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

5382 words

Paula de Miguel Sánchez & Esther Allevi

Research question

Crime rates in the United States differ markedly across states and over time, even in the presence of common national trends. This project examines how crime co-evolves with economic conditions and socials characteristics at the state level, and whether these relationships differ between property and violent crime.

The core objective of the analysis is to document robust within-state associations between crime and key economic and social variables, such as economic activity, unemployment, poverty, income inequality, incarceration and alcohol consumption, using a long state-level panel over the period 1998–2018. By relying on fixed effects models, we focus on how changes within a given state are associated with changes in crime, while controlling for unobserved state characteristics and nationwide shocks.

In a second, complementary step, we conduct a short-run exploratory analysis (2015–2018) to incorporate additional institutional and social indicators that are only available for recent years, including police shootings, immigration flows and public spending. Given the limited time dimension, this part of the analysis is descriptive and aims to highlight contemporaneous cross-state correlations rather than causal effects.

Principally, the project asks whether economic and social variables are more strongly associated with property crime than with violent crime, and to what extent crime dynamics exhibit persistence beyond contemporaneous economic conditions.

Brief dataset description

We built our panel by usind nine different sources :

- Property and Violent crime rates, state population (FBI/CORGIS, 1960–2019) ;
- Income inequality providing top income shares by state (WID, 1960–2018);

- 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 for population aged 14+ (NIAAA, 1970–2022);

- Immigration flows covering legal residents, refugees and asylees (DHS, 2013–2023);

- Police shootings (Washington Post, 2015–2024);

- Public spending on education, welfare, health, corrections and police (Census Bureau, 2017–2023);

- Food assistance participation (USDA Food Environment Atlas, 2012-2023).

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

Data analysis

This 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, police shootings, public spending and food access are only available for recent years (2015-2018). Considering the lack of historical depth, but also the value of these contemporary indicators, we conduct an exploratory analysis, that should be interpreted accordingly.

Long-run analysis (1998–2018)

Descriptive analysis

National long-run trends in crime (1998-2018)

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. Crime trends across U.S. states (1960–2018)
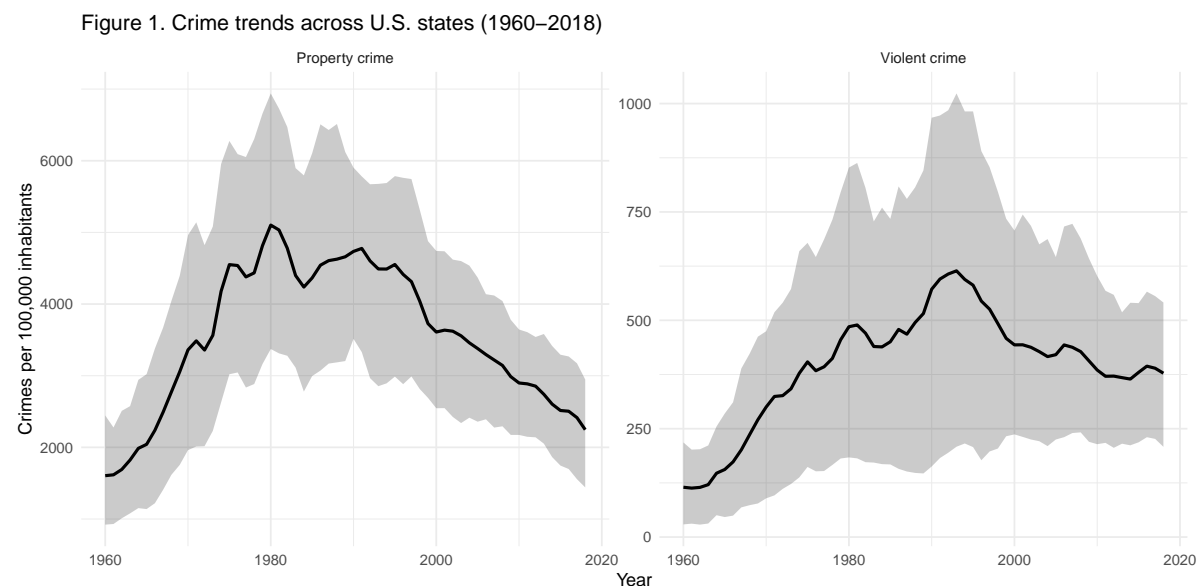

Figure 1: National crime trends and cross-state dispersion (mean and 10th–90th percentiles, 1960–2018)

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.

Due to data availability constraints, our focus will be on the period 1998-2018, in the midst of a crime decline in the US states.

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

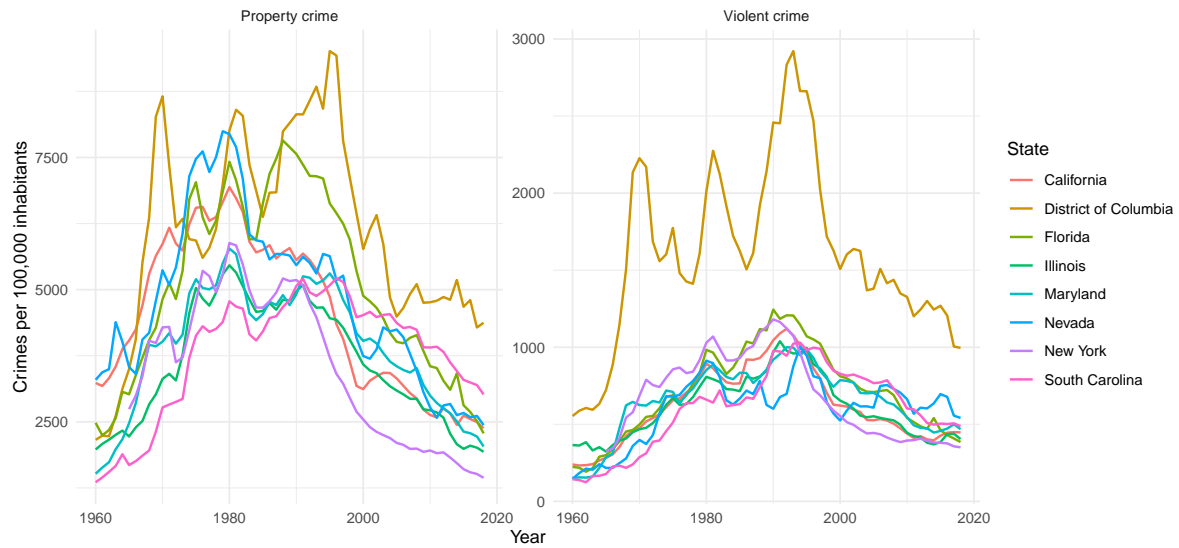Figure 2. Crime rates in the eight highest–crime states (1960–2018)

Figure 2: Violent and property crime rates in the eight highest-crime U.S. states (1960–2018)

Figure 2 displays the crime trends for the eight states with the highest crime rate per 100 000 inhabitants between 1960 and 2018. Despite differences in levels, all states exhibit similar trends : a strong increase until the late 1980s or early 1990s, followed by a sustained decline. Cross-state heterogeneity also remains pronounced.

The District of Columbia stands out, both for its higher levels and for its volatility, <hich might reflect its unique urban and institutional characteristics. This makes it difficult to compare it with US states. While it is included in this descriptive analysis, it is excluded from the regression sample due to missing values for key variables.

Descriptive co-movement with economic structure

Figure 3: Raw co-movement between crime and GDP (pooled state–year observations)

The pooled scatter plot reveals a non-linear relationship between property crime and state GDP. Property crime increases at low GDP levels before declining at higher levels, consistent with the opportunity cost hypothesis. However, substantial dispersion around the fitted curve indicates that GDP alone cannot explain cross-state differences in property crime.

In contrast, violent crime shows a weaker association to GDP. While crime rises sharply at low GDP, the relationship flattens at higher levels with no clear monotonic decline. This suggests violent crime is less tied to economic performance and more driven by social, institutional, and demographic factors.
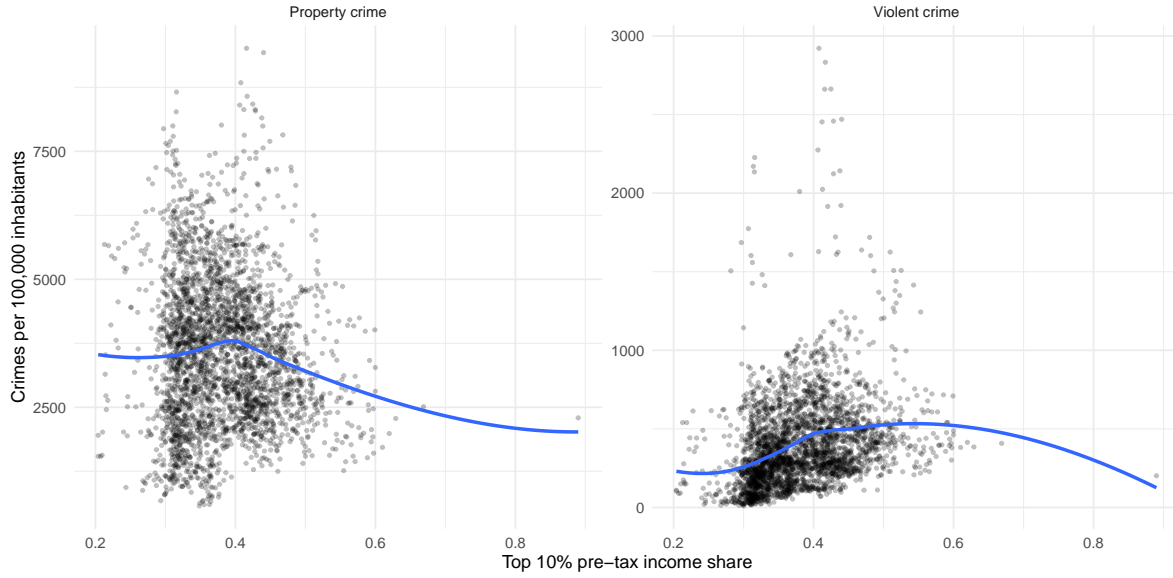
Figure 4: Raw co-movement between crime and income inequality (pooled state–year observations)

Property crime exhibits a non-linear and noisy relationship with inequality: highest at intermediate levels, declining at high inequality. This suggests inequality matters for property crime, though not monotonically. Wide dispersion indicates important unobserved state characteristics.

Violent crime shows a stronger association with inequality, rising sharply as top incomes rises from low to intermediate levels before declining slightly at very high inequality. This pattern aligns with theories linking violent crime to social fragmentation and relative deprivation rather than economic incentives.

The 5 following figures presents pooled scatter plots relating crime rates to key socio-economic variables across state-year observations.
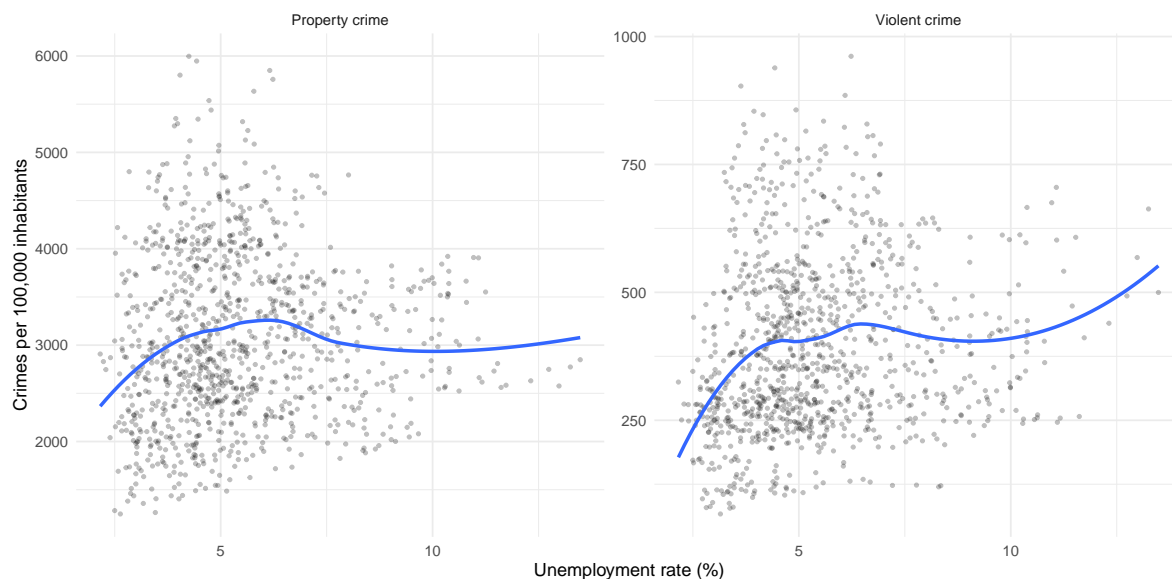
Figure 5: Raw co-movement between crime and unemployment rate (pooled state–year observations)

The relationship between unemployment and crime appears weak and non-linear. Property crime peaks at intermediate unemployment levels, while violent crime rises modestly with unemployment. The substantial dispersion suggests unemployment alone is not a strong crime predictor.

Figure 6: Raw co-movement between crime and percentage of poor (pooled state–year observations)

Poverty displays a modest positive association with both crime types, slightly stronger for violent crime. However, substantial dispersion around the LOESS curve indicates poverty alone poorly predicts crime across states.

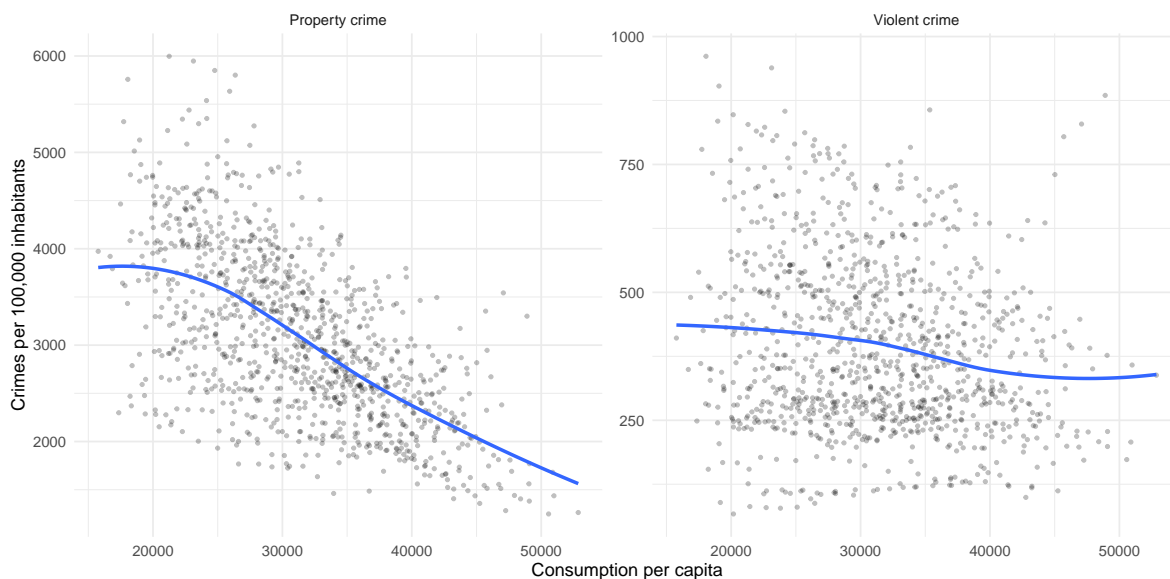Figure 7: Raw co-movement between crime and consumption per capita (pooled state–year observations)

Consumption per capita shows a clear negative association with both crime types. States with higher living standards systematically experience lower property and violent crime rates.
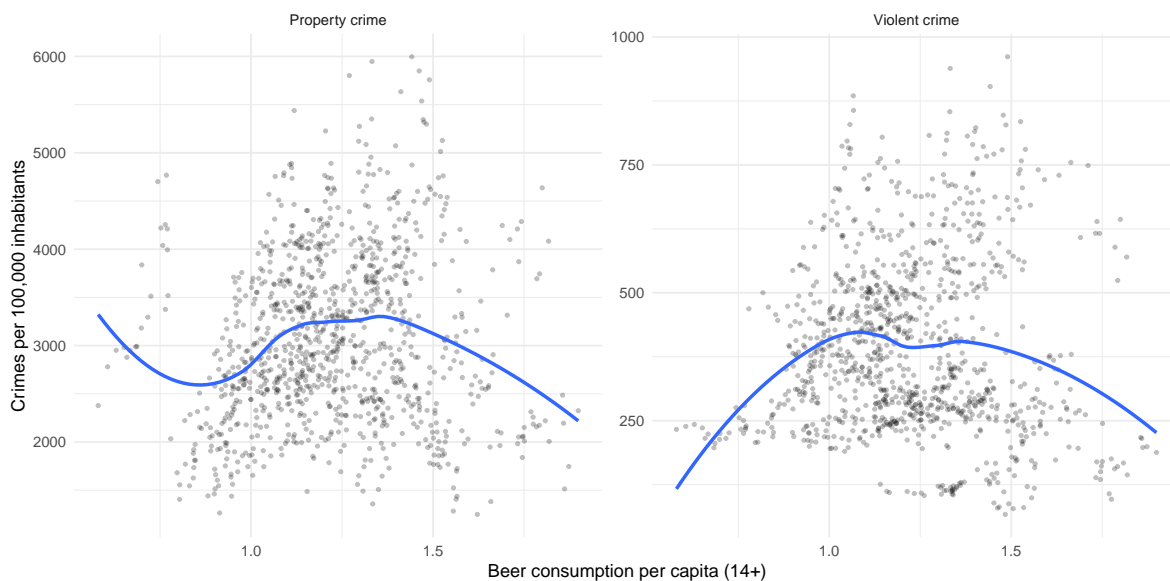

Figure 8: Raw co-movement between crime and beer consumption per capita for people aged 14 and above (pooled state–year observations)

Beer consumption per capita shows a non-linear, inverted-U relationship with crime. Property crime increases with beer consumption up to approximately 1.25 gallons per capita, then levels off. For violent crime, the pattern is similar but noisier. These graphs support the inclusion of alcohol as a criminogenic factor in our regression models.
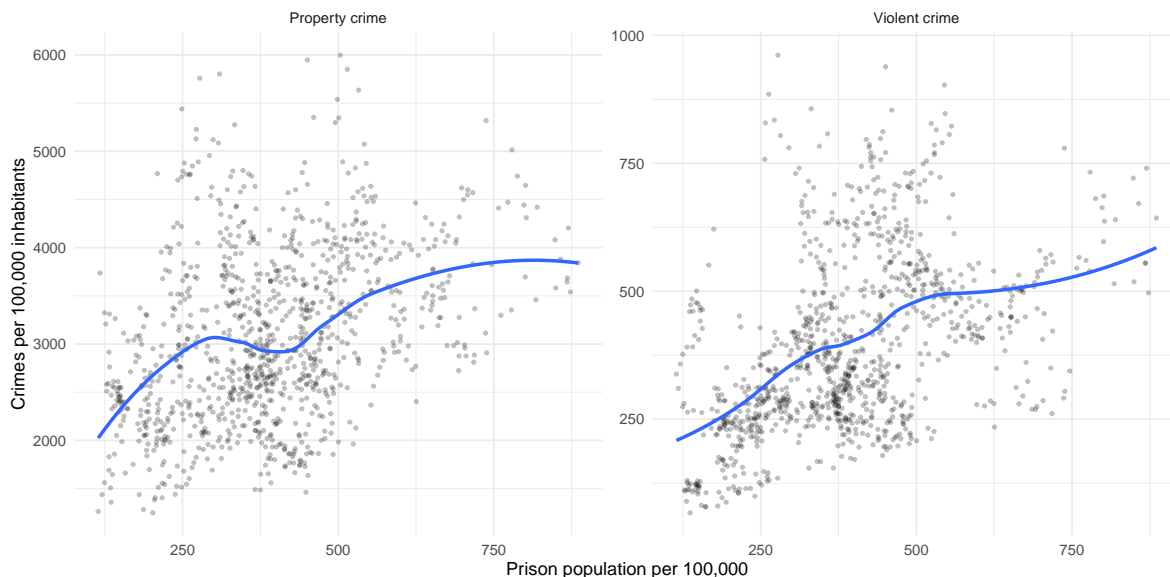

Figure 9: Raw co-movement between crime and prison population per 100 000 inhabitants (pooled state–year observations)

Prison population per 100 000 inhabitants is positively associated to both property and violent crime. States with higher incarceration rates tend to exhibit higher crime, likely reflecting reverse causality rather than failed deterrence.

Several variables display clear raw co-movements with crime, although the relationships are often non-linear and differ between property and violent crime. In particular, beer consumption and economic activity measures exhibit strong positive associations, while the relationship with unemployment and poverty appears weaker and less stable. These descriptive patterns motivate the regression strategy adopted in the following section.

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.

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.

Tests

Before presenting our results, we conduct a series of tests to validate our approach.

The first question is whether state-level heterogeneity exists. We test this using two complementary approaches.

Table 1: F-test and Breusch-Pagan LM test results

Test	Statistic	p-value
F-test (Pooled vs FE)	148.8985	< 0.0001
Breusch-Pagan LM (Pooled vs RE)	3047.7445	< 0.0001

The F-test compares the fixed effects model against pooled OLS. The null hypothesis is that all state fixed effects are jointly equal to zero ($H_0 = \alpha_1 = ... = \alpha_2$). We obtain $F = 148.9$ with $p < 0.001$. Thus, the F-test strongly rejects the null hypothesis that state fixed effects are jointly zero, indicating that pooled OLS is inappropriate.

The Breusch-Pagan Lagrange Multiplier test provides a complementary check. It tests whether the variance of individual effects is zero ($H_0 : \sigma_\alpha^2 = 0$). We obtain $\chi^2 = 3047.7$ with $p < 0.001$, which allows us to reject the null hypothesis.

These tests confirm that individual effects are present and statistically significant. Having established that, we still have to choose between random effects and fixed effects. To do so, we use the Hausman test :

Table 2: Hausman test results (fixed vs random effects)

Test	Statistic	p-value
Hausman (FE vs RE)	595.8735	< 0.0001

Under H_0, both fixed effects and random effects are consistent, but random effects is more efficient ; under H_1, only fixed effects is consistent. We obtain $\chi^2 = 595.9$ with $p < 0.001$, so we can reject H_0.

This indicates that state-specific effects are correlated with our explanatory variables (GDP, inequality, etc.), which makes sense since wealthier states likely have different unobserved characteristics than poorer states.

Next, we test for autocorrelation and heteroscedasticity. Because of the nature of crime according to our hypothesis (persistence), and of our explanatory variables (GDP, inequalities, prison population, etc change slowly over time), we may expect serial correlation. Moreover, heteroscedasticity is also plausible given the variation in state size and crime levels. Smaller states might experience more volatility in their crime rates because small absolute changes translate into large percentage changes.

Table 3: Specification tests for the fixed-effects model

Test	Statistic	p-value
Breusch–Godfrey, AR(1)	527.4423	< 0.0001
Breusch–Pagan	30.5139	< 0.0001

The specification tests reveal two issues, serial correlation and heteroscedasticity. They do not bias our coefficient estimates but invalidate classic standard errors, potentially leading to misleading significance levels. So, we compute robust standard errors, clustered at the state level, and compare them with the correlated and heterodastic standard errors.

Table 4: Comparison of classical vs robust standard errors

Variable	Coefficient	SE	SE_robust	t	t_robust	p	p_robust
Log(GDP)	-1509.764	47.372	169.348	-31.870	-8.915	0.000	0.000
Inequality (top 10%)	-422.111	388.825	710.335	-1.086	-0.594	0.278	0.552

Variable	Coefficient	SE	SE_robust	t	t_robust	p	p_robust
Unemployment rate	-9.029	5.330	6.629	-1.694	-1.362	0.091	0.173
Prison per 100k	0.997	0.248	0.610	4.017	1.633	0.000	0.103
Beer per capita (14+)	1833.486	132.857	457.610	13.800	4.007	0.000	0.000

Classical standard errors are 1.2 to 3.5 times smaller than robust standard errors, leading to inflated significance levels. For instance, prison population per 100 000 inhabitants appears significant with classical errors ($p < 0.001$) but becomes non-significant with robust errors ($p = 0.103$). Given these findings, all subsequent results use robust standard errors clustered by state.

Finally, we test for multicolinearity, using Variance Inflation Factors (VIF).

Table 5: Variance inflation factors for all covariates

	VIF
log_gdp	92.90
Consumption	5.67
unemployment_rate	1.40
percent_poor	1.90
share_pretax_income_p90p100	1.56
prison_per_100k	1.69
beer_per_capita_14plus	1.32
log_pop	86.25

We do not include GDP, consumption and population simultaneously because of the multicollinearity it clearly introduces. Instead, we will estimate alternative specifications to assess which measure is most strongly associated with crime (see following models).

For now, we look at the VIF when we only keep one variable among log(gdp), log(pop) and consumption.

Table 6: Variance inflation factors with log(gdp)

	VIF
log_gdp	3.53
share_pretax_income_p90p100	1.42
unemployment_rate	1.37
prison_population	2.74
percent_poor	1.40

	VIF
beer_per_capita_14plus	1.38

Table 7: Variance inflation factors with log(pop)

	VIF
log_pop	3.33
share_pretax_income_p90p100	1.33
unemployment_rate	1.35
prison_population	2.81
percent_poor	1.35
beer_per_capita_14plus	1.34

Table 8: Variance inflation factors with consumption

	VIF
Consumption	1.29
share_pretax_income_p90p100	1.47
unemployment_rate	1.38
prison_population	1.28
percent_poor	1.48
beer_per_capita_14plus	1.12

Without consumption, population, and GDP being included simultaneously, all VIF values are below 5, the conventional threshold for concern, so our explanatory variables are sufficiently independent.

Table 9: Summary of model specification tests results

Test	Statistic	p_value	Decision
F-test (Pooled vs FE)	148.89854	< 0.001	FE needed
Breusch-Pagan LM	3047.74448	< 0.001	Individual effects exist
Hausman (FE vs RE)	595.87355	< 0.001	FE preferred
Breusch-Godfrey (autocorr.)	527.44225	< 0.001	Autocorrelation present
Breusch-Pagan (heterosced.)	30.51393	< 0.001	Heteroscedasticity present

Models with fixed effects

Our objective is to document robust correlations between socio-economic variables and crime rates, not to establish causal effects. Considering our tests conclusions, we employ fixed effects panel regressions to control for time-invariant state characteristics.

We estimate five panel specifications with state fixed effects to assess the robustness of our findings. All models use robust standard errors clustered by state to account for serial correlation and heteroscedasticity.

Model 1 (Baseline) includes log GDP, income inequality, unemployment rate, prison population per capita and beer consumption. This specification tests the baseline correlations between economic conditions and crime.

$$
\begin{aligned}
Crime_{it} = \alpha_i &+ \beta_1 \log(GDP_{it}) + \beta_2 \cdot Inequality_{it} \\
&+ \beta_3 \cdot Unemployment_{it} + \beta_4 \cdot Prison_{it} \\
&+ \gamma_1 \cdot Beer_{it} + \epsilon_{it}
\end{aligned}
$$

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

We measure inequalities by 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).

Model 2 (Consumption) replaces $log(GDP)$ by $Consumption$. While GDP measures the size of the state economy, per capita consumption better captures household living standards. Comparing the two allows us to assess which dimension of economic conditions matters most for crime.

$$
\begin{aligned}
Crime_{it} = \alpha_i &+ \beta_1 \log(Consumption_{it}) + \beta_2 \cdot Inequality_{it} \\
&+ \beta_3 \cdot Unemployment_{it} + \beta_4 \cdot Prison_{it} \\
&+ \gamma_1 \cdot Beer_{it} + \gamma_2 \cdot \log(Pop_{it}) + \varepsilon_{it}
\end{aligned}
$$

Model 3 (Population) replaces $log(GDP)$ by $log(Pop)$, to test whether the size of the state population is associated with property crime. This allows to compare the influence of the total population per state to the GDP.

$$
\begin{aligned}
Crime_{it} = \alpha_i &+ \beta_1 \log(Pop_{it}) + \beta_2 \cdot Inequality_{it} \\
&+ \beta_3 \cdot Poverty_{it} + \beta_4 \cdot Prison_{it} \\
&+ \gamma_1 \cdot Beer_{it} + \varepsilon_{it}
\end{aligned}
$$

Model 4 (Dynamic) adds lagged crime as an explanatory variable to test the persistence hypothesis (H5) (i.e., the tendency of high-crime states to remain high-crime). If crime exhibits inertia, the lagged coefficient should be positive and significant. We acknowledge that including a lagged dependent variable in fixed effects models introduces Nickell bias, which attenuates the coefficient toward zero when T is small. With T = 20 years, this bias is modest but not negligible. Results from Model 4 should therefore be interpreted cautiously, as the true persistence coefficient may be slightly higher than estimated.

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

Model 5 (Two-ways) adds year fixed effects (δ_t) to control for national trends affecting all states simultaneously (e.g., federal policies, economic cycles).

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

Results

This table condensates all of our regression results for property crime :

	Model 1	Model 2	Model 3	Model 4	Model 5
Log GDP	−1509.732***			−244.226***	−104.716
	(169.415)			(69.330)	(173.391)
Inequality (top 10%)	−421.245	−670.438	−2516.508*	−384.190	−155.350
	(716.024)	(682.399)	(1207.310)	(275.499)	(213.868)
Unemployment rate	−8.948	−9.638	2.247	4.756	1.026
	(7.571)	(8.151)	(9.708)	(2.979)	(8.524)
Prison per 100k	0.997	0.313	0.258	0.069	−0.113
	(0.611)	(0.694)	(0.639)	(0.206)	(0.212)
Beer consumption	1833.025***	1908.015***	1595.683**	454.270**	226.185
	(453.305)	(407.773)	(524.056)	(144.527)	(139.668)
percent_poor	−0.142	0.665	−5.603	−4.088	−2.941
	(7.219)	(7.655)	(8.386)	(3.569)	(3.903)
Consumption		−0.061***			
		(0.006)			
log_pop			−4954.541***		
			(759.530)		
Property crime (t-1)				0.795***	0.811***
				(0.023)	(0.021)
Num.Obs.	1050	1050	1050	1000	1000
R2	0.761	0.752	0.681	0.917	0.716

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

Robust standard errors clustered by state are in parenthesis

This table presents the results of our 5 models for property crime, using robust standard errors clustered by state to account for serial correlation and heteroscedasticity.

The variable `log_gdp` shows a strong and consistent negative correlation with property crime. For instance, in the baseline model, a 10% increase in per state GDP is associated to about 151 fewer property crimes per 100 000 inhabitants ($\beta = -1509$, $p < 0.001$). This effect is attenuated but remains significant when controlling for crime persistence in Model 4 ($\beta = -244$, $p < 0.001$), and becomes non-significant in the two-way fixed effects specification in Model 5, $\beta = -105$, $p > 0.05$. This suggests that the crime-GDP relationship reflects national trends rather than state-specific dynamics.

Personal consumption expenditures, tested as an alternative measure to GDP in Model 2, shows a significant but smaller negative correlation ($\beta = -0.061$, $p < 0.001$). This confirms that household living standards, not just aggregate output, are negatively associated with property crime.

Neither unemployment rate nor poverty rate shows a significant correlation with property crime after controlling for other factors. In a similar fashion, the prison population per 100

000 inhabitants is not significantly correlated to property crime. These null findings contrasts with theoretical expectations. For unemployment, it is not consistent with the literature that shows that, during the 1990s, unemployment dropping caused, in part, an decrease in property crimes (Raphael & Winter-Ebmer, 2001).

Beer consumption per capita is positively and significantly correlated with property crime across all specifications. In model 1, an increase of 1 gallons per capita is associated with approximately 1833 additional property crimes per 100,000 inhabitants ($p < 0.001$). The attenuation in Model 5 could be linked to the absorption of common trends by year fixed effects.

The lagged crime rate is the strongest predictor in dynamic specifications : in Model 4, a one-unit increase in the previous year's property crime rate is associated with a 0.79-unit increase in the current rate ($\theta = 0.795$, $\theta < 0.001$). This coefficient remains stable in Model 5 : a one-unit increase in the previous year's property crime rate is associated with a 0.81-unit increase in the current rate ($\theta = 0.811$, $p < 0.001$), confirming substantial crime inertia.

The baseline model explains 76% of within state variation ($R^2 = 0.761$). Adding lagged crime in models 4 and 5 increases R^2 up to 91%, indicating that past crime is a powerful predictor of current crime.

Thus, for property crime, GDP and consumption are negatively correlated with property crime, which partially confirms our first hypothesis. Alcohol consumption is positively correlated with property crime, supporting H_4. Finally, the crime persistence hypothesis (H_5) is strongly supported by our results.

This second table condensates our results for violent crimes :

	Model 1	Model 2	Model 3	Model 4	Model 5
Log GDP	−34.680			8.430	−11.871
	(27.265)			(5.687)	(14.819)
Inequality (top 10%)	−194.683	−187.984	−241.940	15.952	7.167
	(140.576)	(144.638)	(131.145)	(40.477)	(43.797)
Unemployment rate	−9.234***	−9.210***	−8.969***	−2.792***	0.049
	(2.131)	(2.156)	(1.994)	(0.629)	(0.974)
Prison per 100k	0.333*	0.318*	0.316*	0.079	0.043
	(0.135)	(0.134)	(0.135)	(0.043)	(0.041)
Beer consumption	224.109**	221.493***	218.109***	17.271	7.407
	(69.309)	(67.062)	(63.285)	(29.765)	(28.039)
percent_poor	−0.717	−0.683	−0.843	−0.939	−1.302*
	(1.813)	(1.813)	(1.813)	(0.754)	(0.658)
Consumption		−0.002			
		(0.001)			
log_pop			−115.069		
			(112.220)		
Violent crime (t-1)				0.833***	0.877***
				(0.031)	(0.032)
Num.Obs.	1050	1050	1050	1000	1000
R2	0.315	0.317	0.312	0.828	0.812

* p < 0.05, ** p < 0.01, *** p < 0.001

Robust standard errors clustered by state are in parenthesis

The first three models fail to significantly predict violent crime rates as they explain only about 31% of within-state variation ($R^2 = 0.31$), which is substantially less than for property crime ($R^2 \approx 0.76$). This might indicate that violent offences are driven by factors beyond economic conditions alone.

Unlike property crime, `log_gdp` shows no significant correlation with violent crime. Personal consumption is also non-significant. These findings imply that violent crimes are less motivated by economic need than property crimes, and might be more likely to be driven by conflicts or social disorganization.

Interestingly, unemployment rate is significantly and negatively correlated with violent crime in models 1-4 ($p < 0.001$). A 1 percentage point increase in unemployment rate leads to a decrease of around 9 violent crimes per 100 000 inhabitants for models 1-3 (representing about 2.4% of the average violent crime rate), and to a 2.8 decrease in model 4. This counterintuitive finding may reflect omitted variable bias, and should be interpreted with caution. The effect disappears in Model 5 after controlling for year fixed effects, suggesting it may capture only national time trends and not a structural relationship.

The top 10% income share is not significantly correlated with violent crime ($p > 0.05$ in all models), contrary to theoretical predictions linking inequality to social tension and violence.

Prison population per 100,000 inhabitants shows a small but positive and significant correlation with violent crime (3.2, $p < 0.05$ in Models 1-3). This might not imply that incarceration fail to decrease crime, but could reveal reverse causality issues, as states with higher violent crime rates incarcerate more people.

These findings lead us to reject hypothesis H3 (deterrence). The positive correlation between incarceration and crime does not support the deterrence hypothesis in our panel setting, likely due to reverse causality (states with more crime incarcerate more) overwhelming any potential deterrent effect.

As for beer consumption, it is again strongly and positively correlated with violent crime.

It's interesting to notice that adding lagged crime dramatically improves the models' fit ($R^2 \approx$ 0.8 in models 4 and 5). The crime persistence coefficient is large and highly significant ($\theta = 0.833$). It indicates that violent crime exhibits even stronger inertia than property crime. Indeed, holding other factors constant, a state with 100 more violent crimes per 100 000 the past year will have 83 more violent crime the following year. But in these models, almost all variables that previously were significant lose their significance. This might indicate that their effect was linked to crime inertia. In the two-way fixed effect model (Model 5), only poverty remains slightly significant aside from lagged crime, but it has a negative coefficient.

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

Descriptive analysis

Descriptive overview of crime and new covariates

We start the short-run analysis by describing crime rates and contemporaneous socio-economic variables over the period 2015–2018, which corresponds to the years for which the richest set of covariates is jointly available. Given the short time dimension, the objective is to document cross-state variation rather than long-run trends.

Table 10: Summary statistics, short-run panel (2015–2018)

variable	N	Mean	SD	P10	Median	P90
Violent crime rate	204	385.38	179.59	217.87	360.90	555.60
Property crime rate	204	2419.80	672.01	1573.97	2428.65	3249.19
Log GDP	204	12.30	1.04	10.92	12.29	13.56
Top 10% pretax income share	204	0.46	0.05	0.41	0.46	0.53
Lawful permanent residents per million	204	2480.04	1553.79	952.88	2043.32	4497.99

Police shootings per million	201	3.57	2.23	1.39	3.01	6.63
Public spending: education (in millions)	102	20.24	24.73	3.07	13.50	39.36
Public spending: welfare (in millions)	102	13.56	19.21	2.41	7.99	28.64
Public spending: health (in millions)	102	2.06	3.70	0.27	0.87	4.33
Public spending: corrections (in millions)	102	1.58	2.37	0.23	0.79	2.75
Public spending: police (in millions)	102	2.28	3.28	0.33	1.30	3.87
WIC participants (2015)	204	128.33	149.49	14.00	98.00	247.00
School lunch participants (2015)	204	307.67	259.14	1.00	280.00	676.00
School breakfast participants (2015)	204	202.71	177.15	24.00	180.00	423.00
Unemployment rate (%)	204	4.40	1.08	3.00	4.38	6.01
Poverty rate (%)	204	12.25	3.04	8.70	11.75	16.20
Prison population per 100k	200	376.72	140.15	203.96	369.07	572.60
Beer consumption per capita	204	1.13	0.22	0.90	1.11	1.40

During this period, violent crime averages 385 incidents per 100,000 inhabitants, while property crime is substantially more prevalent, with an average of 2,420 incidents per 100,000 inhabitants. Both indicators display considerable dispersion across states, especially property crime, highlighting strong heterogeneity in crime exposure even in recent years.

Economic conditions vary moderately across states. Log GDP shows limited dispersion, while income concentration, measured by the top 10% pre-tax income share, averages 46%, with meaningful cross-state variation. Labour market conditions appear relatively favourable on average, with unemployment rates around 4.4%, but poverty remains persistent, affecting roughly 12% of the population.

Migration intensity and police-related variables also differ markedly across states. States host on average 2,480 lawful permanent residents per million inhabitants, while police shootings occur at a rate of about 3.6 per million, with substantial dispersion. These differences motivate their inclusion as potential correlates of crime.

Public spending and social programme indicators (education, welfare, health, corrections, police, and food assistance) exhibit large variability but are available only for a subset of observations, reflecting the short coverage of these datasets. Institutional and behavioural factors further show heterogeneity, with prison populations averaging 377 inmates per 100,000 inhabitants and beer consumption per capita displaying moderate cross-state variation.

Contemporary heterogeneity and structural correlates (2015–2018)

To further explore cross-state heterogeneity in the short run, we group state–year observations into quartiles of violent crime intensity and compare average socio-economic and institutional characteristics across these groups. Given the limited time dimension, this exercise is purely descriptive and aims to highlight contemporaneous correlations rather than dynamic effects.

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	

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	

States in the highest violent crime quartile experience, on average, nearly three times the violent crime rate of those in the lowest quartile, and also display substantially higher property crime rates. These high-crime states are characterised by worse labour market and social conditions, with unemployment rates exceeding 5% and poverty rates above 14%, compared to about 3.9% unemployment and 10.5% poverty in low-crime states.

Institutional and behavioural indicators also differ markedly across quartiles. Prison populations increase monotonically with violent crime intensity, rising from about 293 inmates per 100,000 inhabitants in low-crime states to over 470 per 100,000 in high-crime states. Police shootings per million inhabitants are almost twice as high in the highest violent crime quartile, suggesting higher levels of police–citizen confrontation or enforcement intensity in these states.

A similar pattern emerges when states are grouped by property crime quartiles. High property crime states combine higher violent crime rates, higher poverty, larger prison populations, and substantially more police shootings per capita. In contrast, differences in beer consumption per capita are relatively small across quartiles, suggesting that alcohol consumption varies less systematically with crime intensity in the short run.

Public spending and food assistance indicators show considerable dispersion but no clear monotonic pattern across quartiles, reflecting both institutional heterogeneity and the limited availability of these variables in the short-run sample.
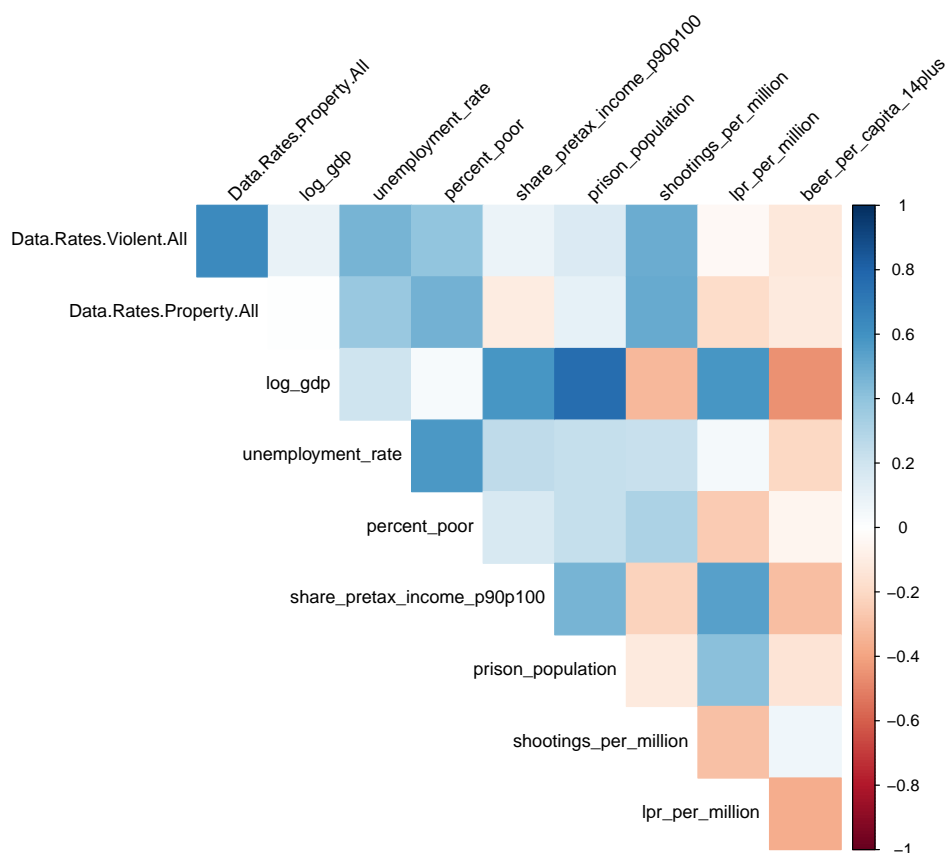
Overall, this quartile-based comparison indicates that, even over a short horizon, states with higher crime rates tend to be characterised by greater economic stress, higher incarceration rates, and more intense police activity. These patterns motivate the regression analysis that

follows, which formalises these associations while controlling for multiple dimensions simultaneously.

Regression analysis

Because the short-run panel only covers four years (2015–2018), we begin with a purely exploratory correlation analysis. The objective is not statistical inference, but rather to provide a preliminary overview of how violent and property crime co-move with a broad set of socio-economic and institutional variables, and to motivate a parsimonious and structured regression design.

Using the short-run panel, we compute pairwise correlations between crime rates and selected economic, social, and institutional indicators, including GDP, income inequality, unemployment, poverty, prison population, police activity, public spending, immigration, and alcohol consumption. The resulting correlation matrix is summarized using a heatmap.

The correlation matrix reveals several clear short-run patterns. Both violent and property crime are positively correlated with unemployment and poverty, with stronger associations for property crime, suggesting a closer link with short-run economic conditions. Crime rates are also positively correlated with police shootings per million inhabitants, indicating that high-crime states tend to experience more intense or confrontational policing environments.

In contrast, log GDP shows only a weak and slightly negative correlation with crime, while income inequality displays modest correlations, suggesting that these variables are not strong short-run predictors of crime on their own. Prison population is positively correlated with both crime and police shootings, consistent with institutional responses to higher crime rather than pure deterrence effects.

Overall, the correlation structure points to a distinction between socio-economic conditions (unemployment and poverty) and institutional intensity (incarceration and policing), motivating a structured and parsimonious regression strategy.

Hypotheses

We focus on four short-run hypotheses, mapped directly into two model blocks:

- **H1 (economic stress)**: States experiencing worse economic conditions—higher unemployment and poverty—exhibit higher crime rates, especially property crime.
- **H2 (inequality)**: Higher income concentration at the top of the distribution is associated with higher crime rates.
- **H3 (institutional stress)**: Higher incarceration levels and policing activity are associated with different crime patterns, reflecting enforcement intensity rather than deterrence per se.
- **H4 (behavioral factors)**: Higher alcohol consumption is positively associated with violent crime.

(H1–H2 are tested in the economic stress model; H3–H4 in the institutional model.)

Economic stress model (pooled OLS + year FE)

Given the short time dimension, we use a parsimonious pooled OLS specification and include year fixed effects to absorb nationwide shocks common to all states.

$$
\begin{aligned}
Crime_{it} = \ & \beta_0 + \beta_1 log(GDP_{it}) + \beta_2\, Unemployment_{it} \\
& + \beta_3\, Poverty_{it} + \beta_4\, Inequality_{it} \\
& + \beta_5\, log(Pop_{it}) + \delta_t + \varepsilon_{it}
\end{aligned}
$$

where: $Crime_{it}$ denotes either violent or property crime in state i and year t, δ_t are year fixed effects capturing nationwide shocks.

This specification captures short-run associations between economic stress and crime without conditioning on potentially endogenous institutional responses.

```
Call:
lm(formula = Data.Rates.Property.All ~ log_gdp + unemployment_rate +
    percent_poor + share_pretax_income_p90p100 + log_pop + factor(Year),
    data = panel_short_inst)

Residuals:
     Min       1Q    Median       3Q      Max
-1250.88  -434.24    33.47   434.08   963.92

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                    865.55    1407.73   0.615  0.54018
log_gdp                        -55.67     384.21  -0.145  0.88511
unemployment_rate              106.62      83.40   1.278  0.20434
percent_poor                    84.04      28.47   2.952  0.00402 **
share_pretax_income_p90p100  -3079.69    1246.70  -2.470  0.01536 *
log_pop                        143.44     397.70   0.361  0.71918
factor(Year)2018               -82.95     112.72  -0.736  0.46369
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 536.4 on 91 degrees of freedom
Multiple R-squared:  0.2911,    Adjusted R-squared:  0.2444
F-statistic: 6.229 on 6 and 91 DF,  p-value: 1.648e-05



Call:
lm(formula = Data.Rates.Violent.All ~ log_gdp + unemployment_rate +
    percent_poor + share_pretax_income_p90p100 + log_pop + factor(Year),
    data = panel_short_inst)

Residuals:
    Min      1Q  Median      3Q     Max
-298.49  -83.32   -1.83   90.42  357.88

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   147.429    351.026   0.420  0.67548
```

```
log_gdp                          63.301     95.806   0.661   0.51046
unemployment_rate                69.697     20.796   3.352   0.00117 **
percent_poor                     13.622      7.099   1.919   0.05815 .
share_pretax_income_p90p100    -378.076    310.872  -1.216   0.22706
log_pop                         -54.328     99.168  -0.548   0.58514
factor(Year)2018                 20.165     28.108   0.717   0.47495
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 133.8 on 91 degrees of freedom
Multiple R-squared:  0.2936,    Adjusted R-squared:  0.2471
F-statistic: 6.305 on 6 and 91 DF,  p-value: 1.427e-05
```

Institutions stress model (pooled OLS + year FE)

We estimate a separate specification focusing on contemporaneous institutional intensity and behavioral factors. These covariates are interpreted as descriptive correlates (not exogenous policy instruments), hence we keep them separate from economic stress variables.

$$
\begin{aligned}
Crime_{it} = \ & \beta_0 + \beta_1\ Prison_{it} + \beta_2\ PoliceShootings_{it} \\
& + \beta_3\ PoliceSpending_{it} + \beta_4\ CorrectionsSpending_{it} \\
& + \beta_5\ Alcohol_{it} + \beta_6\ log(Pop_{it}) + \delta_t + \varepsilon_{it}
\end{aligned}
$$

This model does not aim to measure deterrence effects, but rather documents how crime co-varies with enforcement intensity and social regulation in the short run.

```
Call:
lm(formula = Data.Rates.Property.All ~ prison_population + nb_police_shootings +
    spending_police + spending_corrections + beer_per_capita_14plus +
    log_pop + factor(Year), data = panel_short_inst)

Residuals:
    Min      1Q  Median      3Q     Max
-937.29 -387.62  -63.45  287.98 1276.63

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.948e+03  1.706e+03   1.728   0.0874 .
prison_population   -1.645e-03  4.379e-03  -0.376   0.7081
nb_police_shootings  2.727e+01  5.158e+00   5.287 8.64e-07 ***
spending_police      1.580e-05  8.723e-05   0.181   0.8566
```

```
spending_corrections    -2.415e-04  1.236e-04  -1.953   0.0539 .
beer_per_capita_14plus -4.265e+02  2.943e+02  -1.449   0.1509
log_pop                 -1.597e+01  1.036e+02  -0.154   0.8778
factor(Year)2018        -1.816e+02  1.070e+02  -1.697   0.0932 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 528.7 on 90 degrees of freedom
Multiple R-squared:  0.3189,    Adjusted R-squared:  0.2659
F-statistic: 6.019 on 7 and 90 DF,  p-value: 9.071e-06




Call:
lm(formula = Data.Rates.Violent.All ~ prison_population + nb_police_shootings +
    spending_police + spending_corrections + beer_per_capita_14plus +
    log_pop + factor(Year), data = panel_short_inst)

Residuals:
    Min      1Q  Median      3Q     Max
-238.97 -113.68  -28.77   84.11  519.96

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             6.670e+02  4.885e+02   1.365   0.1755
prison_population       3.774e-05  1.254e-03   0.030   0.9761
nb_police_shootings     3.664e+00  1.477e+00   2.481   0.0149 *
spending_police         1.975e-05  2.497e-05   0.791   0.4312
spending_corrections   -4.935e-05  3.540e-05  -1.394   0.1667
beer_per_capita_14plus -9.334e+01  8.427e+01  -1.108   0.2710
log_pop                -1.474e+01  2.965e+01  -0.497   0.6203
factor(Year)2018       -1.591e+01  3.064e+01  -0.519   0.6049
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 151.4 on 90 degrees of freedom
Multiple R-squared:  0.1053,    Adjusted R-squared:  0.03566
F-statistic: 1.512 on 7 and 90 DF,  p-value: 0.1731
```

These results should therefore be interpreted as descriptive associations rather than causal effects.

Robustness: Two-way fixed effects (TWFE) for the economic stress model

As a robustness check, we also estimate a two-way fixed effects version of the economic stress model (state FE + year FE). With only four years, this specification is demanding and is used mainly to assess the robustness of the pooled OLS associations to unobserved state-level heterogeneity.

```
Twoways effects Within Model

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

Unbalanced Panel: n = 50, T = 1-2, N = 98

Residuals:
    Min.  1st Qu.   Median  3rd Qu.     Max.
-127.901  -26.545    0.000   26.545  127.901

Coefficients:
                              Estimate Std. Error t-value Pr(>|t|)
log_gdp                      -425.8314 1197.8196 -0.3555  0.72399
unemployment_rate             129.2804   59.9948  2.1549  0.03695 *
percent_poor                   -3.7092   10.2212 -0.3629  0.71850
share_pretax_income_p90p100   378.5470 1083.8363  0.3493  0.72863
log_pop                     -2612.6020 2363.7286 -1.1053  0.27533
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:     271220
Residual Sum of Squares: 235310
R-Squared:      0.13241
Adj. R-Squared: -1.0037
F-statistic: 1.28197 on 5 and 42 DF, p-value: 0.28974
```

The short-run regressions provide limited and fragile explanatory power, reflecting the very short time dimension of the panel. The pooled OLS models explain at most 29–32% of the variation in property crime ($R^2 \approx 0.29$–0.32) and around 29% for violent crime, indicating that a large share of cross-state crime differences remains unexplained.

For property crime, poverty and income inequality are the only consistently significant covariates. Poverty is positively associated with property crime ($\beta \approx 84$, $p = 0.004$), and the top 10% income share is also significant ($p = 0.015$). GDP, unemployment, and population size are not statistically significant. These results suggest correlation rather than a robust structural relationship.

For violent crime, results are even less stable. Unemployment is significant ($\beta \approx 70$, $p = 0.001$), while poverty is only marginally significant ($p \approx 0.06$) and inequality is not significant. Model fit remains modest ($R^2 \approx 0.29$), pointing to weak short-run explanatory content.

Institutional models perform slightly better for property crime ($R^2 \approx 0.32$) but very poorly for violent crime ($R^2 \approx 0.11$). Police shootings per capita are the only robust institutional correlate ($p < 0.05$), while incarceration and spending variables are not significant, likely reflecting reverse causality and multicollinearity.

The two-way fixed effects model further weakens results ($R^2 \approx 0.13$), confirming that short-run within-state variation is insufficient to identify stable relationships. Overall, these findings should be interpreted as descriptive correlations rather than reliable estimates of causal effects.

Conclusion and limitations

Limitations

Several limitations should be kept in mind when interpreting the results.

The long run analysis focuses on 1998–2018, a period of sustained crime decline across U.S. states. The relationships we document may not generalize to periods of rising crime in the United States. Despite the use of fixed effects, the estimated relationships remain associative rather than causal. Key variables such as incarceration, unemployment, and alcohol consumption may be jointly determined with crime. We lack credible instruments or natural experiments to establish causal effects. The positive correlation between prison population and crime, for instance, likely reflects reverse causality rather than failed deterrence. Data constraints exclude social capital, educational quality, and drug markets, potentially biasing estimated coefficients.

As for the short-run analysis, it is constrained by a very limited time dimension. This restricts statistical power and makes highly saturated specifications, such as two-way fixed effects models, imprecise. As a result, short-run regression results should be interpreted as descriptive correlations rather than structural effects. Several institutional variables included in the short-run models—police shootings, public spending, and prison population are likely endogenous responses to crime levels. Their coefficients therefore capture contemporaneous co-movement rather than deterrence or policy effects.

Finally, crime exhibits strong persistence over time, especially for violent crime. Once lagged crime is included, most contemporaneous covariates lose significance, indicating that short-run variations are dominated by inertia. This limits the ability of static models to identify independent short-term drivers of crime.

Conclusion

This project examines how crime in U.S. states co-evolves with economic and social conditions, and whether these relationships differ between property and violent crime.

The empirical evidence reveals a clear distinction between the two crime types. Property crime is systematically associated with economic conditions, particularly poverty and, to a lesser extent, unemployment. These variables are statistically significant in multiple specifications and the corresponding models explain a substantial share of within-state variation. This pattern is consistent with economic theories of crime, according to which material deprivation increases incentives for financially motivated offences.

In contrast, violent crime shows a much weaker connection to economic variables. GDP, inequality and consumption are generally insignificant, and explanatory power remains limited in static models. Once lagged crime is introduced, model fit increases sharply, with a large and highly significant persistence coefficient, indicating that violent crime is driven primarily by inertia rather than contemporaneous economic conditions.

Institutional variables such as incarceration rates and police activity are positively correlated with crime, but their interpretation is ambiguous and likely reflects policy responses to crime rather than causal deterrence effects.

Hence, the results show that economic and social conditions matter mainly for property crime, while violent crime is characterised by strong persistence and limited sensitivity to short-run economic fluctuations, providing a coherent and robust answer to the research question.