BUAN 6312.003: Applied Econometrics & Time Series Analysis

Final Project Report

Do more guns reduce crime?

Prepared for:

Dr. Moran Blueshtein

Prepared by:

Angela Chow | axc180045

Aneesa Noorani| axn180021

Rhea Woodcock | vrw180001

April Wu | cxw180004

December 13, 2019

Table of Contents

I. Introduction & Problem Statement

From 1976-1999, a wave of states passed shall-issue laws that mandated the state to grant licenses to carry a concealed handgun, provided the licensee meets certain criteria. This contrasts with may-issue laws, which do not require the state to grant licenses, even if the licensee meets certain criteria. In recent years, several states have converted to unrestricted carry, where no permit is required to carry a handgun.

While there are many contributing factors to this increased leniency in carrying of a concealed weapon (CCW) laws, a major one is the lobbying presence of the National Rifle Association (NRA) since 1975. The NRA is a proponent of the motto "More Guns, Less Crime" and has asserted that the "only way to stop a bad guy with a gun is a good guy with a gun." Proponents of shall-issue laws claim that, from a criminal's perspective, allowing law-abiding citizens to carry concealed handguns increases the risk associated with engaging in unlawful activities, and therefore is a deterrent for criminals. However, if we restrain law-abiding citizens from owning and carrying handguns, criminals (who will find access to arms regardless) will have less fear of encountering armed resistance, and thus more incentive to engage in crime.

Shall-issue law opponents argue that more guns leads to more crime. Even when crime is not intended, armed citizens may be more likely to escalate an angry dispute into a criminal homicide or wounding. A second argument is that allowing law-abiding citizens to carry concealed weapons may initiate an "arms race," where criminals arm themselves with higher-powered weapons and will shoot more quickly when feeling threatened. Third, opponents of shall-issue laws argue that since as many as one million guns are stolen each year, wider access to guns for the law-abiding population will undoubtedly lead to more guns getting into the hands of criminals.

CCW laws are a topic of hot debate. In this report, we explore which side of thought is supported by historical data. We use econometric theory to study what effect shall-issue laws have on state crime.

Our hypothesis is that states that shall-issue laws reduce crime. We explore this hypothesis by modeling the data set in the following ways: 1) pooled ordinary least squares regression, 2) entity-fixed effects, 3) time-fixed effects, and 4) time and entity-fixed effects.

II. Data Description

This dataset is a balanced panel, with data on all 50 US states, plus the District of Columbia, giving us 51 'states.' We have data on each year from 1977 to 1999 for every state. In other words, there are no missing values. Each observation is for a given state for a given year. There are a total of 51 states × 23 years = 1173 observations.

The index/identifier columns are: *year* and *stateid*. The dependent variables of interest are: *vio*, *rob*, and *mur*. The rest of the variables are the independent variables. Below is a table that outlines each of the variables in the dataset and their descriptions. It is important to note that the violent crime rate includes robbery and murder, as well as other types of violent crime.

Variable	**Definition**
vio	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	= 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)

III. Data Preprocessing

In our initial exploratory data analysis, we found that the dependent variable *vio*, violent crime rate, was right-skewed, as shown in the histogram on the left (**Figure 1**). In order to normalize the distribution we generated the variable *lnvio,* which is the natural log of *vio*. As shown in the histogram on the right (**Figure 2**), the distribution of *lnvio* more closely approximates a normal distribution.

By running our regressions on *lnvio*, we improve the interpretability of our models. For example, if a state has a shall-issue law in place (*shall* = 1), we can estimate the percentage change in violent crime per 100,000 population as (100*β_{shall})%. A percentage change is more useful than the number of incidents, as we try to understand the effect *shall* has on crime.

Fig 1: Distribution of *vio*

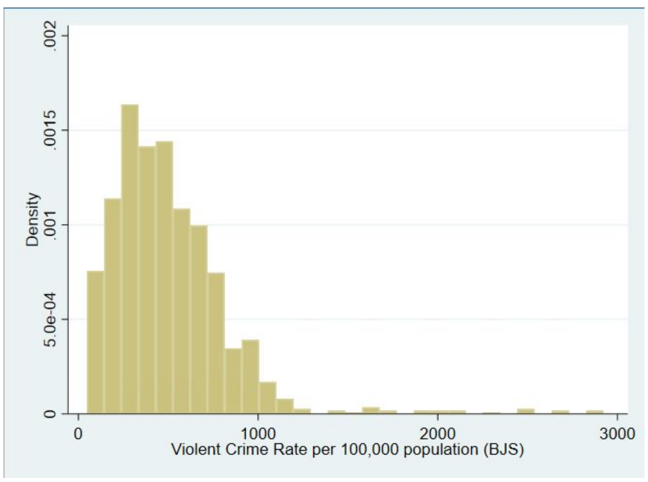
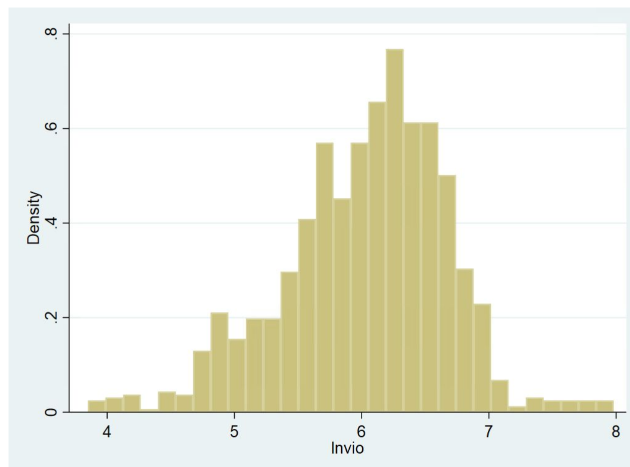


Fig 2: Distribution of *lnvio*



Similarly, we found that *pop* has a right skew (**Figure 3**). We took the natural log, forming a new variable *lnpop* in order to normalize and improve interpretability (**Figure 4)**. Now we can estimate that for every 1% increase in population, violent crime rate is expected to change by β_{lnpop}%.

Fig 3: Distribution of *pop*

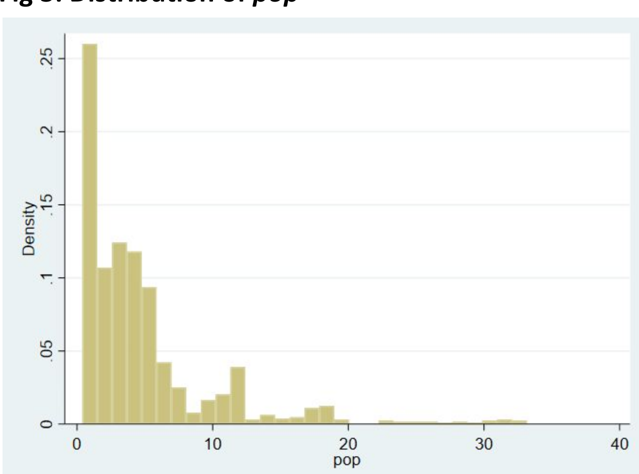
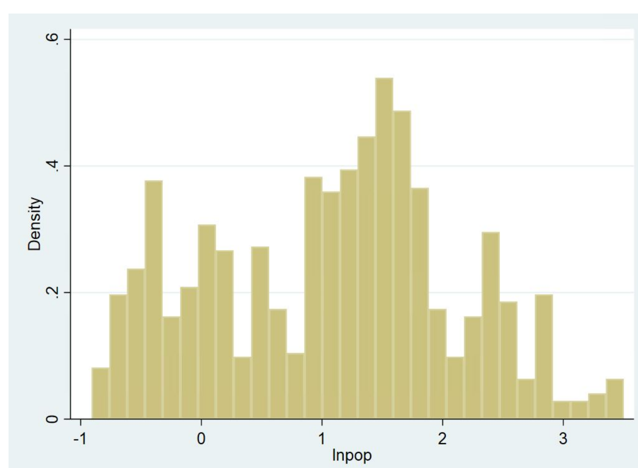


Fig 4: Distribution of *lnpop*



We also considered transforming *density* due to its right skew. However, as discussed in **Section IV: Exploratory Data Analysis**, we found that the skew is caused by a single 'state' - D.C. We decided not to transform *density* for this outlier 'state'.

For the remaining independent variables, there was either no extreme skew found, or there was a skew but it did not make sense to transform the variable by natural log. For example, *pb1064* measures the percent of the population that is black, ages 10 to 64. By taking the natural log of a percentage, we would lose interpretability of the estimate - a percentage change of a percentage is not meaningful.

IV. Exploratory Data Analysis

After reviewing basic summary statistics, we took note of each variable's distribution. Though income data are typically skewed right, the distribution in this dataset is fairly normal (**Figure 5**) and does not need to be log-transformed. This is likely because avg*inc* is a statewide average. Population density is also interesting: while all other observations were clustered around 1,000 people or fewer per square mile, Washington, D.C. has a population density that is 10 times higher (**Figure 6**). This is not surprising, as it is the only 'state' that consists of a single, urban city. As discussed previously, we treated D.C. as an outlier 'state'. Rather than transform *density* for all the observations, we decided to allow our fixed effects (FE) models discussed in **Section V: Data Modeling** to address D.C.'s heterogeneity.

Fig 5: Distribution of *avginc*

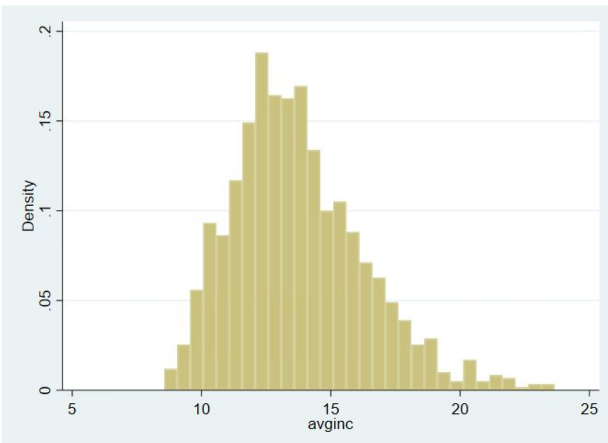
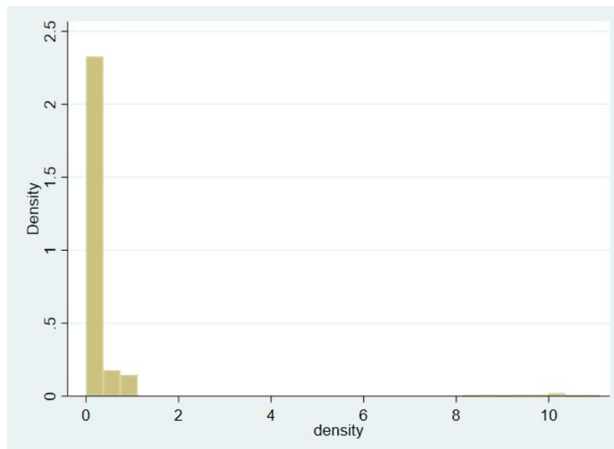


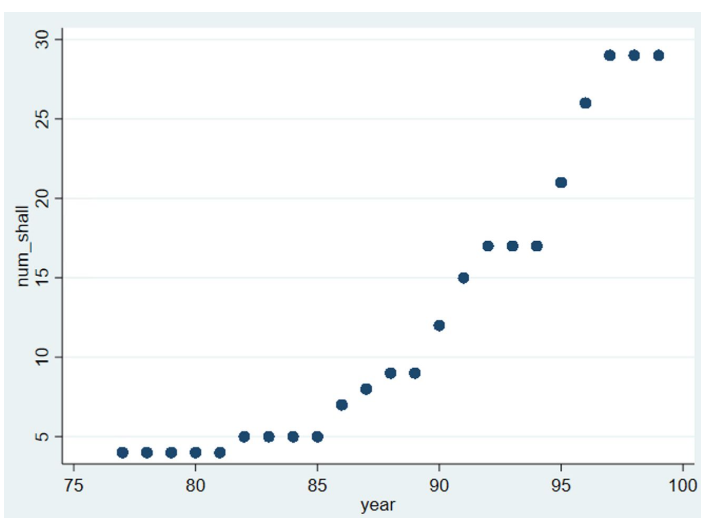
Fig 6: Distribution of *density*



Each of the crime variables (*vio*, *mur*, and *rob*) were particularly right skewed. See discussion of log transformation in **Section III: Data Preprocessing.**

The percentage of states with shall-issue laws trended sharply upward from 1985 to 1997, going from 5 to 29 states (**Figure 7**). This trend indicates that time is an important factor to consider.

Fig 7: Number of states with shall-issue law



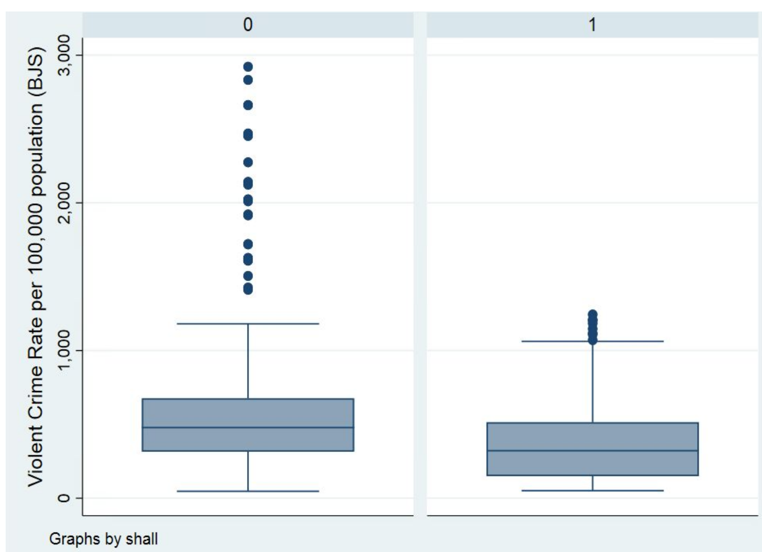
We also generated a correlation matrix (**Figure 8**) to observe the strength and direction of the linear relationship between the variables and note any relationships that stood out. The correlation between *vio* and the *pm1029* was negative, which was unexpected since young men are typically associated with higher crime rates. Additionally, a positive correlation between *vio* and *avginc* was unexpected since it does not fit with the common idea that violent crime is more common in impoverished areas; however, it is important to note that the observations are state-wide averages and do not capture county- or city-level nuances within a state.

Fig 8: Correlation matrix

	vio	mur	rob	incarc~e	pb1064	pw1064	pm1029	pop	avginc	density	shall
vio	1.0000										
mur	0.8265	1.0000									
rob	0.9071	0.7976	1.0000								
incarc_rate	0.7027	0.7096	0.5668	1.0000							
pb1064	0.5698	0.6018	0.5812	0.5308	1.0000						
pw1064	-0.5730	-0.6154	-0.5842	-0.5271	-0.9820	1.0000					
pm1029	-0.1696	0.0150	-0.0860	-0.4463	0.0162	-0.0126	1.0000				
pop	0.3190	0.0999	0.3172	0.0953	0.0581	-0.0654	-0.0975	1.0000			
avginc	0.4080	0.2206	0.4148	0.4615	0.2627	-0.1912	-0.5279	0.2152	1.0000		
density	0.6647	0.7486	0.7818	0.5593	0.5432	-0.5551	-0.0637	-0.0780	0.3433	1.0000	
shall	-0.2069	-0.1794	-0.2125	0.0424	-0.1839	0.2123	-0.2772	-0.1244	-0.0000	-0.1126	1.0000

Though *shall* shows a negative correlation with all three crime variables (*vio*, *mur*, and *rob*), this is misleading since *shall* is a categorical variable thus should be visualized via a box plot. The box plot shows that states with shall-issue laws have a lower median violent crime rate, as well as a lower upper-bound (**Figure 9**).

Fig 9: Box plot of *vio* grouped by *shall*



V. Data Modeling

Before modeling, we considered which models are most appropriate for this data set. We have a panel data set, consisting of observations for 51 states across 23 time periods. Because we do not have randomly selected data from a larger population, but rather complete data on all 51 states, we decided that a random effects model is not appropriate. We also decided that a time series model is not appropriate since we are looking at multiple entities, rather than a single entity, across time.

Below are the signs we expect for the effect each explanatory variable has on the violent crime rate:

- *shall* - negative sign; our hypothesis is that states that do have a shall-issue law in place have lower crimes rates.
- *incarc_rate* - negative or positive sign. Since higher incarceration rates are thought to act as a deterrent for criminals, this sign could be negative. However, there is also a possibility of simultaneous causality bias where if there is higher crime and police are doing their job well, there will be higher incarceration rates. This relationship could cause the sign to be positive. Regardless, we believe this variable to have a high magnitude.
- *density* - positive sign; the more densely populated an area is, the higher the violent crime rate. We know from general knowledge that crime is more prevalent in large urban cities, as compared to suburban or rural areas.
- *avginc* - negative sign; in higher income areas, we expect violent crime rates to be lower.
- *lnpop* - positive sign; similar to *density*, the higher the population, the higher the crime rate.
- *pm1029* - positive sign; a higher percentage of young males is associated with higher crime rates.
- *pw1064* - negative sign; we see from the correlation matrix that a higher percentage of whites is associated with lower crime rates.
- *pb1064* - positive sign; we see from the correlation matrix that a higher percentage of blacks is associated with higher crime rates.

A. Pooled OLS

For our first regression, we used pooled OLS with cluster robust standard errors and regressed *lnvio* on all explanatory variables except *stateid* and *year.* With this initial model, we see the effect of each variable on violent crime rate as if the data were cross-sectional data.

```
. reg lnvio shall incarc_rate density avginc lnpop pm1029 pw1064 pb1064, vce(cluster stateid)

Linear regression                               Number of obs     =      1,173
                                                F(8, 50)          =      68.00
                                                Prob > F          =     0.0000
                                                R-squared         =     0.6260
                                                Root MSE          =     .39623

                                 (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
       lnvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |    -.30425    .096395    -3.16   0.003     -.497865    -.110635
 incarc_rate |    .001628   .0005319     3.06   0.004     .0005597    .0026963
     density |   .0555991   .0413974     1.34   0.185      -.02755    .1387482
      avginc |   .0121265   .0239426     0.51   0.615    -.0359636    .0602166
       lnpop |   .2798641   .0555205     5.04   0.000     .1683479    .3913803
      pm1029 |   .0357077   .0324059     1.10   0.276    -.0293815    .1007969
      pw1064 |   .0359886   .0326003     1.10   0.275    -.0294911    .1014683
      pb1064 |    .086114   .0671312     1.28   0.205     -.048723    .2209509
       _cons |   1.945559   2.171736     0.90   0.375      -2.4165    6.307618
```

In this model, *shall, incarc_rate*, and *lnpop* are significantly different from zero at the 1% level. If a state has a shall-issue law, the state's violent crime rate is estimated to be 30.42% lower. A one unit increase in the *incarc_rate* results in an estimated 0.16% increase in violent crime rate. When the state population increases by 1%, the violent crime rate increases by an estimated 0.27%.

incarc_rate, *avginc,* and *pw1064* have positive signs, which are the opposite of what we expect. Incarceration rate can be a deterrent effect which should decrease violent crime. If the average income is higher or increasing, we should see violent crime decreasing, and we expect that a higher percentage of whites would lead to lower violent crime rates. However, *avginc* and *pw1064* are not statistically significant. Other variables (*density, pm1029,* and *pb1064*) have expected signs, but they aren't statistically significant.

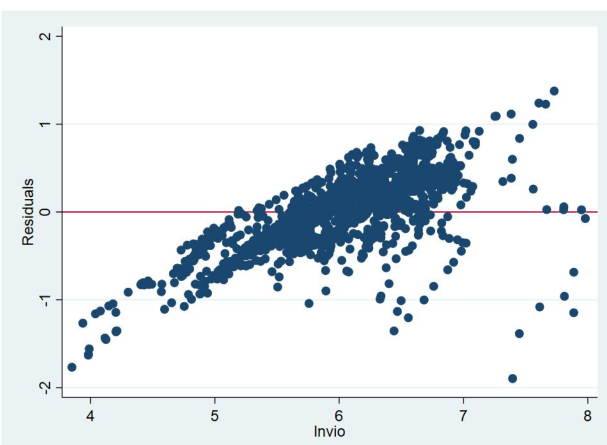
Dropping the variables that are not significant in the first regression model, we get the following result:

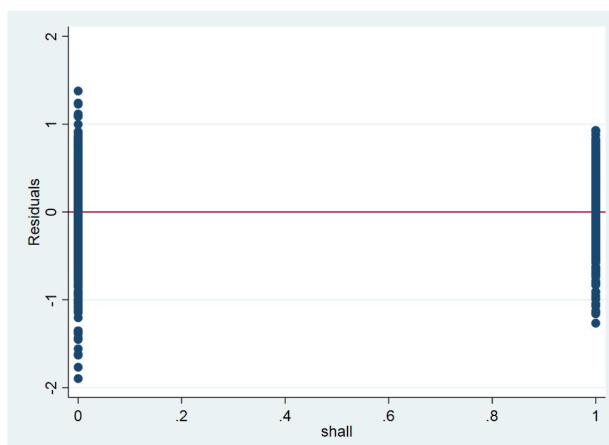
```
. reg lnvio shall incarc_rate lnpop, vce(cluster stateid)

Linear regression                               Number of obs     =      1,173
                                                F(3, 50)          =      41.41
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5753
                                                Root MSE          =     .42133

                               (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
       lnvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -.3889561   .0768116    -5.06   0.000    -.5432367   -.2346756
 incarc_rate |   .0020202   .0002054     9.83   0.000     .0016076    .0024327
       lnpop |   .2480144   .0517969     4.79   0.000     .1439773    .3520516
       _cons |    5.39668   .1161965    46.44   0.000     5.163293    5.630068
------------------------------------------------------------------------------
```

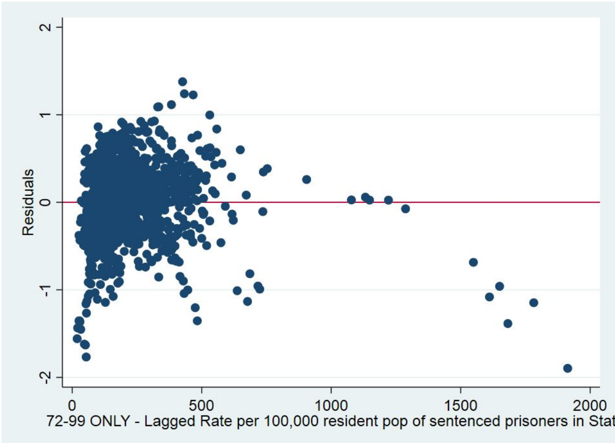
All variables are significantly different from zero at a 1% significance level. The robust standard error didn't change much for all variables: *shall* (0.096 - 0.077 = 0.019), *incarc_rate* (0.000532 - 0.000205 = 0.000327), and *lnpop* (0.0555 - 0.0518 = 0.0037). The coefficients did not change much either. If a state has a shall-issue law, the state's violent crime rate is estimated to decrease by 38.89% (compared with 30.42% in the previous model). A one unit increase in the incarceration rate results in an estimated increase of 0.20% in violent crime rate (compared with 0.16% in the previous model). When the state population increases by 1%, the violent crime rate is estimated to increase by 0.25% (compared with 0.27% in the previous model).

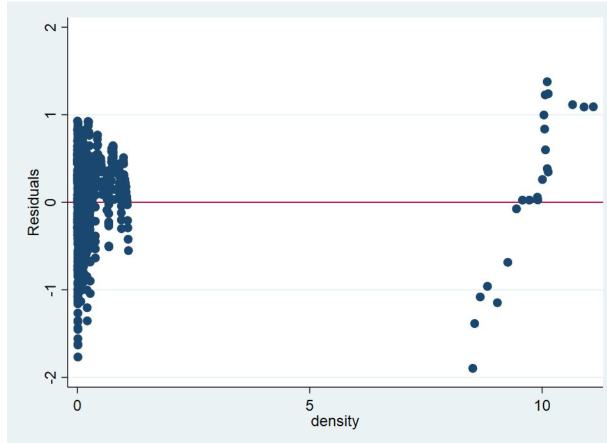
To test for heteroskedasticity, we created residual charts and conducted a White Test.

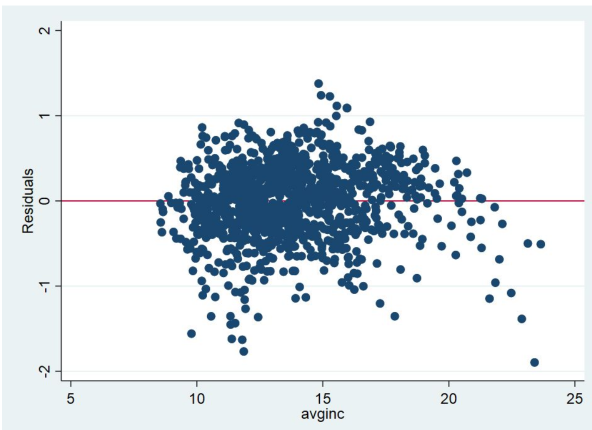
Residual Charts:

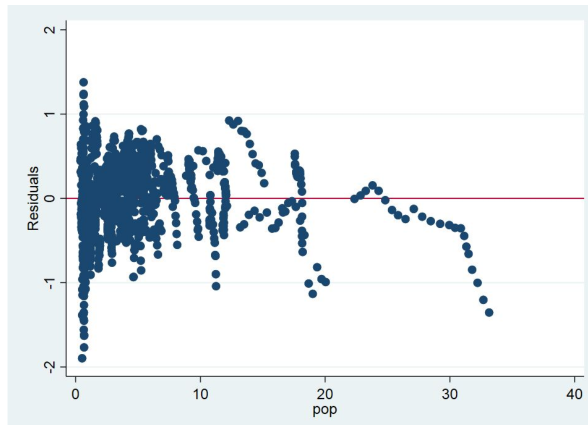


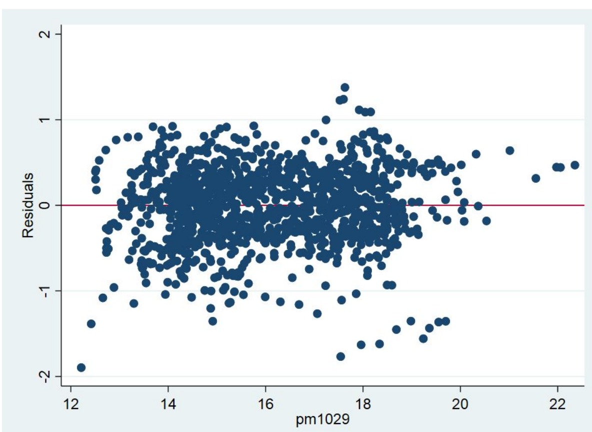


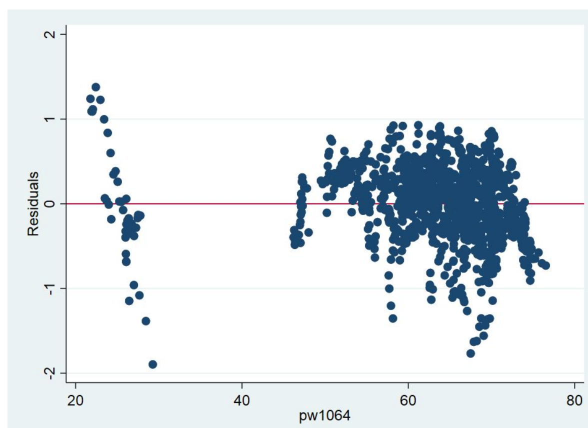

Residuals
2
1
0
-1
-2
0
500
1000
1500
2000
72-99 ONLY - Lagged Rate per 100,000 resident pop of sentenced prisoners in Stat

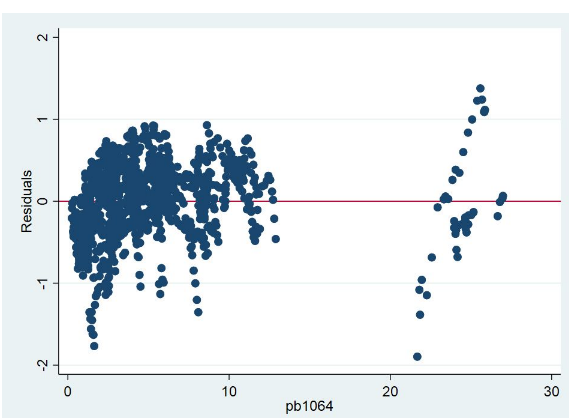

Residuals
2
1
0
-1
-2
0
5
10
density


Residuals
2
1
0
-1
-2
5
10
15
20
25
avginc


Residuals
2
1
0
-1
-2
0
10
20
30
40
pop


Residuals
2
1
0
-1
-2
12
14
16
18
20
22
pm1029


Residuals
2
1
0
-1
-2
20
40
60
80
pw1064


Residuals
2
1
0
-1
-2
0
10
20
30
pb1064

White Test:
First, we run pooled OLS without cluster robust standard errors.

```
. reg lnvio shall incarc_rate density avginc lnpop pm1029 pw1064 pb1064

      Source |       SS           df       MS      Number of obs   =     1,173
-------------+----------------------------------   F(8, 1164)      =    243.54
       Model |  305.883905         8  38.2354881   Prob > F        =    0.0000
    Residual |  182.747654     1,164  .156999702   R-squared       =    0.6260
-------------+----------------------------------   Adj R-squared   =    0.6234
       Total |  488.631558     1,172  .416921125   Root MSE        =    .39623

------------------------------------------------------------------------------
       lnvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |   -.30425    .0304518    -9.99   0.000    -.3639966   -.2445034
 incarc_rate |   .001628    .0000991    16.43   0.000     .0014336    .0018224
     density |  .0555991    .0123738     4.49   0.000     .0313217    .0798765
      avginc |  .0121265    .0070198     1.73   0.084    -.0016465    .0258994
       lnpop |  .2798641    .0123112    22.73   0.000     .2557094    .3040188
      pm1029 |  .0357077    .0099492     3.59   0.000     .0161873     .055228
      pw1064 |  .0359886     .007747     4.65   0.000      .020789    .0511882
      pb1064 |   .086114    .0153761     5.60   0.000      .055946    .1162819
       _cons |  1.945559     .508716     3.82   0.000      .947456    2.943662
------------------------------------------------------------------------------
```

Then we perform a White Test:

```
. estat imtest, white

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

         chi2(43)     =    426.29
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |     426.29     43    0.0000
            Skewness |     138.36      8    0.0000
            Kurtosis |       3.00      1    0.0831
---------------------+-----------------------------
               Total |     567.65     52    0.0000
---------------------------------------------------
```

$X^2 = N \textbf{ X } R^2$ = 426.29, p-value = 0.0000. Null hypothesis: no heteroskedasticity, against the alternative hypothesis: there is heteroskedasticity. P-value (0.000) is less than the alpha value (0.05 or 0.01), so we reject the null hypothesis and conclude that heteroskedasticity exists.

Although we know that heteroskedasticity exists, we decided to use cluster robust standard errors instead of weighted least squares (WLS) for all of our models. Even though we understand that we can develop more efficient estimators by performing WLS, we still chose to use the latter because we do not know the correct form of the variance. If we used WLS, and the assumption about the form of the variance turned out to be incorrect, the estimator would be inefficient and the standard errors would still be incorrect. With the cluster robust standard error method, our estimators are still not efficient, but the standard errors are correct so we can use them for our analysis of significance.

Using the full pooled OLS with cluster robust standard errors, we obtain the AIC and BIC shown below. For comparison purposes, these values have been included in **Section VI: Model Comparisons & Conclusions.**

```
. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,173	-1150.81	-573.9861	9	1165.972	1211.578

Note: N=Obs used in calculating BIC; see [R] BIC note.

In conclusion, in these pooled OLS models, we found that *shall* has a significant effect on reducing crime rate (by a whopping 30-39%). However, given that our data consists of observations across 51 states, we acknowledge that there is significant risk of heterogeneity between the states due to unobserved characteristics. If these unobserved characteristics are correlated with the current explanatory variables, then the pooled OLS would have an endogeneity problem. The effect of the unobserved characteristics would be hidden in the error, causing the existing explanatory variables to be correlated with the error. This results in estimators that are biased and inconsistent. In order to control for state heterogeneity and obtain unbiased and consistent estimators, we move to a FE model.

B. Entity Fixed Effects

When we include entity FE and regress on all other independent variables, excluding *year*, we obtain the results below.

```
. xtreg lnvio shall incarc_rate density avginc lnpop pm1029 pw1064 pb1064, fe cluster(stateid)

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

R-sq:                                           Obs per group:
     within  = 0.2182                                         min =         23
     between = 0.0801                                         avg =       23.0
     overall = 0.0380                                         max =         23

                                                F(8,50)           =      22.09
corr(u_i, Xb)  = -0.4698                        Prob > F          =     0.0000

                                (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
       lnvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -.0382424   .0435764    -0.88   0.384    -.1257682    .0492834
 incarc_rate |  -.0000187   .0002277    -0.08   0.935     -.000476    .0004387
     density |  -.1248095     .12807    -0.97   0.334    -.3820456    .1324266
      avginc |  -.0094166   .0129973    -0.72   0.472    -.0355223    .0166892
       lnpop |  -.1213744   .1626941    -0.75   0.459    -.4481551    .2054063
      pm1029 |  -.0524998   .0215528    -2.44   0.018    -.0957898   -.0092097
      pw1064 |   .0386326   .0131093     2.95   0.005     .0123019    .0649634
      pb1064 |   .1069708   .0310826     3.44   0.001     .0445395    .1694021
       _cons |   4.186536   .8159469     5.13   0.000     2.547658    5.825413
-------------+----------------------------------------------------------------
     sigma_u |   .7017062
     sigma_e |  .16068174
         rho |  .95017732   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

We see that the *shall* coefficient dropped significantly, as compared to the pooled OLS. When a state has shall-issue laws in place, violent crime is estimated to decrease by 3.8%, versus 30.4% in the pooled OLS. However, it is important to note that *shall* is no longer a significant predictor in the model. A more detailed comparison between the pooled OLS and entity FE model is included in **Section VI: Model Comparisons & Conclusions.**

pm1029, pw1064, and *pb1064* are significant at the 5% level. *shall, incarc_rate, density, avginc, and lnpop* are not significant even at the 10% level. If we remove these variables from the model, we obtain the results below.

```
. xtreg lnvio pm1029 pw1064 pb1064, fe cluster(stateid)

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

R-sq:                                           Obs per group:
     within  = 0.2080                                         min =         23
     between = 0.0537                                         avg =       23.0
     overall = 0.0523                                         max =         23

                                                F(3,50)           =       9.35
corr(u_i, Xb)  = 0.0819                         Prob > F          =     0.0001

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

lnvio	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
pm1029	-.042295	.0108692	-3.89	0.000	-.0641265	-.0204636
pw1064	.0337201	.0114407	2.95	0.005	.0107409	.0566994
pb1064	.0707776	.023732	2.98	0.004	.0231104	.1184448
_cons	4.207231	.7103743	5.92	0.000	2.780402	5.63406
sigma_u	.61635506					
sigma_e	.16136709					
rho	.9358531	(fraction of variance due to u_i)				

In this reduced model, we do not see much difference in the coefficients of *pm1029*, *pw1064*, and *pb1064*. This suggests that by excluding the variables, we do not have an omitted variable bias problem in the current model.

For further testing of the significance of *shall* and the other variables, we can conduct an F-test for joint significance. Results are as follows.

```
test shall incarc_rate density avginc lnpop

( 1)  shall = 0
( 2)  incarc_rate = 0
( 3)  density = 0
( 4)  avginc = 0
( 5)  lnpop = 0

      F(  5,    50) =    1.02
           Prob > F =    0.4182
```

The null hypothesis of this test is that the coefficients of *shall, incarc_rate, density, avginc,* and *lnpop* equal zero. The alternative hypothesis is that at least one of the above coefficients does not equal zero. Based on these results (p-value = 0.4182), we do not reject the null hypothesis. We conclude that the variables are *not* jointly significant to the model.

Using the full entity FE model, we obtain the AIC and BIC shown below. For comparative purposes, these values have been included in **Section VI: Model Comparisons & Conclusions.**

```
. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,173	366.1193	510.4836	8	-1004.967	-964.4286

Note: N=Obs used in calculating BIC; see [R] BIC note.

C. Time Fixed Effects

Because we suspect *year* to be a relevant variable, we ran a time FE model, a snapshot of which is shown below.

```
. reg lnvio shall incarc_rate pb1064 pw1064 pm1029 lnpop avginc density year#year, vce(cluster statei
> d)

Linear regression                               Number of obs     =      1,173
                                                F(30, 50)         =     160.20
                                                Prob > F          =     0.0000
                                                R-squared         =     0.6513
                                                Root MSE          =     .38625

                                   (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
       lnvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -.2412168   .1095221    -2.20   0.032    -.4611985   -.0212352
 incarc_rate |   .0018676   .0006825     2.74   0.009     .0004967    .0032385
      pb1064 |   .0950553   .0755011     1.26   0.214    -.0565932    .2467037
      pw1064 |   .0404511   .0365714     1.11   0.274    -.0330049     .113907
      pm1029 |   .0151246   .0654313     0.23   0.818     -.116298    .1465472
       lnpop |    .274411   .0546357     5.02   0.000      .164672    .3841501
      avginc |   .0225757   .0276327     0.82   0.418    -.0329262    .0780776
     density |   .0309668   .0564043     0.55   0.585    -.0823246    .1442582
             |
        year |
         78  |   .0258548   .0202565     1.28   0.208    -.0148316    .0665413
         79  |    .104411   .0334644     3.12   0.003     .0371958    .1716262
         80  |   .1488321   .0452328     3.29   0.002     .0579793    .2396849
         81  |   .1280385    .056752     2.26   0.028     .0140488    .2420282
         82  |   .0724924   .0732327     0.99   0.327    -.0745997    .2195845
         83  |  -.0120522   .0968413    -0.12   0.901    -.2065636    .1824593
```

Some years are individually significant at the 1% level, while others are not significant at all. In an F-test, we find that the years are jointly significant with a p-value of 0, and thus conclude that the years are jointly significant in this regression.

```
. testparm year#year

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

       F( 22,    50) =   21.23
            Prob > F =    0.0000
```

This time FE regression (which is essentially a least squares regression with time indicator variables) demonstrates that time is significant in our research question. Additionally, it shows that *shall* is significant at the 5% significance level. A state with a shall-issue law is expected to have 24% lower violent crime rate. However, expanding further upon time-only FE was disregarded since it is not able to control for state heterogeneity like the entity FE model does. As such, we next perform a time and entity FE model.

Using the above time FE model, we obtain the AIC and BIC shown below. For comparative purposes, these values have been included in **Section VI: Model Comparisons & Conclusions.**

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,173	-1150.81	-532.8638	31	1127.728	1284.815

D. Time & Entity Fixed Effects

In the first variation of this model, we regress *lnvio* on all the other independent variables, besides *rob* and *mur*, since those are already accounted for within *lnvio*.

The results indicate that all the years besides '98 and '99 are significant at the 5% level, and that none of the other independent variables are significant. Specifically, *shall* has a p-value of 0.564 indicating that it does not have a significant effect on violent crime.

```
. xtreg lnvio i.year incarc_rate pb1064 pw1064 pm1029 lnpop avginc density shall, fe
> cluster(stateid)

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

R-sq:                                           Obs per group:
     within  = 0.4200                                         min =         23
     between = 0.1963                                         avg =       23.0
     overall = 0.0685                                         max =         23

                                                F(30,50)          =      54.66
corr(u_i, Xb)  = -0.5307                        Prob > F          =     0.0000

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

lnvio	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
year						
78	.0610252	.0162737	3.75	0.000	.0283385	.093712
79	.1684788	.0252812	6.66	0.000	.1177001	.2192576
80	.2233543	.0355171	6.29	0.000	.1520162	.2946925
81	.2252249	.0397057	5.67	0.000	.1454737	.3049762
82	.2038194	.04721	4.32	0.000	.1089953	.2986435
83	.1690077	.0595131	2.84	0.007	.049472	.2885433
84	.2046856	.0758315	2.70	0.009	.0523735	.3569976
85	.2570377	.0909181	2.83	0.007	.0744233	.4396521
86	.3371815	.1070551	3.15	0.003	.122155	.5522081
87	.3381137	.1225613	2.76	0.008	.0919421	.5842853
88	.4015654	.1366376	2.94	0.005	.1271208	.67601
89	.458197	.1497486	3.06	0.004	.157418	.758976
90	.5654111	.1920148	2.94	0.005	.1797381	.9510841
91	.6195228	.199944	3.10	0.003	.2179234	1.021122
92	.6529918	.2128637	3.07	0.003	.2254425	1.080541
93	.677034	.2210682	3.06	0.004	.2330055	1.121062
94	.6647518	.2302025	2.89	0.006	.2023766	1.127127
95	.6587298	.2400649	2.74	0.008	.1765453	1.140914
96	.6039256	.2513635	2.40	0.020	.099047	1.108804
97	.5843769	.2598807	2.25	0.029	.0623911	1.106363
98	.5296452	.272907	1.94	0.058	-.0185046	1.077795
99	.4712289	.2841429	1.66	0.103	-.0994889	1.041947
incarc_rate	.0000686	.0002022	0.34	0.736	-.0003375	.0004747
pb1064	.021379	.0507923	0.42	0.676	-.0806403	.1233984
pw1064	.0063205	.0235785	0.27	0.790	-.0410382	.0536792
pm1029	.0740327	.0501517	1.48	0.146	-.0267	.1747654
lnpop	-.1479153	.1811246	-0.82	0.418	-.5117147	.2158841
avginc	.0001695	.0164174	0.01	0.992	-.0328058	.0331447
density	-.0622844	.1217831	-0.51	0.611	-.3068929	.1823242
shall	-.0233172	.0401109	-0.58	0.564	-.1038824	.0572479
_cons	4.099443	1.241867	3.30	0.002	1.60508	6.593807
sigma_u	.72465232					
sigma_e	.13978463					
rho	.96412493	(fraction of variance due to u_i)				

Using the above model, we obtain the AIC and BIC shown below. For comparison purposes, these values have been included in **Section VI: Model Comparisons & Conclusions.**

```
. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,173	366.1193	685.6076	30	-1311.215	-1159.196

Note: N=Obs used in calculating BIC; see [R] BIC note.

We then ran a second iteration by dropping *lnpop*, thinking that it might be somewhat captured by *density*. We also dropped *pw1064*, since it might be considered a collinear variable with *pb1064* - the total of these variables cannot exceed 1 (though, the total is likely not near 1).

We see in this model that now even the years ‘98 and ‘99 are significant. In addition, now the *pm1029* variable is significant and also has a positive coefficient. This aligns with our expectations - as the percentage of young males increases, the violent crime rate also increases.

```
. xtreg lnvio i.year incarc_rate pb1064 pm1029 avginc density shall, fe cluster(stat
> eid)

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

R-sq:                                           Obs per group:
     within  = 0.4167                                         min =         23
     between = 0.0858                                         avg =       23.0
     overall = 0.0157                                         max =         23

                                                F(28,50)          =      58.77
corr(u_i, Xb)  = -0.4042                        Prob > F          =     0.0000
```

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

lnvio	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
year						
78	.0618034	.0119651	5.17	0.000	.0377708	.0858361
79	.1694205	.0175883	9.63	0.000	.1340934	.2047477
80	.2238535	.0264812	8.45	0.000	.1706645	.2770425
81	.2265677	.0304326	7.44	0.000	.165442	.2876934
82	.2065383	.0356288	5.80	0.000	.1349758	.2781008
83	.1733954	.0444172	3.90	0.000	.0841808	.2626101
84	.2113716	.0570631	3.70	0.001	.096757	.3259862
85	.2658494	.0686255	3.87	0.000	.1280111	.4036877
86	.3484153	.0839946	4.15	0.000	.1797072	.5171235
87	.3519821	.096904	3.63	0.001	.1573447	.5466195
88	.418224	.1073795	3.89	0.000	.2025458	.6339021
89	.4771882	.118806	4.02	0.000	.2385594	.715817
90	.5987609	.1285733	4.66	0.000	.3405139	.8570079
91	.6546192	.1359641	4.81	0.000	.3815273	.927711
92	.6896927	.1453154	4.75	0.000	.3978181	.9815674
93	.7146689	.1509334	4.73	0.000	.4115104	1.017828
94	.7038881	.1581106	4.45	0.000	.3863137	1.021463
95	.699747	.1613415	4.34	0.000	.375683	1.023811
96	.6465539	.1671505	3.87	0.000	.3108223	.9822855
97	.6283703	.1734131	3.62	0.001	.2800599	.9766807
98	.5752984	.1806568	3.18	0.002	.2124386	.9381582
99	.518767	.1873149	2.77	0.008	.1425339	.8950001
incarc_rate	.0000256	.0002276	0.11	0.911	-.0004314	.0004827
pb1064	.0050607	.0331714	0.15	0.879	-.061566	.0716875
pm1029	.0855713	.032301	2.65	0.011	.0206929	.1504497
avginc	-.0006475	.0153054	-0.04	0.966	-.0313893	.0300944
density	-.1092415	.1272317	-0.86	0.395	-.3647938	.1463109
shall	-.0266708	.0414689	-0.64	0.523	-.1099634	.0566219
_cons	4.256356	.6014034	7.08	0.000	3.048402	5.46431
sigma_u	.69108033					
sigma_e	.14005341					
rho	.96054971	(fraction of variance due to u_i)				

We believe this time and entity FE model to be the most appropriate model for this data set. An entity-fixed component is necessary for this model, since we want to account for the individual-specific, time-invariant characteristics for each state. This also means that the coefficient estimates in the resulting model "depend only on the variation of the dependent and explanatory variables within entities" (Hill, Griffiths, & Lim, 2011), rather than between entities.

We also wanted to include a time-fixed component (where time dummies are added) because time could be an important factor in determining the violence rates. There could be a particular national event, movement, or sentiment that impacts violent crime rates for all states.

Including both time and entity FE also allows us to control for omitted variable bias for 1) variables that vary over time, but not across entities, and 2) variables that vary across entities, but not over time. The time and entity FE model helps account for the issue of endogeneity due to omitted variable bias.

Since we felt the time and entity FE model was most apt for this data set, we decided to explore further models with this type. We regressed *lnmur* (natural log of murder) and *lnrob* (natural log of robbery) on the indicator variables. We did so because we were interested in what effects, if any, the independent variables have on fatal (murders) vs. non-fatal (robberies) types of violence.

When *lnmur* is the variable of interest, the only years that were statistically significant were '80 and '84, and *avginc* was the only other significant independent variable (all at 5% significance levels). This led us to speculate that maybe a particular event occurred in the year 1980 or 1984 that caused murder rates to become significant. However, even though *avginc* is significant, it has a small magnitude (0.05), and is positive, which is inconsistent with our expectations. When we look only at murder crime rates, *shall* is still not a significant predictor.

```
. xtreg lnmur i.year incarc_rate pb1064 pw1064 pm1029 lnpop avginc density shall,
>  fe cluster(stateid)

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

R-sq:                                           Obs per group:
     within  = 0.2870                                         min =         23
     between = 0.2504                                         avg =       23.0
     overall = 0.1861                                         max =         23

                                                F(30,50)          =      87.74
corr(u_i, Xb)  = -0.8506                        Prob > F          =     0.0000

                              (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
       lnmur |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |
         78  |   -.002145   .0320298    -0.07   0.947    -.0664788    .0621889
         79  |   .0576725   .0304028     1.90   0.064    -.0033933    .1187383
         80  |   .0890074   .0400011     2.23   0.031     .0086629    .1693519
         81  |   .0994839   .0503528     1.98   0.054    -.0016527    .2006205
         82  |    .018787   .0550126     0.34   0.734    -.0917091     .129283
         83  |  -.0364706   .0608085    -0.60   0.551    -.1586081    .0856668
         84  |  -.1439343   .0665481    -2.16   0.035       -.2776   -.0102685
         85  |  -.0972829   .0792617    -1.23   0.225    -.2564847    .0619189
         86  |  -.0257576   .0854281    -0.30   0.764     -.197345    .1458298
         87  |  -.0452569   .0919265    -0.49   0.625    -.2298967    .1393828
         88  |  -.0364231   .1112468    -0.33   0.745    -.2598688    .1870227
         89  |  -.0356584     .12044    -0.30   0.768    -.2775692    .2062524
         90  |    .028451   .1507706     0.19   0.851    -.2743807    .3312827
         91  |   .0723395   .1617765     0.45   0.657    -.2525981    .3972772
         92  |    .033796   .1682934     0.20   0.842    -.3042313    .3718233
         93  |   .1193491   .1749053     0.68   0.498    -.2319585    .4706567
         94  |   .0082818   .1808722     0.05   0.964    -.3550106    .3715742
         95  |   .0192433   .1835702     0.10   0.917    -.3494682    .3879549
         96  |   -.052902    .197402    -0.27   0.790    -.4493955    .3435916
         97  |   -.160131   .2007901    -0.80   0.429    -.5634298    .2431678
         98  |   -.224861   .2163674    -1.04   0.304    -.6594478    .2097258
         99  |  -.2945008   .2246274    -1.31   0.196    -.7456781    .1566766
             |
 incarc_rate |  -.0001702   .0003726    -0.46   0.650    -.0009185    .0005781
      pb1064 |   .0215073   .0772697     0.28   0.782    -.1336935    .1767081
      pw1064 |   .0044116   .0191851     0.23   0.819    -.0341229    .0429461
      pm1029 |   .0566797   .0394504     1.44   0.157    -.0225587    .1359182
       lnpop |  -.1688479   .1673388    -1.01   0.318    -.5049579     .167262
      avginc |   .0581811   .0165401     3.52   0.001     .0249593    .0914029
     density |  -.5332878   .3345633    -1.59   0.117    -1.205278    .1387025
       shall |  -.0120109   .0388632    -0.31   0.759    -.0900698    .0660481
       _cons |   .1178213   1.074993     0.11   0.913    -2.041366    2.277009
-------------+----------------------------------------------------------------
     sigma_u |  1.1587646
     sigma_e |   .2033168
         rho |  .97013328   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

With *lnrob* as the variable of interest, the years ‘79 - ‘82 are significant, with no other independent variables of significance. This is consistent with expectations, since we know there was a recession in the early 1980’s, which could have led to more robberies nationwide. *shall,* again, is not a significant predictor of robbery crime rate.

```
. xtreg lnrob i.year incarc_rate pb1064 pw1064 pm1029 lnpop avginc density shall,
> fe cluster(stateid)

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

R-sq:                                           Obs per group:
     within  = 0.2377                                         min =         23
     between = 0.3568                                         avg =       23.0
     overall = 0.3383                                         max =         23

                                                F(30,50)          =      42.61
corr(u_i, Xb)  = 0.3383                         Prob > F          =     0.0000

                              (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
       lnrob |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |
         78  |   .0292667   .0210036     1.39   0.170    -.0129203    .0714536
         79  |   .1312937   .0322186     4.08   0.000     .0665807    .1960066
         80  |    .234877   .0457162     5.14   0.000     .1430532    .3267007
         81  |    .262649   .0492086     5.34   0.000     .1638107    .3614873
         82  |   .2031351    .062168     3.27   0.002      .078267    .3280032
         83  |   .1061304   .0840325     1.26   0.212    -.0626539    .2749147
         84  |   .0617979   .1027058     0.60   0.550    -.1444928    .2680886
         85  |   .0944552   .1236982     0.76   0.449    -.1539999    .3429103
         86  |    .169623   .1483687     1.14   0.258    -.1283843    .4676303
         87  |   .1358651   .1651514     0.82   0.415    -.1958512    .4675814
         88  |   .1694519   .1837637     0.92   0.361    -.1996484    .5385522
         89  |   .2234748   .2087315     1.07   0.289    -.1957747    .6427242
         90  |   .3152448   .2587538     1.22   0.229    -.2044775    .8349671
         91  |   .4292352   .2711989     1.58   0.120    -.1154837    .9739541
         92  |   .4229334   .2869947     1.47   0.147    -.1535125    .9993793
         93  |   .4367543   .2992575     1.46   0.151    -.1643221    1.037831
         94  |   .4485569   .3145637     1.43   0.160    -.1832628    1.080377
         95  |   .4458481   .3248582     1.37   0.176    -.2066489    1.098345
         96  |   .3838101    .341616     1.12   0.267    -.3023459    1.069966
         97  |     .31252    .349837     0.89   0.376    -.3901483    1.015188
         98  |   .2121152   .3606396     0.59   0.559    -.5122508    .9364812
         99  |   .1311218   .3758586     0.35   0.729    -.6238123     .886056
             |
 incarc_rate |   .0000306   .0003344     0.09   0.927    -.0006412    .0007024
      pb1064 |   .0243376   .0826121     0.29   0.770    -.1415937    .1902689
      pw1064 |  -.0080006   .0328335    -0.24   0.808    -.0739487    .0579475
      pm1029 |   .1012235     .07095     1.43   0.160    -.0412837    .2437307
       lnpop |   .1625726   .2577405     0.63   0.531    -.3551144    .6802596
      avginc |   .0156998    .024058     0.65   0.517    -.0326221    .0640217
     density |  -.0813018   .2025748    -0.40   0.690    -.4881852    .3255816
       shall |   .0212216   .0514885     0.41   0.682     -.082196    .1246393
       _cons |   2.823837   1.840882     1.53   0.131     -.873684    6.521357
-------------+----------------------------------------------------------------
     sigma_u |  .80861137
     sigma_e |  .19329572
         rho |  .94594561   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

VI. Model Comparisons & Conclusion

In the table below we compare the estimates for pooled OLS and entity FE. It is apparent that the two sets of coefficient estimates are very different from each other. As noted in **Section V: Data Modeling, Subsection B: Entity Fixed Effects**, the magnitude of *shall*'s effect on violent crime rate decreases tremendously in the entity FE model. We see a 30.4% decrease in violent crime for having shall-issue law in the pooled OLS, as compared to a 3.8% decrease in the FE model. More importantly, *shall* is no longer a significant predictor in the entity FE model. Additionally, the signs for all the variables besides *shall (incarc_rate*, *density*, *avginc*, *lnpop, pm1029, pw1064,* and *pb1064)* are reversed. These differences suggest that endogeneity is present in the pooled OLS, and that ignoring state effects can lead to misleading conclusions.

	Estimates		Robust Standard Errors	
Variable	Pooled OLS	Entity Fixed Effects	Pooled OLS	Entity Fixed Effects
shall	-0.304250	-0.038242	0.096395	0.043576
incarc_rate	0.001628	-0.000019	0.000532	0.000228
density	0.055599	-0.124810	0.041397	0.128070
avginc	0.012127	-0.009417	0.023943	0.012997
lnpop	0.279864	-0.121374	0.055521	0.162694
pm1029	0.035708	-0.052500	0.032406	0.021554
pw1064	0.035989	-0.038633	0.032600	0.013109
pb1064	0.086114	-0.106971	0.067131	0.031083
_cons	1.945559	4.186536	2.171736	0.815947

Throughout this analysis, we moved from model to model based on theory and critical analysis of the dataset itself. We started with the pooled OLS model, which upheld our hypothesis that shall-issue laws reduce violent crime rates; however, we concluded that the model was not sufficiently dynamic to capture the individual heterogeneity of the states, laws, and circumstances. The pooled OLS model was found to have an AIC of 1,166.

We then moved to the entity FE model in order to account for those individual-specific, time-invariant factors so that the coefficient estimates in the resulting model "depend only on the variation of the dependent and explanatory variables within entities" (Hill, Griffiths, & Lim, 2011), rather than between entities. This approach seemed to provide a more reliable result based on the AIC that dropped to -1,005. In this model, *shall* becomes insignificant, which directs us to reject our hypothesis. However, we suspect there are time-related trends that should be accounted for, such as national trends or national political or economic events that influence crime rates.

To check for these time effects, we ran a time FE model. Although the AIC increased to a level similar to the pooled OLS regression, several of the *year* dummy variables became significant. In an F test, they were found to be jointly significant and deemed relevant to the study.

To combine both the state and time variations, we ran a time and entity FE model. This model had the lowest AIC and BIC values of all the previously run models, with an AIC of -1,311, indicating it as our best model and the one that best captures the true relationship between the explanatory variables and violent crime. In addition, this model is also unbiased (does not suffer from omitted variables, simultaneity, or measurement error) and consistent (converging to a true parameter), although not the best (most efficient). Using time and entity FE, the *shall* variable is not significantly different from 0, and therefore does not support our hypothesis. We conclude that there is no evidence to support the assertion that shall-issue laws reduce crime, whether murders, robberies, or overall violent crimes. Neither is there evidence to support the opposing assertion that shall-issue laws increase crime.

One possible further extension of this analysis is including an Instrumental Variables (IV) for any endogenous regressors. We know from comparing the pooled OLS and entity FE models that endogeneity is likely present in the pooled OLS model, since the pooled OLS model greatly overestimated some of the independent variable coefficients. It is possible there is endogeneity due to simultaneous causality bias between *incarc_rate* (independent variable) and *lnvio* (dependent variable). When the incarceration rate goes up, criminals are deterred from committing violent crimes. When the violent crime rate goes up, that leads to a higher incarceration rate. An IV for *incarc_rate* would have to meet three conditions: 1) have a high correlation with *incarc_rate*, 2) not have a direct impact on *lnvio*, and 3) not be correlated with the error term.

One possible IV for *incarc_rate* is drug usage. We know from general knowledge that many incarcerated individuals are drug users, and vice versa. However, drug usage can also lead to increased rates of violent crime, which could potentially violate the second condition outlined above. We would need data on drug usage from a reputable source in order to determine whether it is a suitable IV.

Below are tables that organize all of the above model results for easy visual comparison.

	Model Type	Variables Included	AIC	BIC
1	Pooled OLS	ln(vio) ~ shall + incarc_rate + density + avginc + ln(pop) + pm1029 + pw1064 + pb1064	1165.972	1211.578
2	Entity-Fixed	ln(vio) ~ shall + incarc_rate + density + avginc + ln(pop) + pm1029 + pw1064 + pb1064	-1004.967	-964.4286
3	Time-Fixed	ln(vio) ~ shall + incarc_rate + pb1064 + pw1064 + pm1029 + ln(pop) + avginc + density + year#year	1127.728	1284.815
4	Time & Entity Fixed	ln(vio) ~ shall + incarc_rate + density + avginc + ln(pop) + pm1029 + pw1064 + pb1064 + i.year	-1311.215	-1159.196

VII. References

Ayres, I., & Donohue, J. J. (2003). Shooting Down the More Guns, Less Crime Hypothesis. Retrieved December 11, 2019, from https://digitalcommons.law.yale.edu/fss_papers/1241/.

Hill, R. C., Griffiths, W. E., & Lim, G. C. (2011). *Principles of Econometrics* (4th ed.). Hoboken, NJ: Wiley.