

# **Socio-economic determinants of crime in U.S. states**

Paula de Miguel Sánchez & Esther Allevi

## **Table of contents**

<b>1 Research question</b>	<b>1</b>
<b>2 Brief dataset description</b>	<b>3</b>
<b>3 Data analysis</b>	<b>3</b>
3.1 Part 1. Long-run analysis (1960–2018) . . . . .	3
3.1.1 Descriptive analysis . . . . .	3
3.1.2 Regression analysis . . . . .	4
3.1.3 Tests . . . . .	10
3.2 Part 2. Short-run analysis (2015–2018) - Contemporary correlates of crime . . . . .	10
3.2.1 Descriptive analysis . . . . .	10
3.2.2 Regression analysis . . . . .	10
3.2.3 Tests . . . . .	10
<b>4 Conclusion</b>	<b>10</b>

## **1 Research question**

Our project studies the socio-economic and demographic determinants of crime in the United States at the state level. More precisely, we ask whether economic opportunities and social inequalities are correlated with different types of crime (violent and property crime) across states and over time.

Crime in the United States exhibits substantial variation across states and over time, and understanding the socio-economic factors that contribute to these differences remains a central question in both economics and public policy. Our project aims to investigate how economic conditions, demographic structure, inequality, public spending, and police behaviour correlate

with violent and property crime at the state–year level. The objective is not to establish strict causal relationships—an ambitious task given the observational nature of the data—but rather to document robust associations and explore whether crime levels systematically co-evolve with structural socio-economic indicators.

A first motivation for this research question is the long-standing hypothesis that economic opportunity and crime are linked. Periods of slow economic growth and labour market stress may increase incentives for certain types of offences, particularly property crime. To examine this mechanism, we combine the crime data from the CORGIS/FBI series with annual state GDP from the Bureau of Economic Analysis. GDP provides a measure of overall economic activity and allows us to test whether states experiencing strong or weak economic performance also display different crime patterns.

Income inequality is another potential determinant of criminal activity. The World Inequality Database (WID) provides rich information on income shares and fiscal aggregates at the top and bottom of the distribution. High inequality may generate both economic stress and social fragmentation, potentially amplifying incentives for crime or reducing the perceived legitimacy of institutions. By merging the WID data with the crime series, we can evaluate whether states with higher income concentration or lower fiscal income shares among the majority of the population exhibit systematically different violent or property crime rates.

Migration flows constitute a third dimension of interest. Public debate often assumes—sometimes incorrectly—that immigration influences crime rates. The DHS immigration data enable us to incorporate yearly state-level measures of lawful permanent residents, nonimmigrants, asylees, and refugees. Our goal is not to test politically charged claims, but to empirically examine whether changes in migration patterns are associated with changes in state crime trends once we control for other socio-economic conditions.

In addition, we incorporate information on public expenditure and food assistance (from the Census and USDA Food Environment Atlas), which may proxy for the strength of social safety nets and local investment in welfare-related programmes. States with stronger social services may provide better support for vulnerable populations, potentially mitigating some of the socio-economic pressures that contribute to crime. Including these variables allows us to explore such channels empirically.

Finally, we integrate data on police shootings from the Washington Post database. While this dataset does not measure police activity directly, fatal shootings can serve as a proxy for law-enforcement intensity or tension between police and residents. Investigating whether states with higher levels of fatal shootings also differ in crime trends contributes to a broader understanding of institutional and behavioural dimensions of crime.

How do economic conditions, income inequality, immigration flows, public spending, and police activity correlate with violent and property crime across U.S. states?

## **2 Brief dataset description**

We build our panel from seven sources:

- crime rates (FBI/CORGIS, 1960–2019) distinguishing violent and property offences;
- income inequality (WID, 1960–2018) providing top income shares by state;
- GDP (BEA Regional Accounts, 1997–2023);
- immigration flows (DHS, 2013–2023) covering legal residents, refugees and asylees;
- police shootings (Washington Post, 2015–2024);
- public spending (Census Bureau, 2017–2023) on education, welfare, health, corrections and police;
- and food assistance participation (USDA Food Atlas, 2015).

Full descriptions are provided in the data cleaning document (Phase 2).

## **3 Data analysis**

Our empirical analysis is divided into two parts, reflecting the heterogeneous temporal coverage of our sources. The crime, inequality (WID), and GDP series provide consistent state-year observations spanning nearly six decades (1960–2018), allowing us to examine long-run associations and temporal dynamics. In contrast, immigration flows (2013–2018), police shootings (2015–2018), public spending (2017–2018) and food access (2012–2013) are only available for recent years.

Rather than discarding these valuable contemporary indicators, we adopt a two-stage approach: we first exploit the full historical depth of our panel to study structural relationships between economic conditions and crime, then narrow the window to 2015–2018 to incorporate additional explanatory factors in a richer, shorter-term analysis.

### **3.1 Part 1. Long-run analysis (1960–2018)**

#### **3.1.1 Descriptive analysis**

Evolution of violent crime and property crime in the US per states.

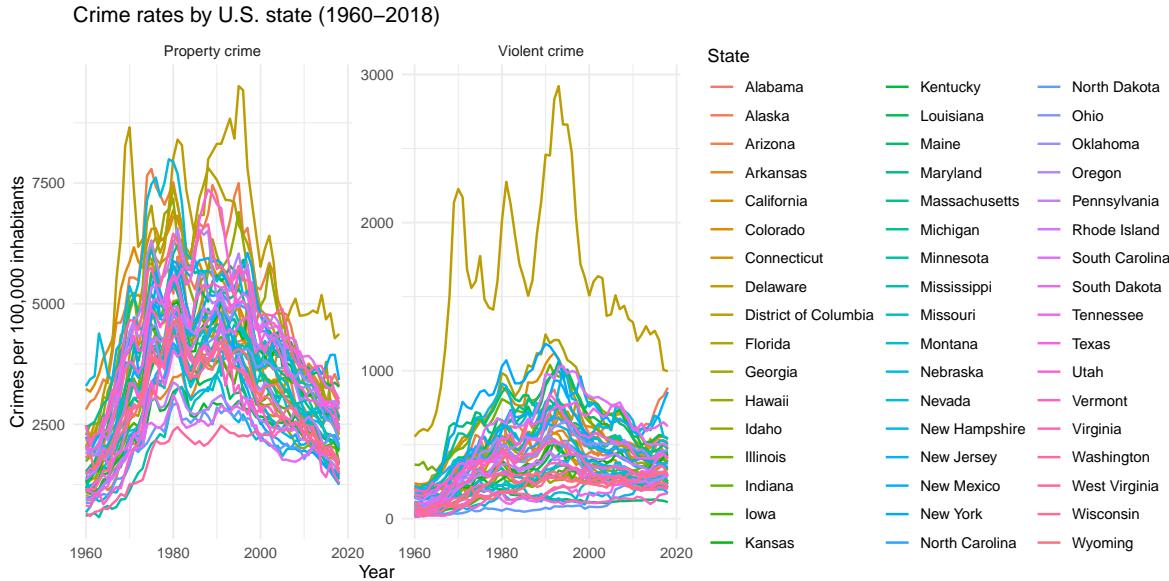


Figure 1: Violent and property crime rates by U.S. state

**ATTENTION :** The colors make it very hard to read, maybe it would be better to have one graph with no colors and the mean in red ?

### 3.1.2 Regression analysis

We first examine the relationship between GDP and violent / property crime rates. Economic theory suggests that GDP should be a stronger predictor of property crime (primarily motivated by financial need) than of violent crime, which is driven by more complex social and psychological factors (Becker, 1968).

**ATTENTION :** remember to delete all of these summaries, it's just so that we can see the results quicker for now !

Oneway (individual) effect Within Model

Call:

```
plm(formula = Data.Rates.Property.All ~ log_gdp, data = panel_reg,
     model = "within", index = c("State", "Year"))
```

Balanced Panel: n = 51, T = 21, N = 1071

Residuals:

	Min.	1st Qu.	Median	3rd Qu.	Max.
-1007.66664	-187.14616	0.71577	189.81172	1226.69115	

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )							
log_gdp	-1875.29	38.69	-48.47	< 2.2e-16 ***							
---											
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'..'	0.1	' '	1

Total Sum of Squares: 332760000

Residual Sum of Squares: 100670000

R-Squared: 0.69748

Adj. R-Squared: 0.68233

F-statistic: 2349.32 on 1 and 1019 DF, p-value: < 2.22e-16

Oneway (individual) effect Within Model

Call:

```
plm(formula = Data.Rates.Violent.All ~ log_gdp, data = panel_reg,
     model = "within", index = c("State", "Year"))
```

Balanced Panel: n = 51, T = 21, N = 1071

Residuals:

	Min.	1st Qu.	Median	3rd Qu.	Max.
-338.3200	-29.2469	-3.0543	28.3988	281.9309	

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )							
log_gdp	-108.2646	7.9294	-13.654	< 2.2e-16 ***							
---											
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'..'	0.1	' '	1

Total Sum of Squares: 5001900

Residual Sum of Squares: 4228400

R-Squared: 0.15465

Adj. R-Squared: 0.11234

F-statistic: 186.421 on 1 and 1019 DF, p-value: < 2.22e-16

The panel regression results reveal a contrast between the two types of crime, that we had planned.

For property crime, GDP explains approximately 69% of the within-state variation ( $R^2 = 0.69$ ), which is consistent with economic theories on crime suggesting that property offences are motivated by financial need (Becker, 1968).

As for violent crime, GDP alone fails to significantly predict violent crime rates ( $R^2 = 0.15$ ), indicating that violent offences are driven by factors beyond economic conditions.

We then add income inequality, using the share of pre-tax income accruing to the top 10% of the distribution, as provided by the World Inequality Database. A higher share indicates greater income concentration at the top, which could be associated with increased social tension and crime rates (Fajnzylber, Lederman, and Loayza, 2002).

Oneway (individual) effect Within Model

Call:

```
plm(formula = Data.Rates.Property.All ~ log_gdp + share_pretax_income_p90p100,
     data = panel_reg, model = "within", index = c("State", "Year"))
```

Balanced Panel: n = 51, T = 21, N = 1071

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-1008.7288	-188.7218	1.6239	188.0136	1223.0143

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )
log_gdp	-1824.561	45.923	-39.7307	< 2e-16 ***
share_pretax_income_p90p100	-874.021	427.873	-2.0427	0.04134 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 332760000

Residual Sum of Squares: 100260000

R-Squared: 0.69871

Adj. R-Squared: 0.68332

F-statistic: 1180.41 on 2 and 1018 DF, p-value: < 2.22e-16

Oneway (individual) effect Within Model

Call:

```
plm(formula = Data.Rates.Violent.All ~ log_gdp + share_pretax_income_p90p100,
     data = panel_reg, model = "within", index = c("State", "Year"))
```

Balanced Panel: n = 51, T = 21, N = 1071

Residuals:

	Min.	1st Qu.	Median	3rd Qu.	Max.
	-338.0478	-29.5385	-3.5873	28.3142	282.6040

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )
log_gdp	-99.8570	9.4185	-10.6022	< 2e-16 ***
share_pretax_income_p90p100	-144.8619	87.7536	-1.6508	0.09909 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 5001900

Residual Sum of Squares: 4217100

R-Squared: 0.15691

Adj. R-Squared: 0.11384

F-statistic: 94.7308 on 2 and 1018 DF, p-value: < 2.22e-16

Finally, we add year fixed effects to control for national trends (e.g., federal policies, economic cycles) that affect all states simultaneously.

Twoways effects Within Model

Call:

```
plm(formula = Data.Rates.Property.All ~ log_gdp + share_pretax_income_p90p100,
     data = panel_reg, effect = "twoways", model = "within", index = c("State",
     "Year"))
```

Balanced Panel: n = 51, T = 21, N = 1071

Residuals:

	Min.	1st Qu.	Median	3rd Qu.	Max.
	-967.51136	-160.83013	-0.55724	148.44822	1279.70476

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )
log_gdp	-214.08	136.20	-1.5718	0.1163
share_pretax_income_p90p100	-676.11	428.76	-1.5769	0.1151

Total Sum of Squares: 77980000

Residual Sum of Squares: 77579000

```
R-Squared:      0.0051377
Adj. R-Squared: -0.066636
F-statistic: 2.57695 on 2 and 998 DF, p-value: 0.076511
```

Twoways effects Within Model

Call:

```
plm(formula = Data.Rates.Violent.All ~ log_gdp + share_pretax_income_p90p100,
     data = panel_reg, effect = "twoways", model = "within", index = c("State",
     "Year"))
```

Balanced Panel: n = 51, T = 21, N = 1071

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-360.729552	-30.286396	-0.092583	25.294463	269.836841

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )
log_gdp	119.753	29.576	4.0490	5.541e-05 ***
share_pretax_income_p90p100	-224.361	93.103	-2.4098	0.01614 *
---				
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'
	0.1	' '	1	

Total Sum of Squares: 3736800  
Residual Sum of Squares: 3658000  
R-Squared: 0.021077  
Adj. R-Squared: -0.049547  
F-statistic: 10.7439 on 2 and 998 DF, p-value: 2.4182e-05

Model 3 shows that the GDP-inequalities-crime relationship is mainly driven by the common temporal trend (all states are becoming wealthier and are seeing a decline in crime at the same time), rather than by dynamics specific to each state.

Table 1: Property crime: Panel regression results

	Model 1	Model 2	Model 3
log_gdp	-1875.288*** (38.690)	-1824.561*** (45.923)	-214.076 (136.202)
share_pretax_income_p90p100		-874.021* (427.873)	-676.111 (428.758)
Num.Obs.	1071	1071	1071
R2	0.697	0.699	0.005
R2 Adj.	0.682	0.683	-0.067

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

For property crime (Table 1), GDP shows a strong negative association ( $-1875$ ,  $p < 0.001$ ), explaining about 70% of within-state variation (Model 1). Adding inequality improves the fit marginally, with higher income concentration associated with lower property crime ( $-874$ ,  $p < 0.05$ ).

As a robustness check, Model 3 adds year fixed effects. The  $R^2$  drops substantially (0.005), indicating that the GDP-crime relationship primarily reflects shared national trends rather than state-specific dynamics. This is consistent with the well-documented “great crime decline” - also visible in our first graph - that affected all U.S. states from the 1990s onward.

Table 2: Violent crime: Panel regression results

	Model 1	Model 2	Model 3
log_gdp	-108.265*** (7.929)	-99.857*** (9.419)	119.753*** (29.576)
share_pretax_income_p90p100		-144.862+ (87.754)	-224.361* (93.103)
Num.Obs.	1071	1071	1071
R2	0.155	0.157	0.021
R2 Adj.	0.112	0.114	-0.050

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

For violent crime (Table 2), the results confirm our theoretical expectations: GDP has a weaker association ( $-100$ ,  $R^2 = 0.15$ ), and inequality is only marginally significant ( $p < 0.10$ ). Economic factors alone cannot explain violent crime patterns.

### **3.1.3 Tests**

## **3.2 Part 2. Short-run analysis (2015–2018) - Contemporary correlates of crime**

Summary statistics :

Data.Rates.Violent_Rates	Data.Rates.Violent_Rates_sd	Data.Rates.Violent_Rates_mean	Data.Rates.Violent_Rates_sd	Data.Rates.Violent_Rates_sd	Data.Rates.Violent_Rates_sd	Data.Rates.Violent_Rates_sd	Data.Rates.Violent_Rates_sd
386.299	180.5363	381460.877499.7	19.55224	24.77274	21996.17	40175.74	

### **3.2.1 Descriptive analysis**

### **3.2.2 Regression analysis**

### **3.2.3 Tests**

## **4 Conclusion**