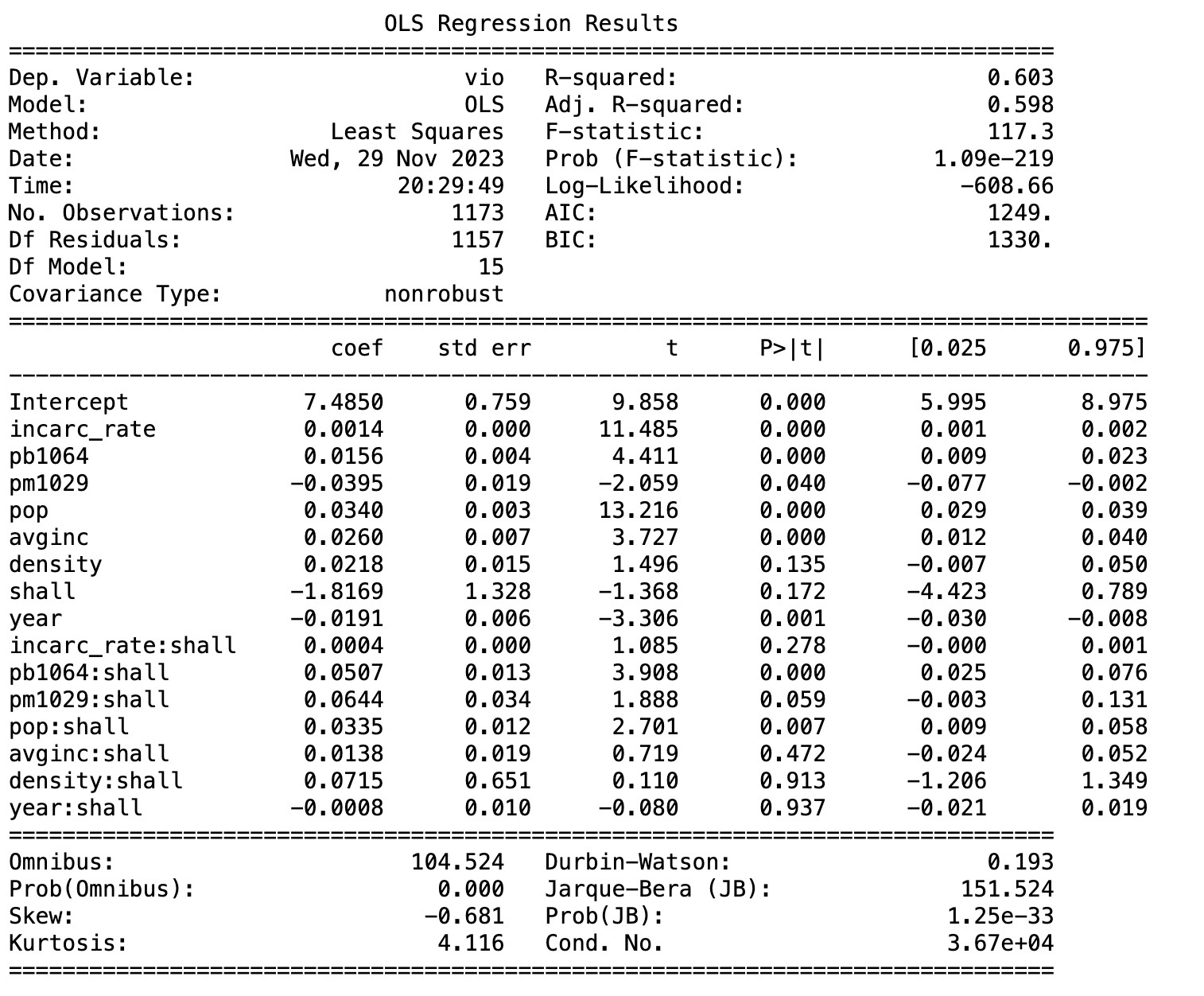
The preceding model successfully captures 59.3% of the overall variability in violent crime, accounting for the influence of explanatory variables. The regression results indicate that, with the exception of 'density' and 'avginc,' all variables significantly differ from zero, as evident from their low p-values, except for 'density' and 'avginc,' which exhibit non-significant t-values and p-values greater than 0.05.

Interpreting estimates from the output is challenging due to the presence of highly correlated explanatory variables ('pb1064' and 'pw1064') in the Ordinary Least Squares (OLS) equation. Additionally, it's noted that state IDs and years should be treated as indicator variables in subsequent regression models. An alternative approach involves removing these variables from the model to evaluate optimal estimators.

The primary insight from the regression analysis is that, in states with shall-issue laws, the violent crime rate is estimated to be 34% lower. This underscores the potential impact of shall-issue laws on reducing violent crime.

This problem can be adressed with help of removal of pw1024 variable that will help us to avoid the problem of multicollinearity as it is highly correlated with vio data. We will carry out the effect of this variable in terms of pb1024, by treating this as an indicator variable as it is positively and strongly correlated with violence rate. Moroever we can also add interaction variable for interaction of shall variable with all other variables.

**(b)** ln(vio) = b0 + b1 incarc\_rate + b2 pb1064 + b3 pm1029 + b4 pop + b5 avginc + b6 density + b7 shall + b8 year + d1 incarc\_rate:shall + d2 pb1064:shall + d3 pm1029:shall + d4 pop:shall + d5 avginc:shall + d6 density:shall + d7 year:shall



The Pooled Ordinary Least Squares (OLS) model, incorporating interactions between 'shall' and all other explanatory variables, successfully accounts for 60.3% of the overall variation in the dataset. However, the marginal increase in explanatory power, merely 1%, compared to our basic model, suggests that the inclusion of numerous interaction terms with 'shall' may not be substantially enhancing the model's performance.

Interestingly, the introduction of these interaction variables has led to the 'shall' variable itself becoming statistically insignificant. This indicates that the presence of interactions may be overshadowing the independent impact of 'shall' on the dependent variable. Furthermore, upon closer examination, it is observed that not all interaction variables are statistically significant. Specifically, only the interactions involving 'popshall' and 'pbshall' exhibit significance, while the others fail to differ significantly from zero. This raises questions about the relevance and necessity of the multitude of interaction terms in explaining the variation in the data.

In summary, while the expanded model with interactions has a slightly higher explanatory capacity, the significance of the 'shall' variable and the limited significance of most interaction terms prompt a critical evaluation of the model's complexity and the meaningfulness of the added interactions.

A more refined version of the above regression model can be written by removing interactions variables that are insignificant.

**(c) Heteroskedasticity Test:**

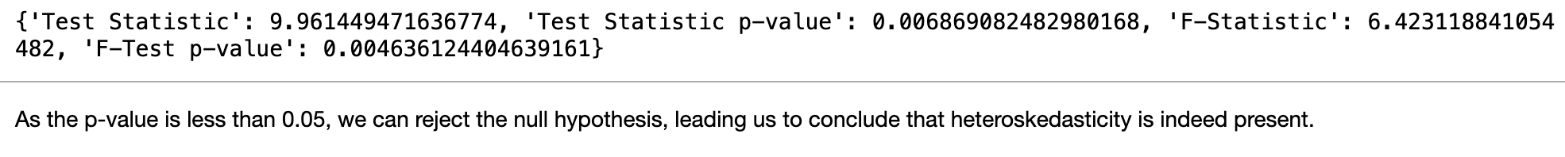
**White Test:**

### H0: Heteroskedasticity present

### H1: Heteroskedasticity not present

The White test is employed to examine whether heteroscedasticity, a violation of the homoscedasticity assumption in regression models, is present. Homoscedasticity assumes that the residuals' variance remains constant across all independent variable values. The null hypothesis of the White test asserts constant error variance (homoscedasticity), while the alternative hypothesis posits varying error variance (heteroscedasticity). If the p-value is below a chosen significance level (e.g., 0.05), the null hypothesis is rejected, indicating evidence of heteroscedasticity.

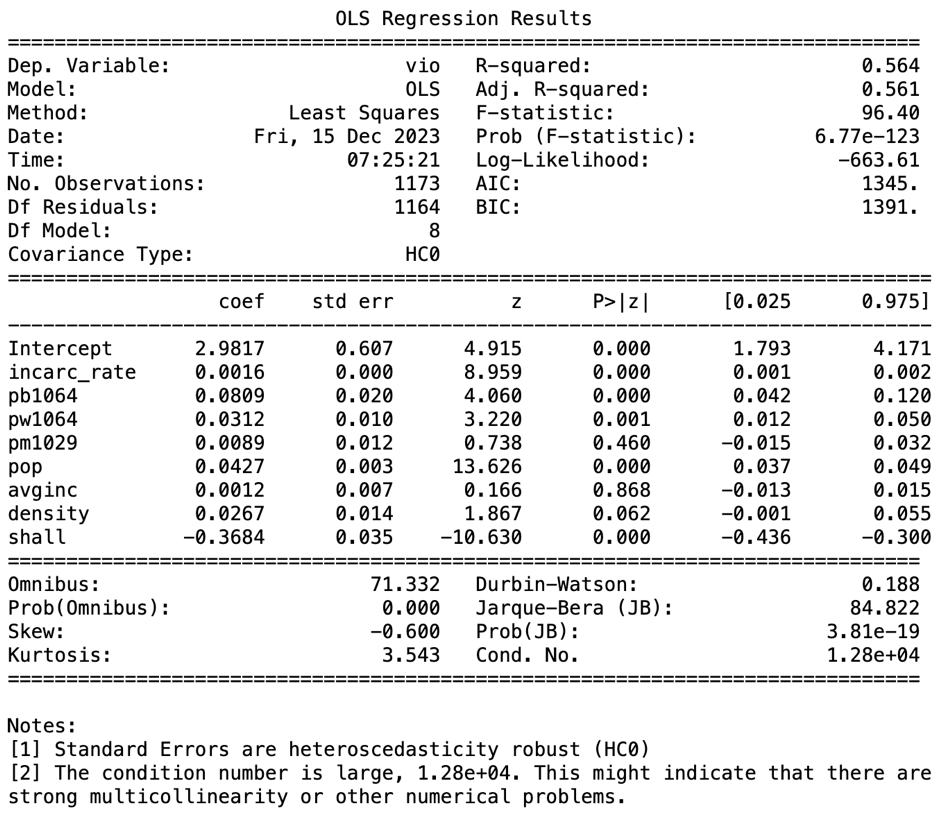
To conduct the White test, the regression model is initially estimated, and the residuals are obtained. Subsequently, a regression of squared residuals on independent variables and their squared terms is performed. The test statistic, derived from the regression output, is compared to a chi-squared distribution with degrees of freedom equal to the number of independent variables plus one.



The Breusch-Pagan test for heteroskedasticity, using White's test, yielded a test statistic of 6and a p-value of 0. This result strongly rejects the null hypothesis, providing evidence of heteroskedasticity in the model. Heteroskedasticity's presence can lead to inefficient estimates and render the pooled least square estimator ineffective.

**(d) Pooled OLS with Robust Standard Errors:**

To enhance the accuracy and reliability of our standard errors for statistical inference, we incorporated cluster-robust standard errors into the Pooled Ordinary Least Squares (OLS) Estimator. This approach enables us to rectify the bias and inconsistency introduced by heteroskedasticity and serial correlation. The ensuing results, adjusted for cluster robust standard errors, are detailed below.



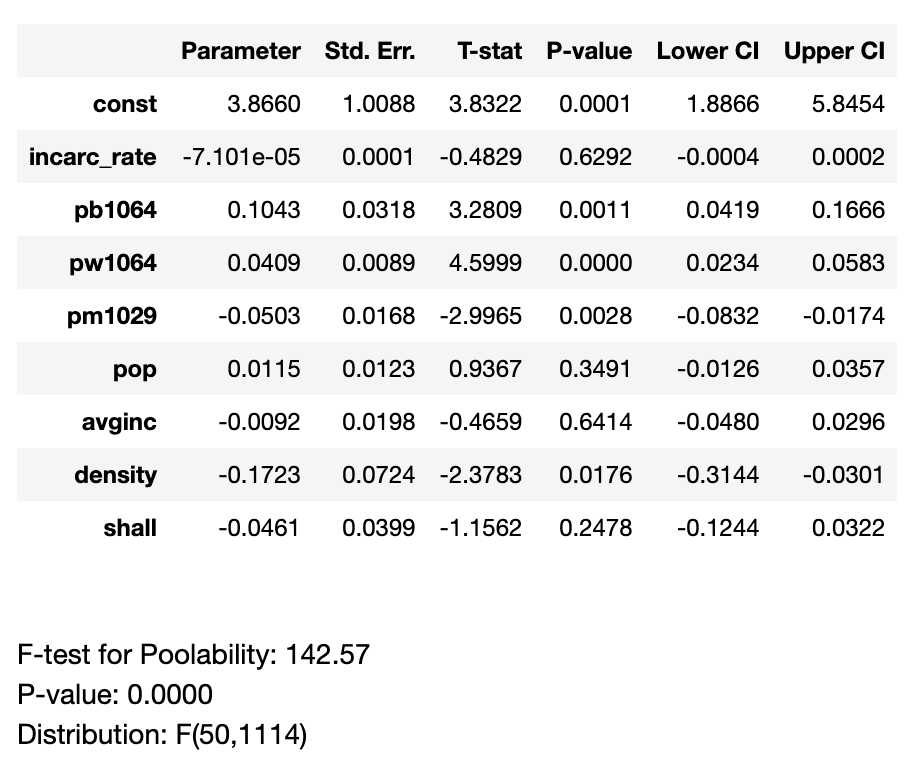
We noticed that the cluster-robust standard errors tended to be greater than those derived from the Pooled Ordinary Least Squares (OLS) estimator. This suggests that the latter method tended to overstate the precision of the estimates. Consequently, we seek a more effective and accurate estimator.

**Heterogeneity:**

The dataset comprises entities (such as states) with diverse characteristics, and their responses to explanatory variables may differ. Factors like varying baseline crime rates, distinct levels of policing or incarceration rates, and diverse cultural attitudes toward gun ownership contribute to this heterogeneity. Modeling the relationship between the dependent variable (violent crime rate) and explanatory variables (e.g., shall-issue laws, incarceration rate) encounters challenges due to potential variations in the effects of these variables across states.

This heterogeneity poses challenges to the assumption of Independence of Observations, suggesting the presence of unobserved factors that systematically influence both independent and dependent variables. Such heterogeneity can lead to serial correlation in errors, potentially biasing standard errors and impacting the statistical significance of estimated coefficients. Consequently, addressing this heterogeneity is crucial for obtaining more accurate and reliable modeling results.

**(e) Entity Fixed effects Model:**

****

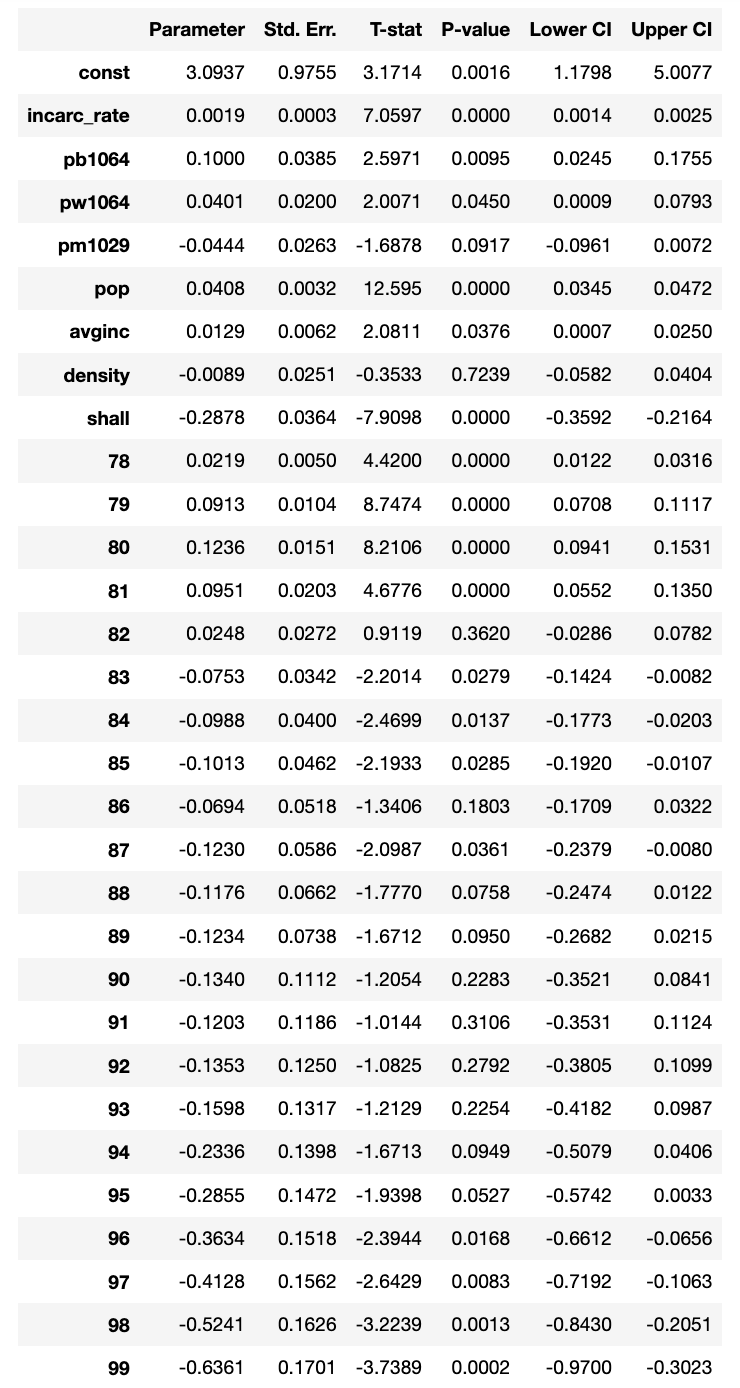
In the Entity Fixed Effects model with the entity represented by 'stateID,' the regression interpretations for each explanatory variable are as follows:

* shall: Holding all other factors constant, the presence of a "shall issue" law is associated with a 4.6% decrease in violence. However, this variable is not statistically significant at the 5% level, with a p-value of 0.2478.
* incarc\_rate: this is not statistically significant at the 5% level.
* pb1064: A one-unit increase in the percentage of the population that is black and aged 10-64 is associated with a 10.43% increase in violence.
* pw1064: A one-unit increase in the percentage of the population that is white and aged 10-64 is associated with a 4.09% increase in violence.
* pm1029: A one-unit increase in the percentage of the population aged 10-29 is associated with a 5.1% decrease in violence.
* pop: but this variable is not statistically significant at the 5% level.
* avginc: this is not statistically significant at the 5% level.
* density: this variable is not statistically significant at the 5% level.

Compared to a pooled OLS model, the entity fixed effects model is more appropriate for estimating the impact of policy interventions like the adoption of "shall-issue" laws. It addresses unobserved state-specific differences by including fixed effects for each state, controlling for time-invariant variations. This model provides more precise estimates by leveraging the variation in policy adoption over time and across states.

However, it's important to note that the fixed effects model assumes a constant relationship between the dependent variable and time-invariant variables across entities, which may not always hold. The Entity Fixed Effects Model is not suitable for estimating the effect of time-invariant explanatory variables.

**(f) Entity and time Fixed effects Model:**

****

The analysis focuses on various variables in a regression model to understand their impact on the violent crime rate. The "shall" variable, indicating the presence of a "shall issue" law allowing concealed weapons, has a coefficient of -0.028. This suggests that states with such laws experience a 2.8% lower violent crime rate. Year fixed effects (I(year)) show a consistent positive trend, indicating an overall idecrease in the violent crime rate from 1977 to 1999. The higher incarceration rates are associated with lower violent crime rates. Other demographic variables such as the percentage of the population that is black, white, or male, along, average income (avginc), and (density), do not show statistically significant effects on the violent crime rate.

Despite the presence of a "shall issue" law seemingly indicating a lower crime rate, the analysis concludes that this effect is not statistically significant. The estimated 2% reduction in crimes in states with such laws is not statistically different from zero, suggesting that the "shall" law does not have a significant impact on the overall crime rate. This finding aligns with the results obtained from the entity fixed model, reinforcing the insignificance of the "shall" law's influence on the crime rate.

The Wald test was employed to compare the Entity Fixed Effects Model with the Entity and Time Fixed Effects Model, differing in the incorporation of a time variable. Model 1 comprises solely state fixed effects, while Model 2 includes both state and time fixed effects. The test yielded a chi-square value of 363.96 with 22 degrees of freedom and a p-value <<<0.05. This result indicates that Model 2 significantly outperforms Model 1. The incorporation of time fixed effects has notably enhanced the model's explanatory power for the dependent variable, ln\_vio, compared to utilizing only entity fixed effects. Consequently, it is advisable to proceed with the Entity and Time Fixed Effects Model for further analysis.

## 4.2 Regression with Murder as a Dependent Variable

In our analysis of murder as a dependent variable, we found that murder, violence, and robberies were all highly correlated. This meant that we could only select one of these variables as a dependent variable and could not use the other two variables as explanatory variables.

**4.2.1 Multi- Collinearity Test:**

We check multi-collinearity in the data to see whether we have two more predictor variables correlated to each other. This might create the problem of collinearity in the model we perform. Thus to check this we conducted the test.

A computer screen with white text

Description automatically generated

Looking at the above result of multi – collinearity we can say that variable “pb1064” & “pw1064” are highly corelated to each other. To remove the issue of multi collinearity we further dropped “pb1064” variable from our dataset.

After dropping one of the corelated we variable we ran the multi collinearity test again to check, whether we still have correlation present in the data.

A screenshot of a computer

Description automatically generated

The new test suggest that we do not have any correlation present after omitting one of the variable.

**4.2.2 Pooled OLS Regression Model:**

A screenshot of a computer

Description automatically generated

Initially in this pooled model, we look at all the other variables except robberies and violence. Additionally, we use the log transformed values of murders and the incarceration rate. The Dataset that we are dealing with has 51 states with 23 observations for each state corresponding to a year.

Here is the Regression Equation for the model:

log(mur) = -3.0974 -0.2214 shall + 0.6863 log(incar\_rate) -0.0405 avginc + 0.0329 pop -0.0106 pw1064 +0.1489 pm1029 + 0.0983 density

The R2 value of 65.3% indicates that the independent variables are able to explain a significant amount of the variation in the dependent variable. All of the explanatory variables are also significant, which means that they are all able to explain a statistically significant amount of the variation in the dependent variable.

The Durbin Watson Test value of 0.467 suggests that there is positive autocorrelation in the model.

The results of this analysis suggest that the independent variables are able to explain a significant amount of the variation in the dependent variable.

**Interpretation of the Model:**

Although the R2 is relatively high, this model might not be the best model for our job. This is explained partly by looking at all the variables in the model and also by seeing the warning that there might be strong multicollinearity in the model as well as the presence of autocorrelation in the model.

1. **Shall law**: The coefficient of -0.22 suggests that if Shall Law is present in a state, then the murder rate decreases by 22%. This seems to be a very high number and is possibly biased.
2. **Incarceration rate:** The coefficient of 0.686 suggests that as the incarceration rate increases by 1% the murder rate increases by 0.686%. We expected that as the incarceration rate increases, the murder rate should decrease. However, it might be the case that as the murder rate or the overall crime rate in the state is higher, the incarceration rate would be forced to be higher instead of it being a policy of the state.
3. **Average Income:** As the Average income increases by 1000$, the murder rate decreases by 0.04%. This is in line with what we expected that if the income of a state is higher the crime rate should naturally be lower.
4. **Population:** As the population increases by 1 million, the murder rate increases by 3%. This is again in line with what we expected.
5. **Percentage of White Males:** As the population of white males increases by 1% in a state, the murder rate increases by 1%.
6. **Percentage of Males:** As the population of Males increases by 1% in a state, the murder rate increases by 14%. Looking at all the demographic information, we can say that the model is biased because the values are very high.

**4.2.2 Heteroskedasticity**

As previously mentioned, our analysis involves panel data, and thus, there might be a potential issue of Heteroskedasticity across the dataset. When such issues arise, it's crucial to acknowledge that the model remains linear, unbiased, and consistent but may no longer be the best or most effective. To refine the model, it becomes essential to assess the presence of these problems in the data. Heteroskedasticity, in particular, signifies that the error is not constant across variables, thereby violating the assumption of constant standard errors. To investigate this, we conduct a heteroskedasticity test.

Our approach involves two methods for testing heteroskedasticity:

**i) Informal Approach:**

We begin with an informal examination of heteroskedasticity by visually inspecting graphical representations of the residuals. Residuals, representing the disparities between observed and predicted values in a regression model, offer an initial insight into the potential presence of heteroskedasticity.

**ii) Formal Approach:**

Transitioning to a more rigorous evaluation, we employ a formal approach to rigorously test for heteroskedasticity in our data. This involves employing statistical methods to provide a comprehensive assessment of the presence and nature of heteroskedasticity. The results from both the White test and the Bausch-Pagan test strongly suggest the presence of heteroskedasticity in the data.

For the White test, the substantial LM Statistic of 222.1640 and an exceptionally low p-value of 1.714e-29 indicate a high likelihood of heteroskedasticity. Similarly, the Bausch-Pagan test supports this finding with a notable LM Statistic of 29.21 and an extremely low p-value of 0.00013, further suggesting the presence of heteroskedasticity.

In summary, the elevated LM statistics and the significantly low p-values from both tests consistently signal the existence of heteroskedasticity in the data. These results imply that the variance in error terms across different variables is not uniform. To address this issue, a recommended corrective action involves performing **"Pooled OLS Regression with Robust Errors".**

**4.2.4 Pooled OLS Regression with Cluster Robust Errors:**

**Here is the Final Regression equation:**

A screenshot of a computer

Description automatically generated

With an R2 value of 65.3%, the current model maintains its ability to explain a substantial portion of the variation in the dependent variable, consistent with the earlier OLS model. Notably, all explanatory variables are statistically significant, except for "pw1064," indicating that each variable, excluding "pw1064," significantly contributes to explaining the variation in the dependent variable.

For panel data analysis, we recommend employing Pooled OLS Regression with Cluster Robust Errors. This approach helps account for potential correlations within clusters and enhances the robustness of the regression analysis, especially when dealing with panel data.

**4.2.5 Fixed Effect Model – Entity Fixed effects**

In our analysis, we employed a fixed effect model to account for individual unobservable heterogeneity among various states in the U.S., while if these characteristics remain constant over time. This approach allows us to explore the assumption of no correlation between the distinctive characteristics of each year.

The fixed effects model focuses on examining variations within entities, specifically years, rather than between entities. This implies that the model delves into the variations occurring within a particular year across different entities. By isolating the year-specific effects, we gain insights into how these time periods evolve across various entities, enhancing the precision of our analysis.

A screenshot of a black screen

Description automatically generated

A table of numbers on a black background

Description automatically generated

**Interpretation of the Model:**

The R^2 value of 0.5922 signifies that approximately 59.22% of the variance in the natural logarithm of the murder rate is accounted for by the model. In simpler terms, the model, incorporating individual-specific effects and various factors, captures about 59.22% of the total variability observed in murder rates within different entities and time periods. Among the variables, log\_incarc\_rate, density, and avginc emerge as the only statistically significant contributors.

i) log\_incarc\_rate: The coefficient suggests that 1 percent increase in incarceration rate is associated with a 0.19 percent increase in murder rate.

ii) Density: The coefficient indicates that a one-unit increase in population density is linked to a 0.8% decrease in the murder rate. However, this is not significant.

iii) Avginc: The coefficient implies that a one-unit increase in average income corresponds to a 1% increase in the murder rate.

iv) shall: If shall law is present in a state then the murder rate decreases by 28%.

v) pb1064: If the percentage of blacks between age of 10 and 64 increase by 1 unit then the murder rate increases by 10%

vi) pw1064: If the percentage of women between the age of 10 and 64 increases by 1 unit then the murder rate decreases by 4%.

For all other variables, the p-values exceed 0.05, leading us to conclude their insignificance in the model.

**4.2.6 Fixed Effect Model – Time & Entity Fixed effects**

To mitigate challenges arising from unobserved heterogeneity, particularly factors like economic recessions and federal laws that consistently influence murder rates across all entities and evolve over time, our approach involves the implementation of a Time and Entity Fixed Model. This model addresses issues of endogeneity in the presence of such unobserved factors, aiming to produce unbiased estimates by accounting for both temporal and cross-sectional variations.

A screenshot of a computer screen

Description automatically generated

**Interpretation of the Model:**

The R^2 value of 0.1191 indicates that approximately 11.91% of the variance in the murder rate is explained by the model. Simply put, the model, considering individual-specific effects and various factors, captures about 11.91% of the overall variability observed in murder rates across different entities and time periods. Notably, density, avginc, and pm1029 stand out as the only statistically significant variables, while others do not contribute significantly. It is noteworthy that, in this Fixed Effect Entity and Time variant model, log\_incarc\_rate is deemed insignificant with a significance cut-off of 0.05.

i) **Density:** The coefficient suggests that a one-unit increase in population density is associated with a substantial 46.6% decrease in the murder rate.

ii) **Average Income** : The coefficient indicates that a one-unit increase in population density is linked to a 0.8% decrease in the murder rate.

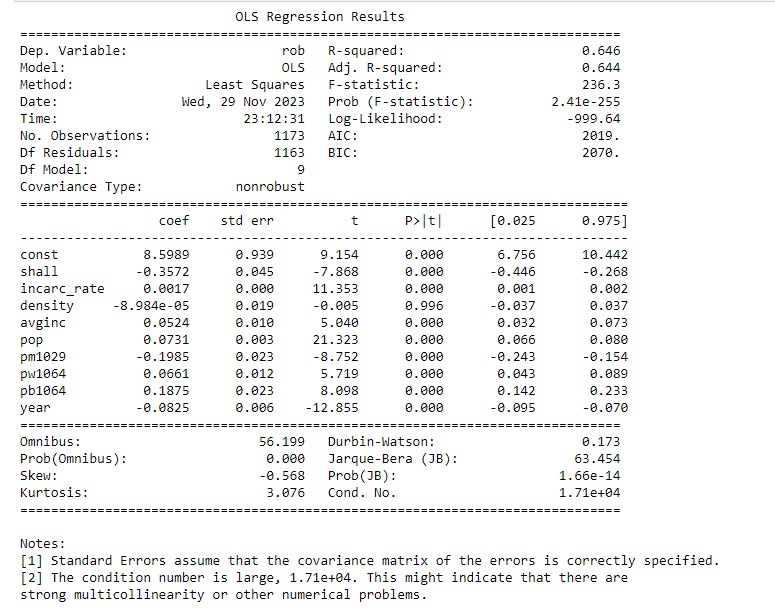
iii) **Percentage of Males:** As the population of Males increases by 1% in a state, the murder rate increases by 0.7 units. For all other variables, the p-values exceed 0.05, leading us to conclude their insignificance in the model.

The model can control for factors that differ across entities but are constant over time. Additionally, the model can eliminate bias from unobservable differences in variables that occur over time but are constant over entities

## 4.3 Regression with Robbery as a Dependent Variable

In this sub analysis we look at robbery as the dependent variable and rest of the independent variables as the explanatory variables.

**4.3.1 OLS Regression**



**Interpretation of the Model:**

The equation based on the regression results for the natural logarithm of the robbery rate is as follows:

*ln(rob)=8.5989−0.3572×shall+0.0017×incarc\_rate−0.00008984×density+0.0524×avginc+0.0731×pop−0.1985×pm 1029+0.0661×pw1064+0.1875×pb1064−0.1786×shall−0.0825×year*

The model explains approximately 64.6% of the variation in the natural logarithm of the robbery rate (ln(rob)), as indicated by the R-squared value.

**Significance of Variables:**

**Shall Law (shall):** The coefficient of -0.3572 suggests that, holding other variables constant, the presence of a Shall Law is associated with a 35.72% decrease in the robbery rate.

**Incarceration Rate (incarc\_rate):** A one-unit increase in the incarceration rate is associated with a 0.17% increase in the robbery rate.

**Population (pop):** A one-unit increase in population is associated with a 7.31% increase in the robbery rate.

**Average Income (avginc):** A $1000 increase in average income is associated with a 5.24% increase in the robbery rate.

**Percentage of Population Male Ages 10-29 (pm1029):** A one-unit increase in the percentage of the population that is male, ages 10-29, is associated with a 19.85% decrease in the robbery rate.

**Percentage of Population White Ages 10-64 (pw1064):** A one-unit increase in the percentage of the population that is white, ages 10-64, is associated with a 6.61% increase in the robbery rate.

**Percentage of Population Black Ages 10-64 (pb1064):** A one-unit increase in the percentage of the population that is black, ages 10-64, is associated with an 18.75% increase in the robbery rate.

**Year (year):** A one-unit increase in the year is associated with an 8.25% decrease in the robbery rate.

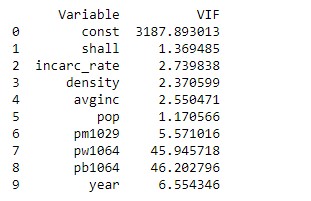
The density variable does not appear to be statistically significant in predicting the natural logarithm of the robbery rate.

The presence of Shall Law, incarceration rate, population, average income, percentage of population male ages 10-29, percentage of population white ages 10-64, percentage of population black ages 10-64, and the year are all statistically significant predictors of the natural logarithm of the robbery rate.

The Durbin-Watson test indicates positive autocorrelation, implying that there may be correlated errors in the model.

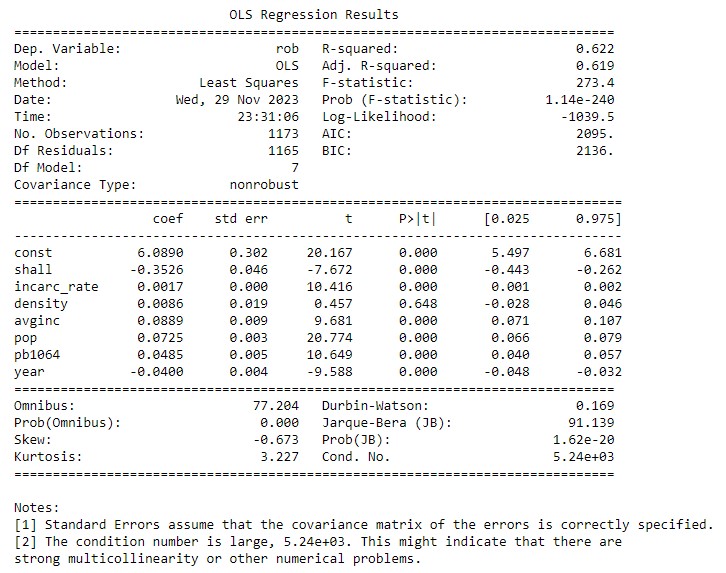
**4.3.2 Multi- Collinearity Test:**

To address the multicollinearity we shall look at the VIF values.



As pw1064 and pb1064 have high VIF values we shall drop them and perform a regression analysis.

Here are the results



**Interpretation of the Model:**

The equation based on the regression results for the natural logarithm of the robbery rate is as follows:

*ln(rob)=6.0890−0.3526×shall+0.0017×incarc\_rate+0.0086×density+0.0889×avginc+0.0725×pop+0.0485×pb1064− 0.0400×year*

The model explains approximately 62.2% of the variation in the natural logarithm of the robbery, as indicated by the R-squared value.

**Significance of Variables:**

**Shall Law (shall):** The coefficient of -0.3526 suggests that, holding other variables constant, the presence of a Shall Law is associated with a 35.26% decrease in the robbery rate.

**Incarceration Rate (incarc\_rate):** A one-unit increase in the incarceration rate is associated with a 0.17% increase in the robbery rate.

**Population (pop):** A one-unit increase in population is associated with a 7.25% increase in the robbery rate.

**Average Income (avginc):** A $1000 increase in average income is associated with an 8.89% increase in the robbery rate.

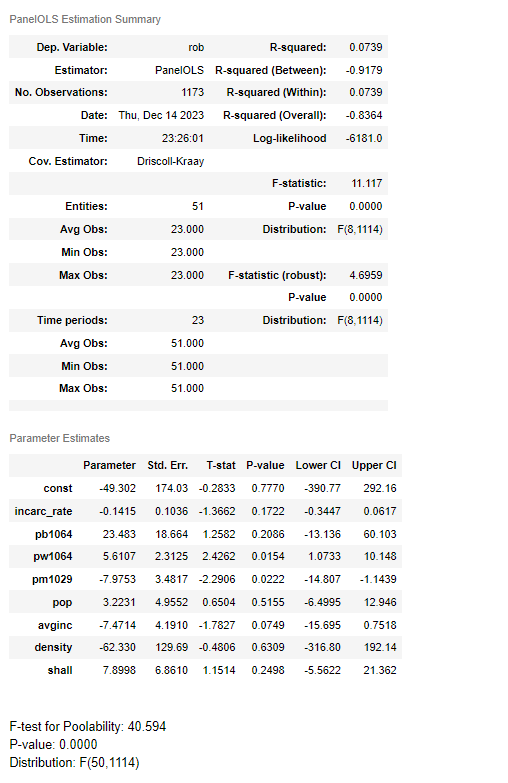
**Percentage of Population Black Ages 10-64 (pb1064):** A one-unit increase in the percentage of the population that is black, ages 10-64, is associated with a 4.85% increase in the robbery rate.

**Year (year):** A one-unit increase in the year is associated with a 4.00% decrease in the robbery rate.

The density variable does not appear to be statistically significant in predicting the natural logarithm of the robbery rate.

The presence of Shall Law, incarceration rate, population, average income, percentage of the population black ages 10-64, and the year are all statistically significant predictors of the natural logarithm of the robbery rate.

**4.3.3 Fixed Effect Model – Time & Entity Fixed effects**



The R^2 value of 0.0739 indicates that approximately 7.39% of the variance in the murder rate is explained by the model. Simply put, the model, considering individual-specific effects and various factors, captures about 7.39% of the overall variability observed in murder rates across different entities and time periods. Notably, pb1064 stand out as the only statistically significant variables, while others do not contribute significantly. It is noteworthy that, in this Fixed Effect Entity and Time variant model, log\_rob is deemed insignificant with a significance cut-off of 0.05

# 5. Conclusion

Throughout our in-depth analysis of the dataset, we faced various challenges like heteroskedasticity, serially correlated errors, and endogeneity. Our focus was on understanding the intricate relationships between variables. We further ran Pooled OLS Model, Entity Fixed Model, and Time and Entity Fixed Model—to understand between our dependent variable and independent variables. We tried to perform these tests to remove errors and problems in the model.

In conclusion we can say that the Time Entity Fixed Effect model performs the best and explains each crime rate with independent variables. Further this model helps us in rectifying the problem of heteroskedasticity, endogeneity and multi -collinearity.

In overall of our analysis, we can see that Incarceration rate has a significant effect on each crime rate in US over 51 states. Moreover, a higher percentage of young males in states exhibited a correlation with a decrease in the each crime rate. Intriguingly, the overall crime rate demonstrated a general decline from 1977 to 1999. All the models have a negative effect on the dependent variables – violence, murder, robbery. Through this we suggest that each state should tightening and improving the policies can really help in reducing the crime rate in US over 51 states.