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

Paula de Miguel Sánchez & Esther Allevi

Table of contents

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 (2012-2023) 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 Long-run analysis (1960–2018)

3.1.1 Descriptive analysis

3.1.1.1 National long-run trends in crime (1960-2018)

To establish a benchmark for the subsequent analysis, we first document the long-run evolution of violent and property crime across U.S. states. Figure 1 displays the average crime rate across states together with the 10th–90th percentile range, providing a synthetic view of both aggregate dynamics and cross-state heterogeneity.

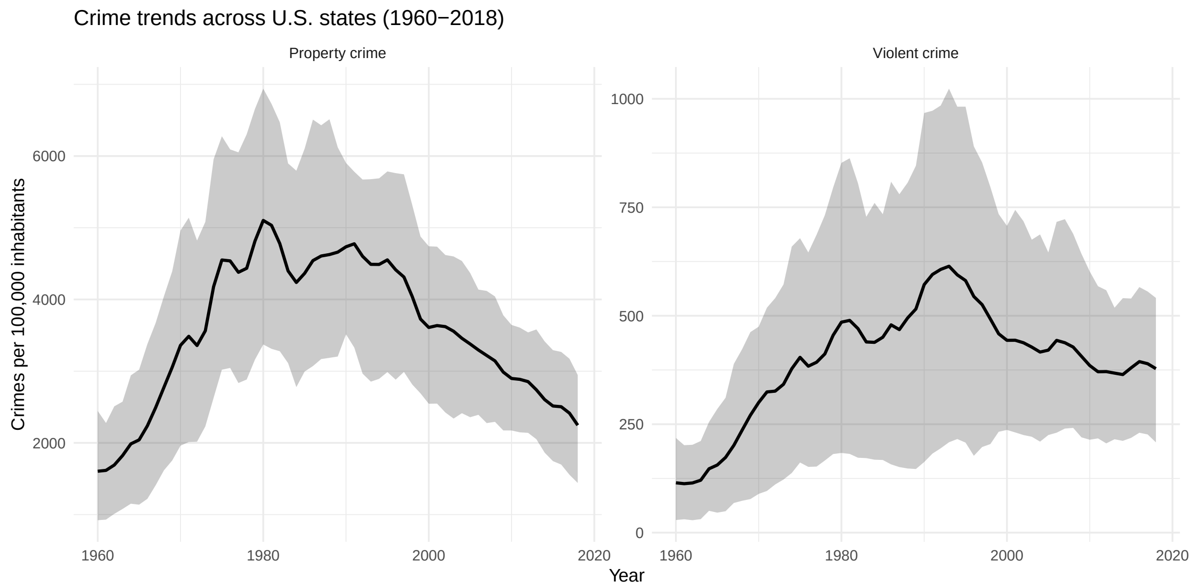

Figure 1: National crime trends and cross-state dispersion (mean and 10th–90th percentiles, 1960–2018)

Both crime categories exhibit a pronounced common time profile, with a strong increase from the 1960s to the late 1980s or early 1990s, followed by a sustained decline.

Despite these shared national trends, cross-state heterogeneity remains large throughout the entire period, as illustrated by the wide percentile bands. Evolution of violent crime and property crime in the US per states.

Property crime displays a particularly strong rise and fall, whereas violent crime peaks later and declines more gradually, suggesting distinct underlying mechanisms. The persistence of substantial dispersion across states indicates that national trends alone cannot account for observed crime dynamics and highlights the relevance of state-specific factors. These descriptive findings motivate the use of panel models with both state and time fixed effects in the subsequent analysis.

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

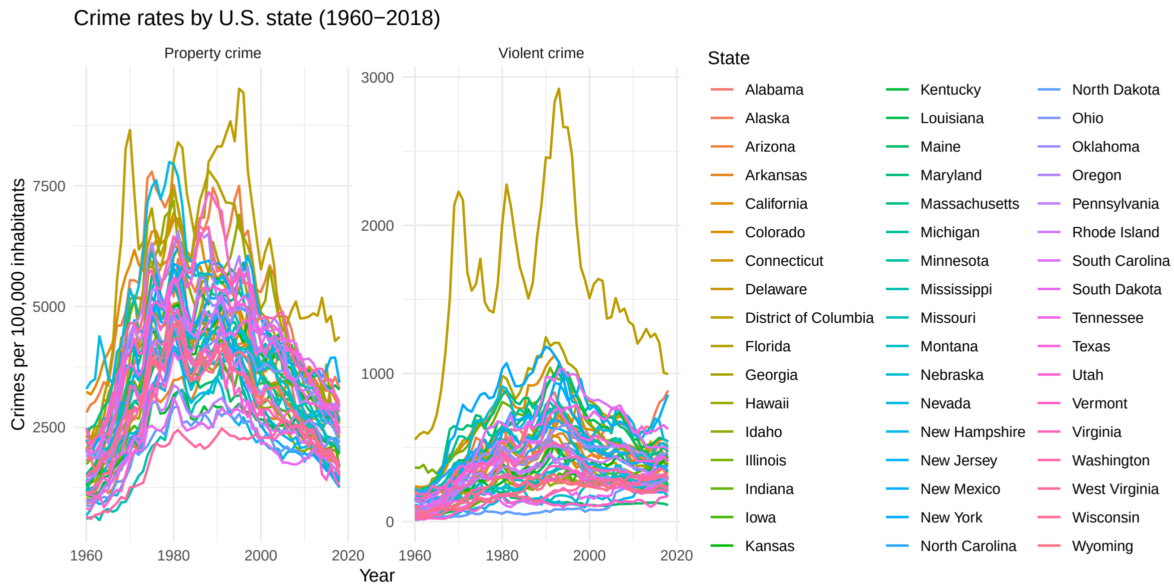

Figure 2: Violent and property crime rates by U.S. state

3.1.1.3 Evolution of violent crime and property crime among high-crime states

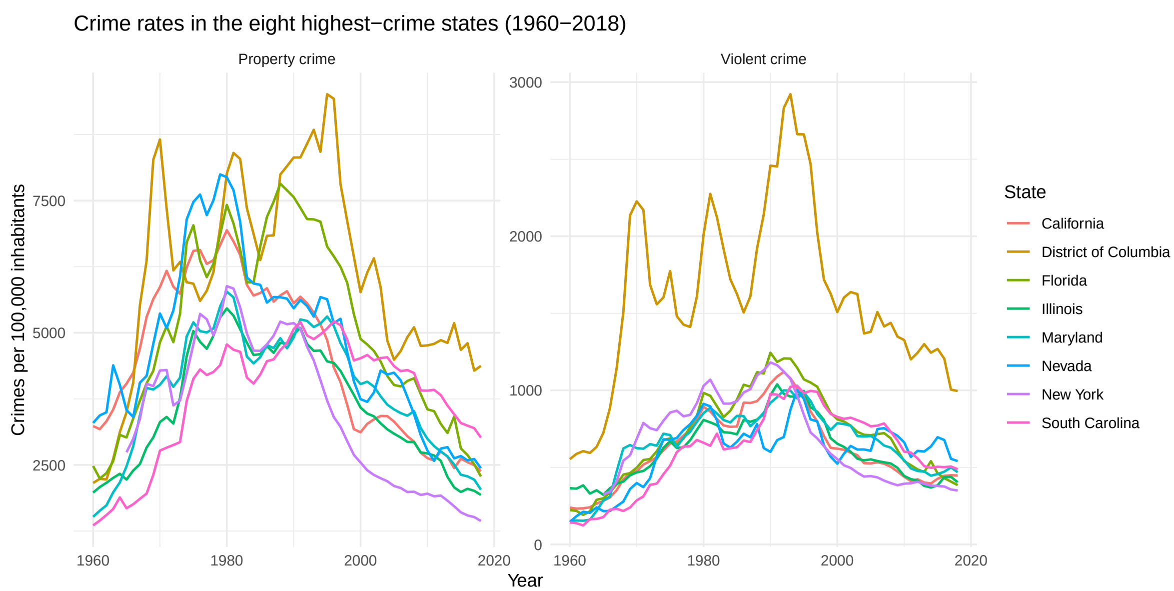

Figure 3: Violent and property crime rates in the eight highest-crime U.S. states (1960–2018)

3.1.1.4 Descriptive co-movement with economic structure

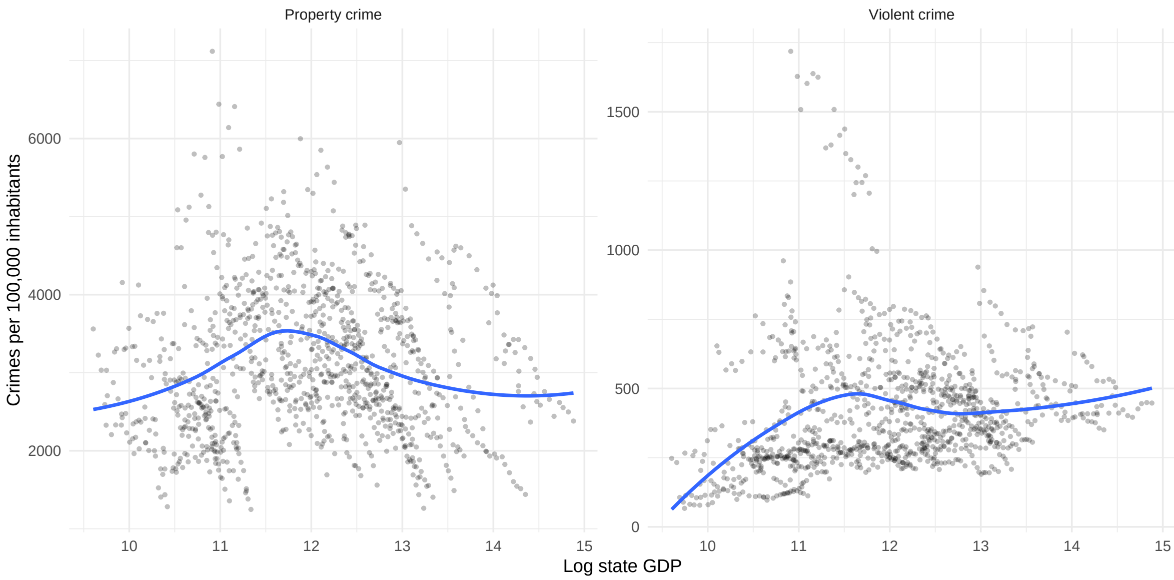

Figure 4: Raw co-movement between crime and GDP (pooled state–year observations)

Property crime: The pooled scatter plot reveals a clear non-linear relationship between property crime and state GDP. At low to intermediate levels of GDP, property crime rates tend to increase, before declining for higher levels of economic activity. The downward-sloping segment at higher GDP levels suggests that economically stronger states tend to experience lower property crime rates, consistent with the hypothesis that improved economic opportunities reduce incentives for financially motivated offences. The dispersion of observations around the fitted curve remains substantial, indicating that GDP alone cannot account for cross-state differences in property crime.

Violent crime: In contrast, the relationship between violent crime and GDP is markedly weaker. While violent crime rates increase sharply at low GDP levels, the association flattens and becomes less systematic as GDP rises. At higher levels of economic activity, violent crime exhibits no clear monotonic decline. This pattern suggests that violent crime is less directly tied to economic performance than property crime and is likely driven by additional social, institutional, and demographic factors.

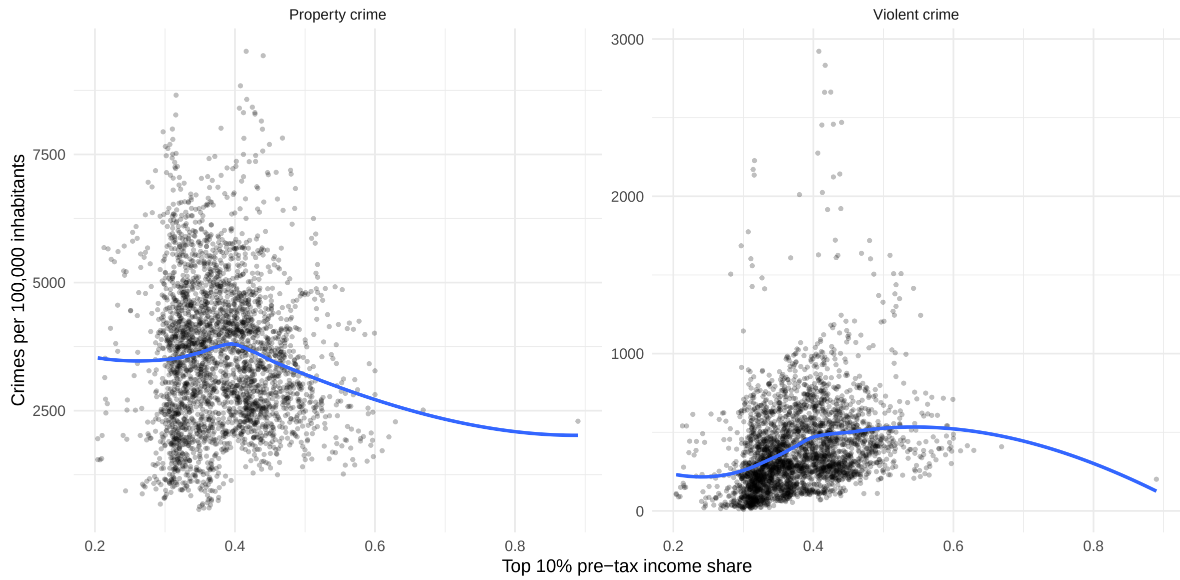

Figure 5: Raw co-movement between crime and income inequality (pooled state–year observations)

Property crime: The relationship between property crime and income inequality is non-linear and noisy. Property crime rates are highest at intermediate levels of inequality and decline at the upper end of the income share distribution. This inverted-U pattern suggests that inequality may matter for property crime, but its effect is neither simple nor monotonic. The wide dispersion of observations highlights the importance of unobserved state-specific characteristics.

Violent crime: It displays a more pronounced association with inequality. Crime rates increase strongly as the top income share rises from low to intermediate levels, before declining slightly at very high levels of inequality. This pattern is consistent with theories linking violent crime to social fragmentation, relative deprivation, and institutional stress, rather than purely economic incentives.

3.1.2 Regression analysis

Taken together, the descriptive evidence suggests that crime rates in U.S. states co-evolve with economic conditions and income inequality, but that these relationships are heterogeneous and potentially confounded by persistent state-specific characteristics and common national shocks. To account for these factors and move beyond raw correlations, we now estimate panel regressions with state and year fixed effects, which exploit within-state variation over time while controlling for unobserved heterogeneity.

3.1.2.1 Fixed effects (within)

We estimate the association between economic conditions, inequality, and crime using fixed effects panel regressions:

$$Crime_{it} = \alpha_i + \beta_1 \log(GDP_{it}) + \beta_2 \cdot Inequalities_{it} + \epsilon_{it}$$

where α_i denotes state fixed effects. The coefficients are identified from within-state variation over time.

First, we examine the bivariate relationship between GDP and crime (Model 1). Second, we add income inequality to test whether income concentration is independently associated with crime (Model 2). Third, as a robustness check, we include year fixed effects to control for national trends such as federal policies (Model 3).

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***	−1824.561***	−214.076
	(38.690)	(45.923)	(136.202)
share_pretax_income_p90p100		−874.021*	−676.111
		(427.873)	(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 ($\beta \approx -1875$, $p < 0.001$), explaining about 70% of within-state variation (Model 1). Adding inequality improves the fit marginally, with higher income concentration associated with lower property crime ($\beta = -874$, $p < 0.05$).

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***	−99.857***	119.753***
	(7.929)	(9.419)	(29.576)
share_pretax_income_p90p100		−144.862+	−224.361*
		(87.754)	(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 ($\beta \approx -100$, $R^2 = 0.15$), and inequality is only marginally significant ($p < 0.10$). Economic factors alone cannot explain violent crime patterns.

3.1.2.2 Between model

```
Oneway (individual) effect Between Model

Call:
plm(formula = Data.Rates.Property.All ~ log(gdp) + share_pretax_income_p90p100,
    data = panel_reg, model = "between", index = c("State", "Year"))

Balanced Panel: n = 51, T = 21, N = 1071
Observations used in estimation: 51

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1274.029  -648.313    43.586   529.392  2121.850

Coefficients:
                            Estimate Std. Error t-value Pr(>|t|)
(Intercept)                 1914.631   1276.458  1.5000   0.1402
log(gdp)                      93.164    118.709  0.7848   0.4364
share_pretax_income_p90p100  178.818   2660.767  0.0672   0.9467
```

```
Total Sum of Squares:    26549000
Residual Sum of Squares: 26034000
R-Squared:      0.019407
Adj. R-Squared: -0.021451
F-statistic: 0.474992 on 2 and 48 DF, p-value: 0.62478
```

3.1.2.3 Random model

```
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

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

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

Effects:
                   var  std.dev share
idiosyncratic  98483.1    313.8 0.155
individual    537683.5    733.3 0.845
theta: 0.907

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-1195.2741  -222.4411     7.5929   231.8969  1525.2828

Coefficients:
                             Estimate Std. Error  z-value  Pr(>|z|)
(Intercept)                 22167.489    515.360  43.0136 < 2.2e-16 ***
log_gdp                     -1520.572     48.409 -31.4106 < 2.2e-16 ***
share_pretax_income_p90p100 -1855.523    477.300  -3.8875 0.0001013 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    337580000
Residual Sum of Squares: 136690000
R-Squared:      0.59509
Adj. R-Squared: 0.59433
Chisq: 1569.64 on 2 DF, p-value: < 2.22e-16

Oneway (individual) effect Random Effect Model
```

```
   (Swamy-Arora's transformation)

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

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

Effects:
                   var  std.dev share
idiosyncratic  4142.50    64.36 0.094
individual    39971.31   199.93 0.906
theta: 0.9299

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-277.6049  -34.3691   -7.2163   28.2783  353.9061

Coefficients:
                             Estimate Std. Error z-value Pr(>|z|)
(Intercept)                 1530.5528    98.6469 15.5155  < 2e-16 ***
log_gdp                      -86.9506     9.1164 -9.5378  < 2e-16 ***
share_pretax_income_p90p100 -170.6205    87.8124 -1.9430  0.05201 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    5225600
Residual Sum of Squares: 4544400
R-Squared:      0.13034
Adj. R-Squared: 0.12872
Chisq: 160.073 on 2 DF, p-value: < 2.22e-16
```

3.1.2.4 Test de Hausman

```
    Hausman Test

data:  Data.Rates.Property.All ~ log_gdp + share_pretax_income_p90p100
chisq = 102.52, df = 2, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

```
   Hausman Test

data:  Data.Rates.Violent.All ~ log_gdp + share_pretax_income_p90p100
chisq = 76.953, df = 2, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

3.2 Short-run analysis (2015–2018) - Contemporary correlates of crime

3.2.1 Descriptive analysis

3.2.1.1 Descriptive overview of crime and new covariates

Summary statistics :

Table 3: Summary statistics, short-run panel (2015–2018)

variable	N	Mean	SD	P10	Median	
Violent crime rate (per 100k)	204	385.38	179.59	217.87	360.90	
Property crime rate (per 100k)	204	2419.80	672.01	1573.97	2428.65	
Log GDP	204	12.30	1.04	10.92	12.29	
Top 10% pretax income share	204	0.46	0.05	0.41	0.46	
Lawful permanent residents per million	204	2480.04	1553.79	952.88	2043.32	
Police shootings per million	201	3.57	2.23	1.39	3.01	
Public spending: education	102	20243296.58	24730036.07	3067044.90	13495795.00	3936
Public spending: welfare	102	13560265.53	19210452.25	2406766.90	7986219.50	2863
Public spending: health	102	2060307.16	3696803.27	273763.20	871195.50	433
Public spending: corrections	102	1577760.88	2368807.34	226187.10	788440.00	275
Public spending: police	102	2276882.65	3282188.59	334300.60	1304563.50	387
WIC participants (2015)	204	128.33	149.49	14.00	98.00	
School lunch participants (2015)	204	307.67	259.14	1.00	280.00	
School breakfast participants (2015)	204	202.71	177.15	24.00	180.00	

3.2.1.2 Contemporary heterogeneity and structural correlates (2015–2018)

Table 4: Contemporary socio-ecor

violent_quartile	Mean_violent_crime	Mean_property_crime	Log_GDP	Inequality_top10	LPR_pe
Low crime	210.95	1867.70	11.73	0.45	
Mid-low	304.23	2334.94	12.43	0.45	
Mid-high	409.25	2387.05	12.89	0.48	

High crime	617.10	3089.50	12.14	0.47

Table 5: Contemporary socio-ecor

property_quartile	Mean_property_crime	Mean_violent_crime	Log_GDP	Inequality_top10	LPR_p
Low crime	1617.99	256.91	12.21	0.48	
Mid-low	2118.10	324.05	12.27	0.44	
Mid-high	2677.65	429.27	12.56	0.47	
High crime	3265.47	531.31	12.15	0.46	

3.2.2 Regression analysis

4 Conclusion and limitations