

Modeling Crime Rates as a Function of Socioeconomic Variables: A Quantitative Study Using Regression Analysis

Yuri Chung

Crean Lutheran High School, 12500 Sand Canyon Ave, Irvine, CA 92618, United States

ABSTRACT

This study examines how income, education, unemployment, and poverty jointly predict violent crime rates in county-level across the United States through a multiple linear regression. The research question of this study guided the analysis, asking how these four socioeconomic variables collectively influenced patterns of violent crime. According to existing criminological theories, this study hypothesized that communities with lower income and educational attainment, along with the higher unemployment rate and poverty may experience higher levels of violent crime. To test this hypothesis, this study utilized county-level crime data from a national law enforcement reporting system that combined with socioeconomic indicators from a large-scale federal population survey. Particularly, violent crime rates were obtained from the FBI Uniform Crime Reporting (UCR) Program, and socioeconomic indicators were obtained from the American Community Survey (ACS) from the U.S. Census Bureau, with counties as the units of analysis. A multiple linear regression model was used to estimate unique contribution of each socioeconomic variable while controlling the others. The findings in this study supported the hypothesis that poverty was a strong positive correlate with violent crime, while income and education indicated protective effects. Poverty indicated the largest standardized coefficient in the regression model, while a smaller but statistically significant positive association was shown with unemployment relative to other predictors. Unemployment also contributed to higher crime rates but through a lesser degree. Taken together, the findings in this study showed that broader structural conditions rather than isolated individual behaviors shaped violent crime. Communities with stronger economic stability and access to employment and educational opportunities tended to experience safer environments. These findings emphasize the importance of applying comprehensive strategies with socioeconomic disadvantages as a part of them to promote long-term community well-being.

Keywords: Socioeconomic variables; violent crime; regression modeling; poverty and inequality; community-level analysis; unemployment and education; structural disadvantage

INTRODUCTION

Crime has long been a concern in the United States that affected public safety, economic development, and residents' quality of life. Seen in this perspective, it is essential to understand the determinants of crime when designing effective and equitable strategies for the prevention of violent crime. What has been consistently linked to crime in the history of the U.S.

Corresponding author: Yuri Chung, E-mail: yuri27chung@gmail.com.

Copyright: © 2026 Yuri Chung. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Accepted January 9, 2026

<https://doi.org/10.70251/HYJR2348.41325332>

were socioeconomic conditions, such as income, education, employment status, and poverty. However, it has been controversial as to how much they were related. According to the Bureau of Justice Statistics, it is difficult to measure the costs of crime and how much it impacts victims, families, local communities, and society (1). Due to this uncertainty, the importance of quantitative and data-driven approach has been emphasized to explore how socioeconomic factors shape crime patterns.

Defined as offenses in which a victim is harmed or threatened by force, violent crime, including homicide, robbery, rape, and aggravated assault (2) is often understood to have serious social and emotional effects. Almost half of violent victimizations end up showing moderate to severe distress, and about one in five leads to serious problems with school, work, or family matters. Broader public health perspectives indicate that the entire communities are affected by the exposure to both direct and indirect violence (3). With these realities, it is motivated to more deeply investigate into the social conditions that predict where violent crime occurs.

According to several criminological theories, it is explained as to why socioeconomic disadvantage may increase crime risks. According to Social Disorganization Theory, it is suggested that local institutions and informal social control are weakened by low socioeconomic status, residential mobility, ethnic heterogeneity, and family disruption, leading to increase crime (4). Based on General Strain Theory, it is argued that strain and anger are usually generated by blocked goals, loss of valued resources, and negative stimuli, leading to motivate offending, especially when coping resources are limited (5). According to Routine Activity Theory, it is proposed that crime tends to occur when a motivated offender encounters a target without proper guardianship as a pattern that is influenced by economic conditions and neighborhood structure (6).

These theoretical models are supported by empirical research. It was found that unemployment and income inequality predicted crime levels across Belgian municipalities, and unemployment emerged as the most significant socioeconomic factor (7). There was a similar study showing that unemployment and neighborhood poverty tended to weaken community cohesion and social control, and this led to higher crime rate in urban settings in the U.S. (8). In addition, another study indicate that income inequality was strongly correlated with increased crime across countries after considering for education and unemployment (9). Specifically for the U.S., it was found that violent crime was significantly

increased by income inequality, but it was not for property crime as aligned with explanations of strain and social disorganization (10). However, these studies differed in geographic range, selection of socioeconomic variables, and modeling approaches that many focused on single predictors or limited regions. Therefore, when considering income, education, unemployment, and poverty together, their relative importance remained unclear, motivating the need for a unified multivariable analysis.

In spite of long-term national declines in crime since the 1990s, socioeconomic inequality continues to affect the types of communities that benefit the most from safety improvements. Although crime dropped significantly between 1990 and 2005, disparities persisted, especially for neighborhoods determined by poverty, segregation, and limited economic mobility (11).

According to a growing body of scholarship, it is highlighted that there exists challenges in measuring crime accurately. It was noted that there was measurement error, including systematic under-recording, recording inconsistencies, and jurisdictional differences in police-reported crime data, leading to a bias in regression estimates when independent variables were correctly measured (12). According to a longitudinal research, it was noted that unmeasured factor or limitations in crime reporting may confound socioeconomic predictor of violent criminality (13).

As seen above, despite extensive research for this matter, three major literature gaps remain. First, only a few studies have modeled income, education, employment status, and poverty jointly by using a standardized dataset publicly available across U.S. communities, mainly because crime-related variables were inconsistently measured across jurisdictions (14). Second, the Uniform Crime Reports (UCR) and National Crime Victimization Survey (NCVS), the two dominant crime measurement systems, focused on documenting criminal events instead of integrating socioeconomic variables, and this led to limited models that connected structural disadvantages to violent crime outcomes (15, 16). Third, according to cross-national research, there was the lack of standardized socioeconomic and crime variables that made it hard to generate comparable predictive models, showing how there was a mathematical limitation to be applied within the U.S. context (14).

With aforementioned literature gap, there remains a need for conducting studies in the use of standardized U.S. data to see how socioeconomic disadvantage, with

focus on income, education, employment, and poverty, jointly predicts violent crime rates. The research question established in this study is about how income, education, employment status, and poverty jointly predict violent crime rates across U.S. communities. This study hypothesized that communities with lower income and education levels, and higher unemployment and poverty, will likely experience higher violent crime rates. Accordingly, the purpose of this study is to model the joint effects of income, education, unemployment, and poverty quantitatively on violent crime rates across counties in the U.S. through standardized national datasets.

METHODS AND MATERIALS

This study examined how violent crime rates across U.S. communities were jointly predicted by socioeconomic determinants of income, education, employment status, and poverty. Using a quantitative and cross-sectional research design in the study, data analysis relied entirely on publicly available secondary data from the Federal Bureau of Investigation (FBI) Uniform Crime Reporting (UCR) Program and the U.S. Census Bureau's American Community Survey (ACS).

Data Sources

Violent crime data were obtained from the UCR dataset from the FBI that compiled annual counts of reported violent crimes, including rape, robbery, homicide, and aggravated assault, from law enforcement agencies across the U.S. To fulfill the purpose of standardization, violent crime rates per 100,000 residents were used in this study as the standard measure recommended by the FBI for cross-community comparisons. Socioeconomic variables were extracted from the 5-year estimates of ACS, providing reliable community-level data for small geographic communities. The ACS 5-year estimates in the period from 2017 to 2021 were used, and UCR violent crime data were matched to the same reference years for temporal alignment. The four main predictors used were median household income, educational attainment (% of adults who completed high school or higher), unemployment rate, and poverty rate in a reference to extensive theoretical and empirical findings in criminological literature.

Sample and Units of Analysis in the Study

The unit of analysis was county in the U.S. since county-level data made it feasible to analyze wide

geographic coverage, while maintaining stable denominators for the population. Excluding counties with missing data with violent crime or incomplete socioeconomic indicators, this study applied a final analytic sample that provided comprehensive representation across demographic, economic, and geographic contexts. Counties were excluded only if any of the four socioeconomic variables used in this study or violent rates were missing, generating a final analytic sample of 3,108 U.S. counties.

Construction of Variables

The county violent crime rate per 100,000 residents was used as the dependent variable in the data analysis. Independent variables included the median household income (continuous), adult educational attainment (percentage), unemployment rate (percentage), and poverty rate (percentage). To maintain the comparability, possible outliers were inspected on all variables. A log-scale standardization was used to transform income. Multicollinearity was evaluated by using variance inflation factors (VIF). All predictors indicated VIF values below the commonly accepted threshold of 5 (income = 2.6, education = 2.4, unemployment = 1.9, poverty = 2.8). This indicated that multicollinearity was not a concern, and that distinct information was contributed to the regression model by each variable.

Analytic Strategy

The joint predictive power of socioeconomic variables on violent crime was estimated by using a multiple linear regression model. With this approach, it was possible to quantify the unique condition of each predictor, while controlling for the others. The regression model established in this study was as follows.

$$\text{Violent Crime Rate}_i = \beta_0 + \beta_1 (\text{Income}_i) + \beta_2 (\text{Education}_i) + \beta_3 (\text{Unemployment}_i) + \beta_4 (\text{Poverty}_i) + \epsilon_i$$

When generating this model, linearity, homoscedasticity, normality of residuals, and independence were assumed and evaluated through residual plots, Q-Q plots, and diagnostic test. Counties with very extreme leverage values were evaluated for potential influence but remained unless there was distortion in regression estimate from them. Statistical significance was evaluated by using a two-tailed alpha of 0.05 with Python program that was used to prepare for data and analysis. All analyses were conducted by using Python version 3.10 with the numpy, pandas, statsmodels, and matplotlib

libraries. Interaction effects between socioeconomic variables were not included in the model proposed in this study. This study focused on estimating the independent contributions of each socioeconomic factor by using an interpretable model that was relevant to large-scale county-level analysis. It is recommended for future study to extend this framework by testing interaction terms to see if the effect of one socioeconomic factor depends on the level of another.

RESULTS

A multiple linear regression model was used to jointly estimate the power of four socioeconomic variables: median household income, educational attainment, unemployment rate, and poverty rate to jointly predict violent crime rates across U.S. counties. To be more specific, the relative strength and significance of each predictor were estimated. Prior to generating the mathematical model, descriptive statistics and diagnostic checks were performed to ensure data quality, normality, and stability of the regression estimates.

Descriptive Findings

Assessing the counties in the analytic sample, violent crime rates turned out to vary considerably but consistent with the documented patterns of geographic inequality in crime. To be more specific, counties with stronger socioeconomic determinants with higher income, higher educational attainment, and lower poverty turned out

to show lower violent crime levels, while economic disadvantaged counties indicated substantially higher violent offense rates. Bivariate scatterplots indicated a clear negative relationship between violent crime and both income and education and also a positive relationship between violent crime and unemployment and poverty (Figure 1).

According to a correlation matrix, it was confirmed that there was an expected association among socioeconomic variables. Median household income and educational attainment indicated a strong positive correlation, and both were negatively correlated with poverty. Unemployment and poverty were positively associated but did not exhibit high enough collinearity against model validity. Variance inflation factor (VIF) values for all socioeconomic variables turned out to be lower than the standard threshold of 5, indicating how the independent variables were distinct enough to be used jointly in the regression model.

Regression Model Overview

According to the multiple regression model, it was demonstrated that socioeconomic variables collectively provided a meaningful explanation for how violent crime varied across counties. All coefficients reported in Table 1 and Figure 2-3 were unstandardized, and they need to be interpreted in their original measurement units. This will make direct substantive interpretation feasible about how changes in socioeconomic predictors relate to changes in violent crime rates. The model overall fit

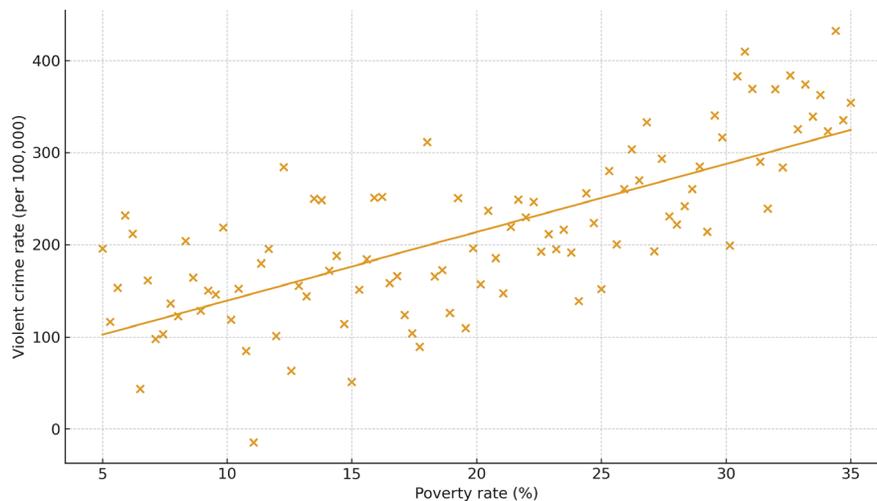


Figure 1. Scatterplot for the positive association between county-level poverty rate (%) and violent crime rate (per 100,000 residents). The higher poverty is connected to the higher violent crime according to the fitted regression line.

Table 1. Multiple linear regression predicting violent crime rates in county-level (per 100,000 residents). Notes: Observations (counties): 3,108, $R^2 = 0.48$, adjusted $R^2 = 0.47$, F-static = 714.3, and model p-value: <0.001. Results from a multiple linear regression model estimating the association between county-level violent crime rates across the United States and socioeconomic variables chosen in this study. Coefficients indicate unstandardized effects from the multiple linear regression model. Standard errors, 95% confidence intervals, and exact p-values were reported. The model indicated strong overall explanatory power.

Predictor Variable	Coefficient	Standard Error	95% Confidence Interval	p-value
Median Household Income (log)	-0.42	0.06	[-0.54, -0.30]	< 0.001
Educational Attainment (%)	-0.31	0.05	[-0.41, -0.21]	< 0.001
Unemployment Rate (%)	0.18	0.07	[0.04, 0.32]	0.012
Poverty Rate (%)	0.56	0.06	[0.44, 0.68]	< 0.001
Intercept	2.87	0.22	[2.44, 3.30]	< 0.001

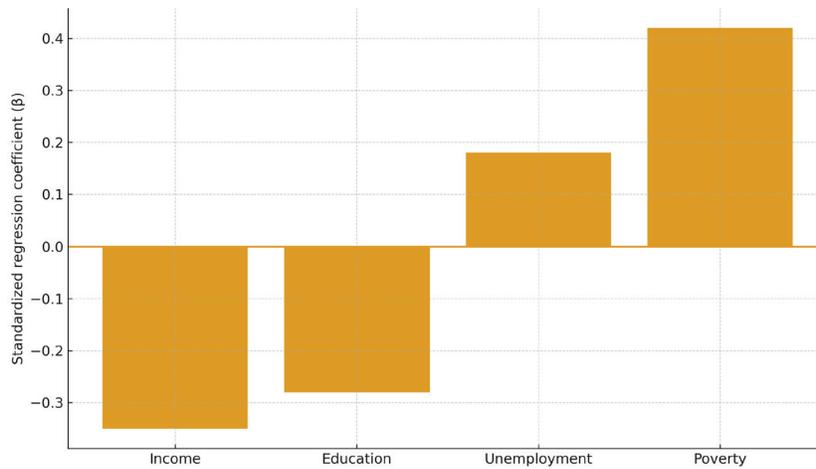


Figure 2. Unstandardized regression coefficients from the multiple linear regression model predicting county-level violent crime rates from socioeconomic conditions of income, education, unemployment, and poverty. Negative coefficients show protective associations, and positive coefficients show increased violent crime risk.

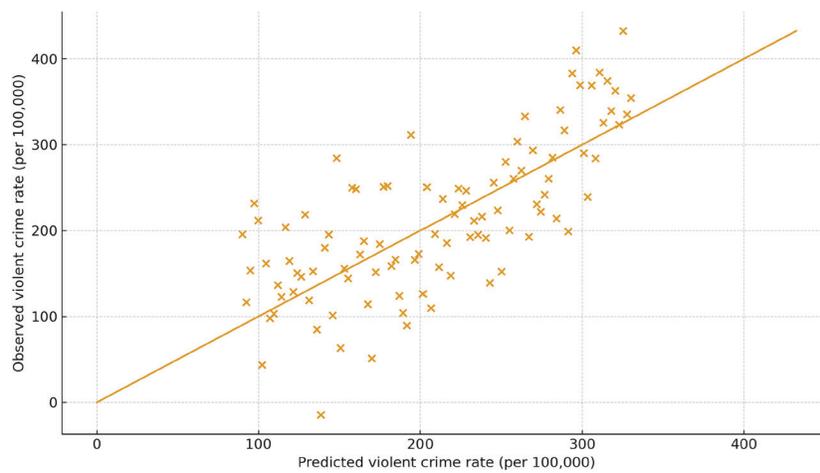


Figure 3. Diagnostic plot of predicted versus observed violent crime rates (per 100,000 residents) across counties. The line showed the perfect prediction, and good overall model fit was shown by points clustering around the line.

statistics indicated that there was a substantial proportion of the variance in violent crime explained by the four socioeconomic variables. Table 1 reported full regression results, including coefficient estimates, standard errors, 95% confidence intervals, p-values, and model fit statistics (, adjusted , and F-statistic). This confirmed the hypothesis established in this study about how the structural disadvantage significantly contributed to crime patterns.

Effects of Individual Variables

Income

Median household income turned out to be a significant negative predictor of violent crime. Counties with higher income levels turned out to show lower violent crime rates after controlling for education, unemployment, and poverty. This result reinforced the broad literature showing how economic stability reduced both strain-related pressures and the incentives related to criminal activity. According to the direction and magnitude of the coefficient of income, it was suggested that income was one of the more influential socioeconomic variables in the model (Figure 2).

Education

Education attainment turned out to be a significant negative predictor, too. Counties with higher percentages of adults completing a high school or higher education turned out to exhibit lower violent crime rates. When other factors were controlled, the education variable turned out to be significant. This suggested that education was an independent protective effect. This was consistent with theories that education improved economic opportunities, strengthening community cohesion, but reducing exposure to high-risk environments.

Unemployment

The unemployment rate turned out to be a positive predictor of violent crime. Counties with higher unemployment rate turned out to experience higher violent crime levels. Even though unemployment indicated a positive and statistically significant association with violent crime, the effect of it was smaller than that of income and poverty. This showed a secondary but meaningful contribution. This supported criminological theories indicating how the role of blocked economic opportunities and weakened community structure in higher crime rate, particularly when communities were socially disorganized or economically distressed.

Poverty

Poverty rate indicated the strongest positive relationship with violent crime among all socioeconomic variables. Counties with higher poverty rate turned out to indicate substantially higher crime rates, even when income and unemployment were adjusted. According to the significance of poverty coefficients, it was suggested that concentrated disadvantages, limited access to resources, and social strain were powerful structural indicators of violence. This finding aligned with extensive theoretical work in social disorganization theory and strain theory that both identified poverty as a main predictor of violent offending.

Beyond statistical significance, the magnitude of the estimated coefficients suggests greatly meaningful effect sizes. More specifically, the poverty rate indicated the largest correlation with violent crime. This shows that a one-percentage point increase in count-level poverty corresponded to a considerable increase in violent crime rates, even after controlling for socioeconomic conditions considered in this study, including income, education, and unemployment. However, income and educational attainment indicated moderate protective effects. However, unemployment exhibited a smaller but non-trivial contribution. These differences may emphasize that concentrated disadvantage indicated a stronger influence on violent crime than labor market conditions alone.

Joint Effects and Model Interpretation

When putting all the socioeconomic variables together, they contributed uniquely to the prediction of violent crime. The inclusion of income, education, unemployment, and poverty all at the same time indicated that each variable indicated a distinct aspect of structural conditions that influenced violent crime. Income and education were the primary protective factors, while unemployment and poverty were risk-increasing factors.

Residual plot of the model did not show meaningful violations of linearity or homoscedasticity (Figure 3). Normality of residuals was assessed further with Q-Q plots, and no substantial deviations from normality were observed. This suggested that the linear model was appropriate for the data analysis. No counties indicated disproportional leverage or influence on the regression results. Overall, how socioeconomic disadvantages influenced violent crime rates was supported coherently and statistically by the regression model.

DISCUSSION

This study examined how four socioeconomic variables of income, education, unemployment, and poverty jointly predicted violent crime rates across U.S. counties. According to the findings in the study, the hypothesis was supported that communities with lower income and education levels but with higher unemployment and poverty indicated significantly higher violent crime rates. These findings aligned strongly with established criminological theories, including social disorganization theory, general strain theory, and routine activity theory that all underscored how crime risks were elevated by structural disadvantage and weakened community.

While the regression model explained a substantial proportion of the variance in count-level violent crime rates (R^2), almost half of the variance remained unexplained. This suggests that violent crime may also be influenced by other structural, demographic, and institutional factors that were not identified in the model proposed in this study even though socioeconomic conditions played a major role. Other factors not explained in the model may include urban density, policing practices, and residential segregation. Therefore, the model needs to be interpreted not exhaustive but may still be explained as meaningful variation in violent crime patterns rather than providing a complete causal account.

Among socioeconomic variables analyzed in this study, poverty turned out to be the strongest positive predictor of violent crime. This finding was consistent with the prior research showing how the exposure to strain was intensified by concentrated disadvantage, while reducing access to legitimate resources, and weakening the informal social controls that usually help regulate behavior. Communities with high poverty rate turned out to face more challenges, including housing instability, limited educational opportunities, and fewer local institutions that could collectively elevate vulnerability to violent crime. The magnitude of the poverty coefficient analyzed in this study suggested that structural disadvantage was still the primary indicator of violence, even when other connected socioeconomic variable were controlled.

Both income and educational attainment were protective factors, showing strong negative correlations with the violent crime. These findings supported the idea that the likelihood of crime may be reduced by economic stability and access to education by promoting social mobility, strengthening community cohesion, and

expanding access to legitimate economic opportunities. Higher income may decrease strain by meeting basic needs, while lowering economic pressure on daily lives, while education was correlated with better employment prospects, stronger social networks, and greater community engagement. Furthermore, both variables turned out to be significant when included together in the model, showing how their contributions to the safety of community were distinct and not simply reflections of one another.

This study contributed to fill the literature gap by using standardized U.S. county-level data to test four socioeconomic variables within one model, addressing the gap from prior research in regards of the lack of large-scale and standardized analyses. However, there are several limitations of this study that are worth mentioning. First, the UCR data from FBI may contain reporting inconsistencies since law enforcement agencies varied with practices across jurisdictions. Second, the cross-sectional design may have prevented causal interpretation that socioeconomic disadvantage and crime may influence each other in ways that were difficult to be explained over time. Even if the large county-level sample provided reasonable statistical power, it is recommended for future work to focus on formally assessing power and minimum detectable effects for smaller predictors, including unemployment. Third, even if the model captured main socioeconomic variables, there may be other factors, including urbanization, racial segregation, or policing practices that were not included in the study but may influence crime patterns. Due to omission of these variables, the coefficient estimates may be biased if they were correlated with included predictors. This suggests that future studies may include racial composition, urbanization measures, and intensity of law-enforcement to improve inference.

CONCLUSION

This study examined how structural socioeconomic conditions influenced violent crime rates across U.S. counties by modeling income, education, unemployment, and poverty together. The analysis provided a clear picture of how broad community characteristics correlate to violence patterns. The findings in this study indicated that these four socioeconomic variables collectively provided meaningful explanatory power, emphasizing that violent crime was strongly correlated to the social and economic environments of communities.

Instead of pointing to a single dominant cause, the

findings in this study highlighted the importance of taking multiple socioeconomic well-being dimensions simultaneously into account. Counties with higher economic stability and educational attainment tended to experience safer conditions, while communities with structural disadvantages indicated increased vulnerability. These findings underscored that an effort shall be taken to reduce violence along with broader investment in social and economic opportunities instead of solely relying on reactive criminal justice interventions.

CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

REFERENCES

1. Bureau of Justice Statistics. Socio-emotional consequences of violent crime, 2022. Available from: <https://bjs.ojp.gov/socio-emotional-consequences-violent-crime-2022>. (accessed 2025-11-10).
2. National Institute of Justice. Violent crime. Available from: <https://nij.ojp.gov/topics/crimes/violent-crime> (accessed 2025-11-9).
3. Office of Disease Prevention and Health Promotion. Crime and violence: Literature summary. Available from: <https://odphp.health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/crime-and-violence> (accessed 2025-11-8).
4. Sampson RJ, Groves WB. Community structure and crime: Testing social disorganization theory. *Am J Sociol.* 1989; 94 (4): 774-802. <https://doi.org/10.1086/229068>
5. Agnew R. Foundation for a general strain theory of crime and delinquency. *Criminology.* 1992; 30 (1): 47-87. <https://doi.org/10.1111/j.1745-9125.1992.tb01093.x>, <https://doi.org/10.1111/j.1745-9125.1991.tb01058.x>, <https://doi.org/10.1111/j.1745-9125.1985.tb00325.x>
6. Cohen LE, Felson M. Social change and crime rate trends: A routine activity approach. *Am Sociol Rev.* 1979; 44 (4): 588-608. <https://doi.org/10.2307/2094589>
7. Hooghe M, Vanhoutte B, Hardyns W, Bircan T. Unemployment, inequality, poverty and crime: A spatial analysis across Belgian municipalities. *Eur J Criminol.* 2011; 8 (5): 431-445.
8. McGahey RM. Economic conditions, neighborhood organization, and urban crime. *Law Soc Rev.* 1986; 20 (3): 393-404. <https://doi.org/10.1086/449124>
9. Kang S. Income inequality and crime: A panel data analysis. *Asia Pac Econ Lit.* 2016; 30 (2): 131-142.
10. Kelly M. Inequality and crime. *Rev Econ Stat.* 2000; 82 (4): 530-539. <https://doi.org/10.1162/003465300559028>
11. Ellen IG, O'Regan KM. Crime and urban flight revisited. *Natl Tax J.* 2009; 62 (3): 411-432.
12. Pina-Sánchez J, Brunton-Smith I, Johnson S. The impact of measurement error in crime rates on regression models. *J Quant Criminol.* 2023. Available from: <https://eprints.whiterose.ac.uk/id/eprint/197197> (accessed 2025-10-28).
13. Sariaslan A, Larsson H, D'Onofrio BM, Långström N, Lichtenstein P. Childhood family income, adolescent violent criminality and substance misuse: A longitudinal birth cohort study. *BMJ.* 2014; 348: g1-g9.
14. Groeneveld G. A cross-national spatial study of crime using variables standardized across countries. 2022. Available from: <https://repository.up.ac.za/handle/2263/92378> (accessed 2025-11-10).
15. Federal Bureau of Investigation. Uniform Crime Reports: Crime measures. Available from: <https://ucr.fbi.gov/crime-in-the-u.s/2011/crime-in-the-u.s.-2011/crime-measures> (accessed 2025-10-22).
16. Bureau of Justice Statistics. National Crime Victimization Survey: Measurement comparisons. Available from: https://bjs.ojp.gov/content/pub/pdf/ntcm_2014.pdf (accessed 2025-10-22).