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 nine sources :

- Crime rates (FBI/CORGIS, 1960–2019) distinguishing violent and property offences, and state population ;
- Income inequality (WID, 1960–2018) providing top income shares by state;
- GDP and consumption per capita (BEA Regional Accounts, 1998–2023);
- Unemployment rate (BLS, 1976-2025);
- Prison population (The Sentencing Project, 1980–2022);
- Poverty rates (Census Bureau, 1980–2023);
- Beer consumption per capita (NIAAA, 1970–2022);
- 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;
- Food assistance participation (USDA Food Environment Atlas, 2012-2023).

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, GDP, consumption, unemployment, prison population, poverty, and alcohol series provide consistent state-year observations between 1998-2018. It allows us to examine the structural relationships between economic conditions and crime using fixed effects panel regressions.

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 a shorter period to incorporate additional explanatory factors in a richer, shorter-term analysis.

3.1 Long-run analysis (1998–2018)

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.

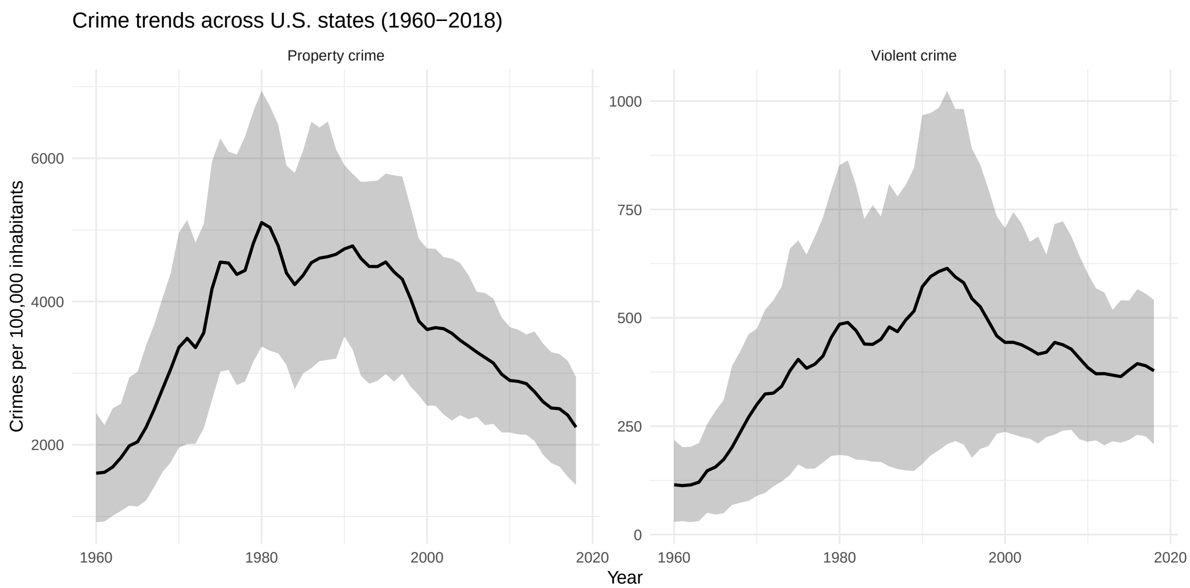

Figure 1: National crime trends and cross-state dispersion (mean and 10th–90th percentiles, 1960–2018)

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.

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.

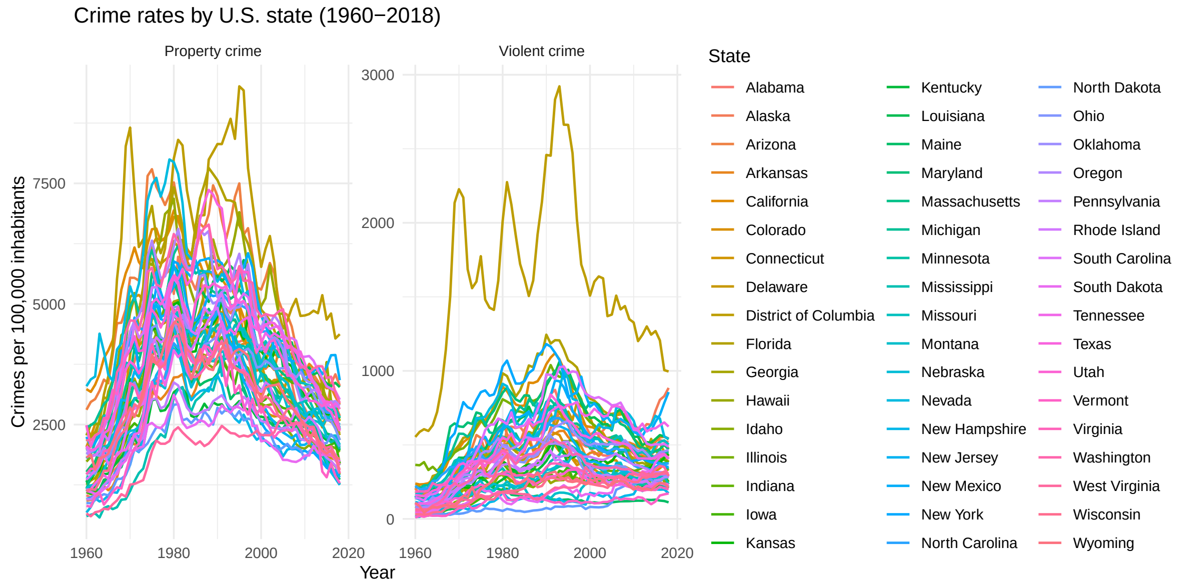

Figure 2: Violent and property crime rates by U.S. state

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

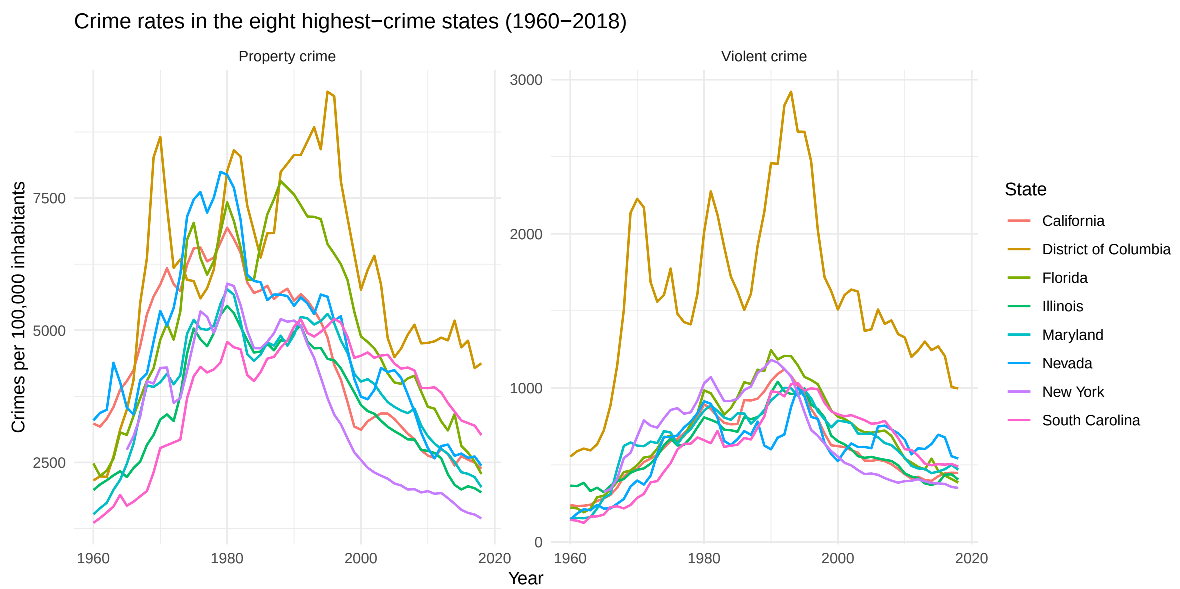

Figure 3: Violent and property crime rates in the eight highest-crime U.S. states (1960–2018)

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.

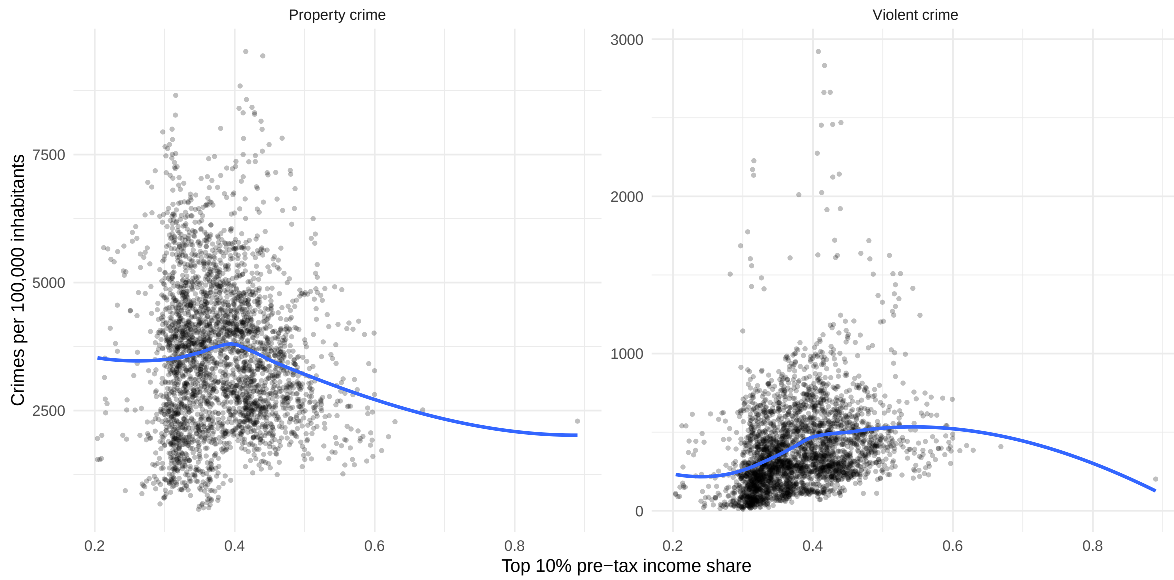

Figure 5: Raw co-movement between crime and income inequality (pooled state–year observations)

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 Hypotheses

We test the following hypotheses:

- **H1 (opportunity cost)** : Higher GDP and lower unemployment are associated with lower crime rates, particularly for property crime.
- **H2 (inequality)** : Greater income concentration at the top of the distribution is associated with higher crime rates.
- **H3 (deterrence)** : Larger prison populations are associated with lower crime rates because of incapacitation of criminals.
- **H4 (alcohol as a criminogenic factor)** : Higher alcohol consumption is associated with higher violent crime rates.
- **H5 (crime persistence)** : Crime rates exhibit temporal inertia, i.e. past crime is a strong predictor of current crime (it encourages or discourages entry, it owerwhelms law enforcers…).

We estimate different models for property and violent crime, as property crime is more directly linked to economic incentives (Becker, 1968), while violent crime may be more linked to specific social factors.

3.1.2.2 Fixed effects

First we test for the model that best suits our data.

Our baseline model looks like this :

$$Crime_{it} = \alpha_i + \beta_1 \log(GDP_{it}) + \beta_2 \cdot Inequality_{it} + \beta_3 \cdot Unemployment_{it} + \beta_4 \cdot Prison_{it} + \gamma_1 \cdot Beer_{it} + \gamma_2 \cdot \log(Pop_{it}) + \epsilon_{i}$$

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

Model 2 : we replace $log(GDP)$ by $Consumption$ to test another measure of economic conditions.

$$Crime_{it} = \alpha_i + \beta_1 \log(Consumption_{it}) + \beta_2 \cdot Inequality_{it} + \beta_3 \cdot Unemployment_{it} + \beta_4 \cdot Prison_{it} + \gamma_1 \cdot Beer_{it} + \gamma_2 \cdot \log(P$$

Model 3 : we replace $Unemployment$ by $Povertyrate$ to test another measure of economic difficulties

$$Crime_{it} = \alpha_i + \beta_1 \log(GDP_{it}) + \beta_2 \cdot Inequality_{it} + \beta_3 \cdot Poverty_{it} + \beta_4 \cdot Prison_{it} + \gamma_1 \cdot Beer_{it} + \gamma_2 \cdot \log(Pop_{it}) + \epsilon_{it}$$

To look into crime persistence (i.e., the tendency of high-crime states to remain high-crime), we estimate a dynamic panel model including lagged crime rates as an explanatory variable (Model 4).

$$Crime_{it} = \alpha_i + \theta \cdot Crime_{i,t-1} + \beta_1 \log(GDP_{it}) + \beta_2 \cdot Inequality_{it} + \beta_3 \cdot Unemployment_{it} + \beta_4 \cdot Prison_{it} + \gamma_1 \cdot Beer_{it} + \gamma_2$$

```
Oneway (individual) effect Within Model

Call:
plm(formula = Data.Rates.Property.All ~ crime_lag + log_gdp +
    share_pretax_income_p90p100 + unemployment_rate + prison_population +
    beer_per_capita_14plus + log_pop, data = filter(panel_reg,
    !is.na(crime_lag)), model = "within", index = c("State",
    "Year"))

Balanced Panel: n = 50, T = 20, N = 1000

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-678.2384  -81.4457   -2.7366   85.6378  927.5653

Coefficients:
                               Estimate  Std. Error t-value  Pr(>|t|)
crime_lag                    8.0257e-01  1.8023e-02 44.5307 < 2.2e-16 ***
log_gdp                     -3.1195e+02  4.7477e+01 -6.5706 8.279e-11 ***
share_pretax_income_p90p100 -3.7873e+02  2.1747e+02 -1.7415   0.08192 .
unemployment_rate            2.4786e+00  3.0042e+00  0.8250   0.40955
prison_population           -2.3842e-03  1.3956e-03 -1.7084   0.08789 .
beer_per_capita_14plus       5.4072e+02  8.6938e+01  6.2196 7.483e-10 ***
log_pop                      4.4552e+02  1.7872e+02  2.4928   0.01284 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    268820000
Residual Sum of Squares: 22114000
R-Squared:      0.91774
Adj. R-Squared: 0.91285
F-statistic: 1502.86 on 7 and 943 DF, p-value: < 2.22e-16
```

The model explains 91,7% ($R^2 = 0.9177$) of the variation in crime rates within states over time.

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

$$Crime_{it} = \alpha_i + \delta_t + \theta \cdot Crime_{i,t-1} + \beta_1 \log(GDP_{it}) + \beta_2 \cdot Inequality_{it} + \beta_3 \cdot Unemployment_{it} + \beta_4 \cdot Prison_{it} + \gamma_1 \cdot Beer_{it}$$

```
Twoways effects Within Model

Call:
plm(formula = Data.Rates.Property.All ~ crime_lag + log_gdp +
    share_pretax_income_p90p100 + unemployment_rate + prison_population +
    beer_per_capita_14plus + log_pop, data = panel_reg, effect = "twoways",
    model = "within", index = c("State", "Year"))

Balanced Panel: n = 50, T = 20, N = 1000

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-708.4732  -74.7897   -3.0698   72.1411  853.4436

Coefficients:
                               Estimate  Std. Error t-value Pr(>|t|)
crime_lag                    8.0543e-01  1.8496e-02 43.5449  < 2e-16 ***
log_gdp                     -1.0277e+02  9.5126e+01 -1.0803  0.28027
share_pretax_income_p90p100 -1.4307e+02  2.2488e+02 -0.6362  0.52479
unemployment_rate            7.5586e-01  6.7233e+00  0.1124  0.91051
prison_population           -2.9297e-03  1.3162e-03 -2.2258  0.02627 *
beer_per_capita_14plus       2.1815e+02  9.3317e+01  2.3377  0.01961 *
log_pop                      3.0999e+01  1.7679e+02  0.1753  0.86085
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    64776000
Residual Sum of Squares: 18343000
R-Squared:      0.71682
Adj. R-Squared: 0.69383
F-statistic: 334.132 on 7 and 924 DF, p-value: < 2.22e-16

Twoways effects Within Model

Call:
plm(formula = Data.Rates.Violent.All ~ crime_lag + log_gdp +
```

```
    share_pretax_income_p90p100 + unemployment_rate + prison_population +
    beer_per_capita_14plus + log_pop, data = panel_reg, effect = "twoways",
    model = "within", index = c("State", "Year"))

Balanced Panel: n = 50, T = 20, N = 1000

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-160.12092  -25.58584    0.42004   22.65859  187.07184

Coefficients:
                               Estimate  Std. Error t-value  Pr(>|t|)
crime_lag                    7.8254e-02  6.1273e-03 12.7714 < 2.2e-16 ***
log_gdp                      9.4061e+01  3.1512e+01  2.9849  0.002911 **
share_pretax_income_p90p100 -1.9019e+02  7.4494e+01 -2.5532  0.010835 *
unemployment_rate           -1.2508e+00  2.2272e+00 -0.5616  0.574526
prison_population            8.1964e-04  4.3603e-04  1.8798  0.060452 .
beer_per_capita_14plus       1.7567e+02  3.0913e+01  5.6827 1.777e-08 ***
log_pop                      7.5780e+01  5.8564e+01  1.2940  0.196000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    2673900
Residual Sum of Squares: 2013000
R-Squared:      0.24717
Adj. R-Squared: 0.18606
F-statistic: 43.3374 on 7 and 924 DF, p-value: < 2.22e-16
```

First, we examine the 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 !

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

$$Crime_{it} = \alpha_i + \beta_1 \log(GDP_{it}) + \beta_2 \cdot Inequalities_{it} + \beta_3 \cdot Consumption_{it} + \beta_4 \cdot Unemployment_{it} + \beta_5 \cdot Population_{it} + \epsilon_i$$

```
Oneway (individual) effect Within Model

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

Balanced Panel: n = 50, T = 20, N = 1000

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-173.9448  -25.7728    0.2792   25.2172  256.2941

Coefficients:
                               Estimate  Std. Error  t-value  Pr(>|t|)
log_gdp                      1.2691e+02  2.6608e+01   4.7698 2.135e-06 ***
share_pretax_income_p90p100 -1.4085e+02  7.1121e+01  -1.9804   0.04795 *
Consumption                 -5.7171e-03  1.0630e-03  -5.3783 9.481e-08 ***
unemployment_rate           -7.3870e+00  9.5694e-01  -7.7194 2.969e-14 ***
Data.Population             -4.5076e-05  3.3298e-06 -13.5372 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    3443100
Residual Sum of Squares: 2369700
R-Squared:      0.31177
Adj. R-Squared: 0.27244
F-statistic: 85.6161 on 5 and 945 DF, p-value: < 2.22e-16
```

```
Twoways effects Within Model

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

Balanced Panel: n = 50, T = 20, N = 1000

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-725.9977 -141.7998    0.0851  132.4228 1371.9751

Coefficients:
                               Estimate  Std. Error t-value  Pr(>|t|)
log_gdp                     -2.4226e+02  1.5031e+02 -1.6117  0.107359
share_pretax_income_p90p100  3.5297e+02  3.9543e+02  0.8926  0.372288
Consumption                  6.9814e-02  8.7775e-03  7.9537 5.261e-15 ***
unemployment_rate            3.0649e+01  1.1367e+01  2.6963  0.007138 **
Data.Population             -1.2736e-04  1.6635e-05 -7.6561 4.813e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    64776000
Residual Sum of Squares: 56210000
R-Squared:      0.13225
Adj. R-Squared: 0.063841
F-statistic: 28.2253 on 5 and 926 DF, p-value: < 2.22e-16

Twoways effects Within Model

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

Balanced Panel: n = 50, T = 20, N = 1000

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
```

```
-158.31216  -22.64855    0.94243   21.66623  249.92022

Coefficients:
                              Estimate  Std. Error  t-value  Pr(>|t|)    
log_gdp                     1.9806e+02  2.9353e+01   6.7473 2.651e-11 ***
share_pretax_income_p90p100 -8.7611e+01  7.7223e+01  -1.1345    0.2569    
Consumption                 9.7004e-04  1.7142e-03   0.5659    0.5716    
unemployment_rate           1.6463e+00  2.2199e+00   0.7416    0.4585    
Data.Population            -4.4152e-05  3.2486e-06 -13.5911 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    2673900
Residual Sum of Squares: 2143700
R-Squared:      0.19827
Adj. R-Squared: 0.13507
F-statistic: 45.8007 on 5 and 926 DF, p-value: < 2.22e-16
```

Table 1: Property crime: Panel regression results

	Model 1	Model 2	Model 3	Model 4
log_gdp	−581.975***	−242.255	126.915***	198.056***
	(153.216)	(150.306)	(26.608)	(29.353)
share_pretax_income_p90p100	−381.136	352.972	−140.849*	−87.611
	(409.535)	(395.429)	(71.121)	(77.223)
Consumption	−0.043***	0.070***	−0.006***	0.001
	(0.006)	(0.009)	(0.001)	(0.002)
unemployment_rate	−6.463	30.649**	−7.387***	1.646
	(5.510)	(11.367)	(0.957)	(2.220)
Data.Population	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Num.Obs.	1000	1000	1000	1000
R2	0.708	0.132	0.312	0.198
R2 Adj.	0.691	0.064	0.272	0.135

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

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 = 50, T = 20, N = 1000

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-680.43 -155.96    4.50  139.28 1301.63

Coefficients:
                            Estimate Std. Error t-value Pr(>|t|)
log_gdp                      -132.08     135.16 -0.9772   0.3287
share_pretax_income_p90p100  -347.58     416.23 -0.8351   0.4039

Total Sum of Squares:    64776000
Residual Sum of Squares: 64657000
R-Squared:      0.0018412
Adj. R-Squared: -0.07337
F-statistic: 0.856818 on 2 and 929 DF, p-value: 0.42485

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 = 50, T = 20, N = 1000

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-206.57571  -27.88499   -0.82886   24.02349  262.37118

Coefficients:
                            Estimate Std. Error t-value  Pr(>|t|)
log_gdp                      140.011     27.029  5.1800 2.722e-07 ***
share_pretax_income_p90p100 -198.062     83.237 -2.3795   0.01754 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Total Sum of Squares:    2673900
Residual Sum of Squares: 2585800
R-Squared:      0.032939
Adj. R-Squared: -0.039929
F-statistic: 15.8211 on 2 and 929 DF, p-value: 1.7514e-07
```

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 2: Property crime: Panel regression results

	Model 1	Model 2	Model 3
log_gdp	−1836.553***	−1795.205***	−132.076
	(41.622)	(49.027)	(135.160)
share_pretax_income_p90p100		−677.120	−347.583
		(425.157)	(416.225)
Num.Obs.	1000	1000	1000
R2	0.672	0.673	0.002
R2 Adj.	0.655	0.656	−0.073

+ 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 3: Violent crime: Panel regression results

	Model 1	Model 2	Model 3
log_gdp	−82.645***	−74.269***	140.011***
	(7.779)	(9.161)	(27.029)
share_pretax_income_p90p100		−137.166+	−198.062*
		(79.444)	(83.237)
Num.Obs.	1000	1000	1000
R2	0.106	0.109	0.033
R2 Adj.	0.059	0.061	−0.040

+ 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.

3.1.2.3 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 = 50, T = 20, N = 1000
Observations used in estimation: 50

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-1084.7000  -595.0844     1.3857   531.6803  1159.9835

Coefficients:
                            Estimate Std. Error t-value Pr(>|t|)
(Intercept)                  1935.01    1143.12  1.6928  0.09712 .
log(gdp)                      170.45     108.90  1.5652  0.12424
share_pretax_income_p90p100 -2164.95    2461.35 -0.8796  0.38356
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    21475000
```

```
Residual Sum of Squares: 20411000
R-Squared:      0.049544
Adj. R-Squared: 0.0090995
F-statistic: 1.22498 on 2 and 47 DF, p-value: 0.30297
```

3.1.2.4 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 = 50, T = 20, N = 1000

Effects:
                   var  std.dev share
idiosyncratic  92673.3    304.4 0.177
individual    429636.2    655.5 0.823
theta: 0.8967

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-1153.6297  -229.2911     3.5301   219.5673  1358.9825

Coefficients:
                             Estimate Std. Error  z-value  Pr(>|z|)
(Intercept)                 20562.010    551.952  37.2533 < 2.2e-16 ***
log_gdp                     -1387.157     51.507 -26.9316 < 2.2e-16 ***
share_pretax_income_p90p100 -1935.988    484.470  -3.9961  6.44e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    273400000
Residual Sum of Squares: 126440000
R-Squared:      0.53753
Adj. R-Squared: 0.5366
Chisq: 1158.82 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 = 50, T = 20, N = 1000

Effects:
                   var  std.dev share
idiosyncratic  3235.76    56.88 0.131
individual    21555.35   146.82 0.869
theta: 0.9137

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-191.7032  -32.3857   -7.9087   27.2613  253.6604

Coefficients:
                             Estimate Std. Error z-value  Pr(>|z|)
(Intercept)                 1128.7751    93.0321 12.1332 < 2.2e-16 ***
log_gdp                      -54.5504     8.6227 -6.3264  2.51e-10 ***
share_pretax_income_p90p100 -188.7634    79.2884 -2.3807   0.01728 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    3618000
Residual Sum of Squares: 3337400
R-Squared:      0.077563
Adj. R-Squared: 0.075712
Chisq: 83.8325 on 2 DF, p-value: < 2.22e-16
```

3.1.2.5 Test de Hausman

```
    Hausman Test

data:  Data.Rates.Property.All ~ log_gdp + share_pretax_income_p90p100
chisq = 93.351, 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 = 73.085, df = 2, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

Our estimates capture associations, not causal effects. Indeed, high crime may reduce GDP or increase prison population, which implies reverse causality. Moreover, there are omitted variables in our work, for instance time-varying factors like policing strategies, population structure and demographic shift may confounf our estimates.

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 4: 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 5: Contemporary socio-econ

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 6: Contemporary socio-econ

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