Exploring the Correlation between Income Inequality and Crime Rates in a Cross-National Context

AmirSadra DaneshvarKakhki[1*]

[1]Graduate School of Management and Economics (GSME), Sharif University of Technology

Abstract

The intricate connection between income inequality and crime rates has long perplexed economists, politicians, and society at large. In our forthcoming research, we aim to elucidate, analyze, and scrutinize the variables at play in this complex issue. Subsequently, we will evaluate the hypothesis positing a correlation between them, utilizing data drawn from various countries. In this pursuit, we have selected homicide and theft rates as proxies for crime rates and explored their relationship with the Gini coefficient—a measure of income inequality—across the years 2003 to 2020. To ensure robust findings, we have incorporated additional control variables, including education levels, urbanization rates, gross national income per capita, and economic growth. Finally, employing Stata software, we have constructed a multivariable regression model to address this multifaceted problem. The results derived from our research indicate a general positive relationship between income inequality and crime when considering the data as a panel. However, it is worth noting that this relationship may exhibit variations and inconsistencies in different countries.

Keywords: *Income Inequality, Crime Rates, Multivariable Regression Model, Panel Data Econometrics*

1 Introduction

The relationship between income inequality and crime has long been a focal point in the economic literature on criminal behavior, as evidenced by Fajnzylber et al. (2002). Half a century ago, Nobel laureate economist Becker (1968) posited a foundational theory in his influential article "Crime and Punishment: An Economic Approach," asserting that all criminal activity is fundamentally driven by economic considerations, with criminals engaging in rational decision-making. Following this, an article by Economist (2018) contended that potential offenders weigh the expected benefits of unlawful actions against the probability of getting apprehended and facing penalties.

According to this perspective, Scorzafave et al. (2009) suggests that the crime rate hinges on factors such as the risks associated with arrest and the severity of penalties, as well as the disparity between the potential gains from criminal acts and the related opportunity costs. In essence, this net benefit is theoretically represented by the wealth gap between affluent and disadvantaged populations.

Moreover, the nexus between inequality and crime has also been scrutinized through sociological lenses by Krohn (1976). In broad terms, these sociological frameworks stem from the consistent observation that lower-class individuals and residents of lower-class neighborhoods consistently exhibit higher official crime rates compared to other demographic groups. One prominent sociological paradigm exhibited by Adeleye (2014), known as the "relative deprivation" theory, posits that inequality fosters social tensions. The sense of deprivation and injustice compels economically disadvantaged individuals to seek redress and satisfaction through various means, including resorting to criminal activities against both their impoverished peers and the affluent, marked by Harrendorf et al. (2010).

Economist (2018) shows that recent Gallup polling data has partially validated Becker's theory. In a survey encompassing 148,000 respondents across 142 countries, individuals were queried about their perceptions of crime and their feelings of safety, based on four key criteria: trust in local law enforcement, personal safety when traveling alone, experiences of theft or property loss, and incidents of sexual assault within the past year. Correlating these responses with the level of income inequality, as gauged by the Gini coefficient, in each respective nation revealed a robust and positive correlation, as illustrated in Figure 1.

These straightforward associations do not encompass the entirety of variations in individuals' perceptions of crime levels. Subsequently, scholars like Szalavitz (2018) have expanded upon Becker's theory to explore the extent to which the affluent showcase their wealth. Hicks et al. (2014) revealed a compelling trend over 20 years: U.S. states characterized by the highest disparities in expenditure—about items like clothing, jewelry, automobiles, and dining out—also experienced the most pronounced instances of vio-

*Acknowledgments: I would like to thank my supervisor, Dr. Narges Tahmasebi, for their guidance and support.

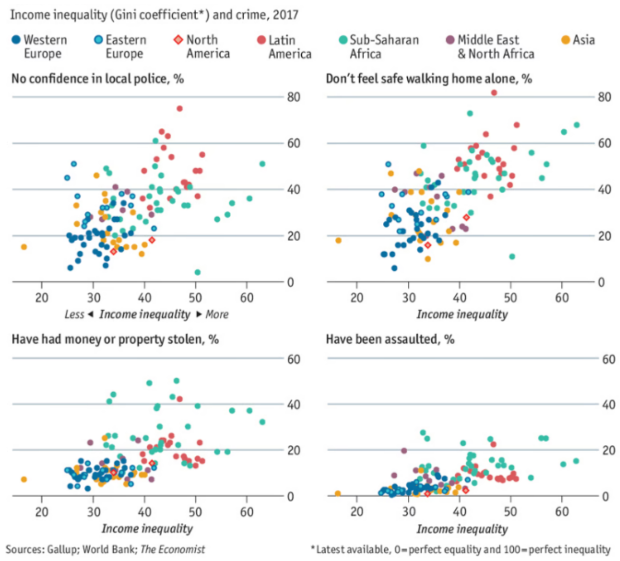

Figure 1: *Illustrating the effect of income inequality on the results of the survey, Economist (2018)*

lent crime.

To delve more precisely into the mechanics of this phenomenon and elucidate the correlation mentioned above, we found it imperative to construct a regression model. To examine the crime rate, we opted for two variables, namely, the homicide rate and robbery rate, as they serve as apt proxies. This selection was driven by their close association with overall crime levels while sidestepping the endogeneity puzzle, as seen in Fajnzylber et al. (2002). Moreover, we have employed the Gini index as a proxy to elucidate the degree of income inequality. Supplementary variables have been judiciously chosen to mitigate potential distortions and tailor the model; a comprehensive discussion will ensue in Section 2.

2 Data

2.1 Data Sources, Structure, and Information

In our study, we focused on two dependent variables: HomicideRate and RobberyRate, both of which measure the occurrence of these crimes per 100,000 people. The data source for HomicideRate was the World Bank, which provided data for 140 countries from 1968 to 2020. Additionally, we obtained RobberyRate data from United Nations sources, covering 88 countries from 1990 to 2020.

For our independent variables, we initially conducted a regression analysis with only the Gini coefficient, which we also sourced from the World Bank. However, this preliminary analysis resulted in a weak model that failed to explain variations in the dependent variables across all countries adequately.

To enhance the model's predictive power and mitigate the impact of omitted variables, we incorporated two additional factors: economic growth (GDPGrowth) and gross national income per capita (GNIperCapita), both obtained from the World Bank. This multivariable regression yielded significantly improved results, although further refinements were still possible.

Finally, we introduced two more explanatory variables into our model: average years of education (AverageSchooling) and the rate of urbanization (Urbanization), covering the years 1980 to 2020, sourced from Our World in Data. These additions notably enhanced the model's accuracy. Table 1 presents a summary of the data and their respective sources.

Table 1: *Data information*

Variable Name	Available years	Countries available (count)	Source
Homicide Rate	1968-2020	140	World Bank
Rubbery Rate	1990-2020	88	UN
Gini Index	1968-2020	140	World Bank
GNI per Capita	1968-2020	140	World Bank
GDP Growth	1968-2020	140	World Bank
Average Schooling	1980-2020	122	OWID
Urbanization	1980-2020	122	OWID

2.2 Selecting and Cleaning Data

In handling the diverse data sources at our disposal, our primary objective was to combine them and rectify any discrepancies. Our initial course of action involved excluding data preceding the year 2000, primarily due to the prevalence of numerous missing values. Additionally, we initially considered 32 countries from the global dataset but ultimately narrowed it down to 11 countries whose data demonstrated greater consistency across various sources. These countries comprised Mexico, Colombia, France, Germany, Greece, Hungary, Italy, Mexico, Netherlands, Norway, and the United States.

However, even within this subset of 11 countries, there remained instances of missing data for homicide and robbery rates until 2003. Consequently, we applied a filtering process to establish our final timeframe, encompassing 2003 to 2020. The conclusive dataset is provided as an appendix for further examination.

Nevertheless, some gaps persisted in the data, prompting us to employ interpolation techniques in Stata to supplement our dataset and ensure completeness. Furthermore, our data's suitability for panel analysis was an interesting aspect. Although this lay outside our formal training, driven by personal interest and curiosity, we ventured into panel analysis in addition to single-country analysis. Detailed insights into these aspects can be found in Sections 3 and 4 of this study.

3 Single Country Analysis

3.1 Country Selection

As previously mentioned, our dataset includes 11 countries from different regions: *France, Germany, Greece, Hungary, Italy, Netherlands, Norway, United States, Mexico, Mexico, and Colombia*. To simplify the analysis and avoid redundancy, we grouped these countries into three distinct categories based on socioeconomic and geopolitical factors.

- **Group1; United States, Brazil, Colombia**: These countries all exhibit high levels of income inequality. Despite the U.S.'s stronger institutions, it still faces significant inequality linked to crime in urban areas. Brazil and Colombia deal with more acute challenges of organized crime and corruption, often exacerbated by poverty and weaker governance. While these countries differ in economic development, they share social issues connected to inequality and institutional weaknesses, making them more prone to higher crime rates than stable countries.
- **Group 2; Germany, France, Netherlands, Norway**: This group consists of highly developed economies with low levels of income inequality and strong welfare systems. Germany, France, the Netherlands, and Norway all have robust social safety nets, contributing to lower crime rates and more excellent social stability. The combination of stable institutions and effective policies makes these countries more resistant to crime driven by poverty or inequality. Compared to Groups 1 and 3, they maintain much stronger governmental structures, allowing them to better manage social issues and economic challenges.
- **Group 3; Mexico, Greece, Hungary, Italy**: Countries in this group face moderate to high-income inequality and weaker institutional frameworks than Group 2. Mexico shares similarities with Greece, Hungary, and Italy, which face significant economic challenges like unemployment and slow growth. While Mexico struggles with organized crime driven by inequality, the European countries in this group grapple with financial instability and governance issues. Despite being part of the EU, Greece, Hungary, and Italy have weaker institutional capacities than their Northern European counterparts, leading to greater economic vulnerabilities and higher susceptibility to social unrest.

We selected the United States, Mexico, and Germany as representative countries of these three groups. This choice was motivated by their roles as economic and social leaders within their respective groups, which offer a broad range of contexts for analyzing the relationship between income inequality and crime.

1. **United States**
   (a) *High-income inequality:* The United States exhibits one of the highest levels of income inequality among developed nations, making it a critical case for examining its relationship with crime.
   (b) *Global economic leader:* As the largest economy in the world, U.S. economic policies have widespread global influence, adding to its importance in this analysis.
   (c) *Societal diversity and urbanization:* The U.S. has a highly diverse population and extensive urbanization, providing varying contexts for examining inequality and crime patterns.
2. **Mexico**
   (a) *Emerging market challenges:* Mexico's higher income inequality and crime rates, particularly driven by drug-related violence, offer a stark contrast to both the U.S. and Germany.
   (b) *Latin American representative:* Mexico's position within Latin America and its social and economic challenges make it an important country for comparison.
   (c) *Institutional struggles and corruption:* Mexico faces significant institutional weaknesses, with corruption and governance issues contributing to high crime levels, making it an essential case for this study.
3. **Germany**
   (a) *Relatively low-income inequality:* Germany's lower income inequality makes it a valuable contrast as we explore whether this correlates with lower crime rates compared to other countries.

(b) *Strong social welfare system:* Germany's extensive social safety net helps mitigate the effects of income inequality, potentially reducing crime rates.

(c) *Economic stability:* As Europe's largest economy, Germany's economic power offers insights into how strong institutions and low inequality impact crime.

Focusing on these three countries allows us to avoid redundancy and maintain a clear focus on the relationship between income inequality and crime across distinct but comparable contexts rather than getting lost in each country's unique nuances.

3.2 United States

Before delving into data analysis from the United States, we conducted a preliminary examination to assess the extent of missing data. The findings of this assessment are presented in Table 2.

Table 2: *United States data information*

Variable	Missing	Total	Percent Missing
Country	0	18	0.000
Year	0	18	0.000
HomocideRate	3	18	16.670
RobberyRate	1	18	5.560
GiniIndex	1	18	5.560
GNIperCapita	0	18	0.000
GDPGrowth	0	18	0.000
AverageSch g	0	18	0.000
Urbanization	0	18	0.000

You can observe that all variables, with the exception of HomicideRate, RobberyRate, and GiniIndex, are complete and devoid of any issues. These three variables have also been filled in and are prepared for further processing using interpolation. Before delving into a detailed analysis of this dataset, it would be beneficial to take a moment to review its summary information in Table 3.

The data clearly suggest that, on average, the American economy exhibits positive and substantial growth. Simultaneously, the incidence of theft in the country is notably high.

Turning our attention to the relationship between the homicide rate and the Gini coefficient within this nation, our analysis indicates a weak positive correlation between these two variables, even without controlling other factors as seen in Figure 2.

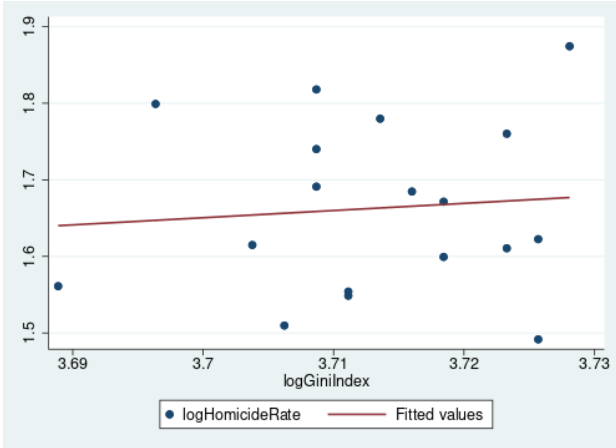

Figure 2: *Correlation between the Gini coefficient and homicide rates in the United States*

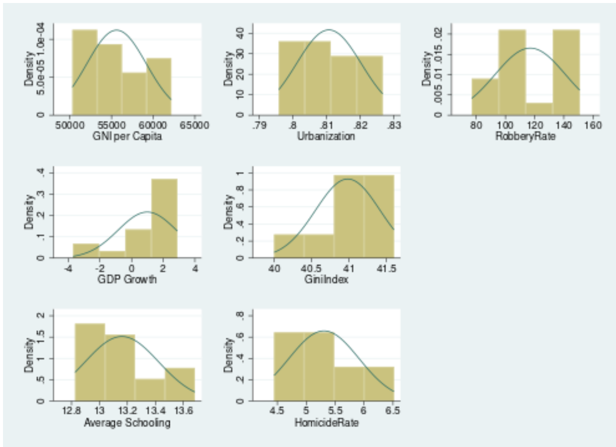

Figure 3: *United States data distribution*

In the above analysis, we examined whether the data for this specific country follows a normal distribution. When the data showed skewness, we addressed it by applying the natural logarithm, ensuring the Ordinary Least Squares (OLS) regression assumptions were not violated. For the United States, the variables Homicide Rate, Average Schooling, Gini Index, and GDP Growth exhibited skewness, so we used their natural logarithms in our modeling.

Now, let's present the American regression models based on these results:

Model 1:

$$\log(\text{HomicideRate}) = \beta_0 + \beta_1 \log(\text{GiniIndex})$$

Model 2:

$$\log(\text{HomicideRate}) = \beta_0 + \beta_1 \log(\text{GiniIndex}) + \beta_2 \,\text{GNIperCapita} + \beta_3 \,\text{GDPGrowth}$$

Table 3: *United States data summary*

Variable	Obs	Mean	Std. Dev.	Min	Max
GiniIndex	18	40.983	.43	40	41.6
RobberyRate	18	117.001	24.12	76.887	150.876
HomicideRate	18	5.306	.606	4.445	6.517
Urbanization	18	.811	.01	.796	.827
AverageSchooling	18	13.16	.64	12.825	13.683
GDPGrowth	18	.99	1.846	-3.698	2.896
GNIperCapita	18	55602.523	3562.465	50303.713	62173.485
Year	18	2011.5	5.339	2003	2020

Model 3:

$$\begin{aligned}\log(\text{HomicideRate}) = \beta_0 &+ \\ &\beta_1 \log(\text{GiniIndex}) + \\ &\beta_2\, \text{GNIperCapita} + \\ &\beta_3\, \text{GDPGrowth} + \\ &\beta_4 \log(\text{AverageSchooling}) + \\ &\beta_5\, \text{Urbanization}\end{aligned}$$

You can find the results of the regressions applied to this country using the above models in Table 4.

Table 4: *Regression of the homicide rate in the US*

VARIABLES	Model One	Model Two	Model Three
logGiniIndex	0.937	4.408	-2.027
	(2.670)	(3.970)	(2.868)
GNIperCapita		-1.45e-05	6.14e-05**
		(1.17e-05)	(2.50e-05)
GDPGrowth		-0.00713	-0.0368**
		(0.0154)	(0.0150)
logAverageSchooling			15.77**
			(5.763)
Urbanization			-57.40***
			(11.85)
Constant	-1.815	-13.89	11.70
	(9.913)	(14.27)	(15.57)
Observations	18	18	18
R-squared	0.008	0.128	0.721

Note: *p<0.1; **p<0.05; ***p<0.01

We can observe that the third model, which incorporates a more significant number of controlled variables, explains approximately 72% of the dependent variable's variations. This is indeed a substantial and desirable level of explanation. Furthermore, the positive coefficient associated with logGiniIndex suggests a direct relationship between this variable and the homicide rate. However, it's worth noting that this relationship is not statistically significant. The explanations and interpretations of the remaining variables are precise. Nonetheless, it's important to highlight that all of these variables are statistically significant in the third and last model.

Now, we can focus our attention on the Robbery Rates. The following models are also calculated for this variable:

Model 1:

$$\begin{aligned}\log(\text{RobberyRate}) = \beta_0 &+ \\ &\beta_1 \log(\text{GiniIndex})\end{aligned}$$

Model 2:

$$\begin{aligned}\log(\text{RobberyRate}) = \beta_0 &+ \\ &\beta_1 \log(\text{GiniIndex}) + \\ &\beta_2\, \text{GNIperCapita} + \\ &\beta_3\, \text{GDPGrowth}\end{aligned}$$

Model 3:

$$\begin{aligned}\log(\text{RobberyRate}) = \beta_0 &+ \\ &\beta_1 \log(\text{GiniIndex}) + \\ &\beta_2\, \text{GNIperCapita} + \\ &\beta_3\, \text{GDPGrowth} + \\ &\beta_4 \log(\text{AverageSchooling}) + \\ &\beta_5\, \text{Urbanization}\end{aligned}$$

In this new regression analysis (results depicted in Table 5), it becomes evident that, again, the third model offers a significantly superior explanation of the variations in the dependent variable compared to the other two models. This enhanced performance can be attributed to carefully controlling and eliminating distortions caused by omitted variables. Notably, in this particular instance, the coefficient for logGiniIndex fails to reach statistical significance and, somewhat surprisingly, exhibits a negative relationship, contradicting the expected evidence.

Moreover, Figure 4 shows that the error distribution in the regressions of all three models closely

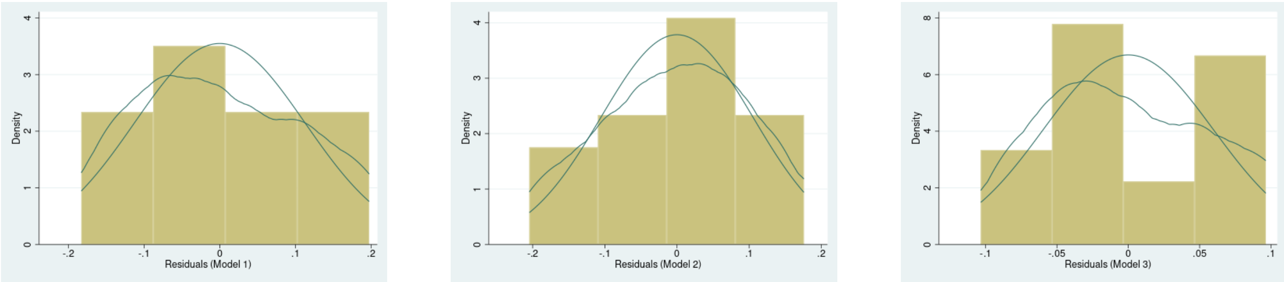

Figure 4: *United States error distributions*

Table 5: *Regression of the robbery rate in the US*

VARIABLES	Model One	Model Two	Model Three
logGiniIndex	-1,217**	595.1	-220.0
	(485.5)	(403.5)	(221.4)
GNIperCapita		-0.0072***	0.00626***
		(0.0012)	(0.00193)
GDPGrowth		1.978	-4.533***
		(1.570)	(1.158)
logAverageSchooling			-1.098**
			(444.8)
Urbanization			-2,367**
			(914.9)
Constant	4,637**	-1,695	5,338***
	(1,803)	(1,451)	(1,202)
Observations	18	18	18
R-squared	0.282	0.803	0.964

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: *Mexico data information*

Variable	**Missing**	**Total**	**Percent Missing**
Country	0	18	0.000
Year	0	18	0.000
HomocideRate	3	18	16.670
RobberyRate	8	18	44.440
GiniIndex	0	18	0.000
GNIperCapita	0	18	0.000
GDPGrowth	0	18	0.000
AverageSch g	0	18	0.000
Urbanization	0	18	0.000

resembles a normal distribution, which is a favorable outcome. Generally, in the United States, there doesn't appear to be a robust and precise correlation between the income inequality index and the prevalence of crime.

3.3 Mexico

Like the US, we conducted an initial assessment to gauge the volume of available data before delving into the data from Mexico. The findings from this evaluation are presented in Table 6.

All variables, except for Robberyrate and GiniIndex, are complete and free of issues. These two variables have also been filled in and are now ready for interpolation processing. Before delving into a detailed analysis of this dataset, it would be beneficial to review its summary information (Table 7).

We can observe that Mexico experiences consistently modest but positive economic growth. Conversely, the country faces an average theft rate of nearly four times that of the United States.

Furthermore, examining the correlation between the homicide rate and the Gini coefficient in Mexico reveals an intriguing pattern. Notably, there appears to be a negative relationship between these two variables, even without controlling for other factors.

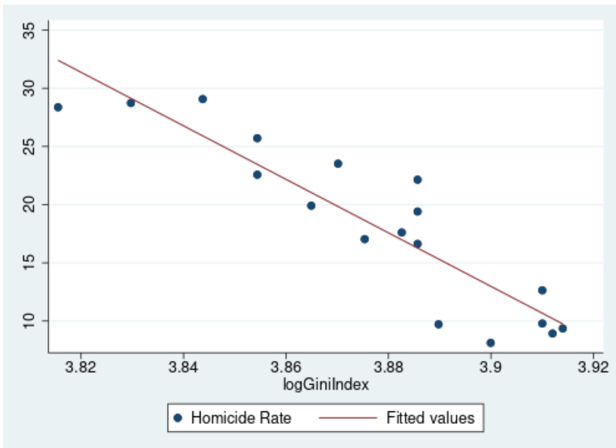

Figure 5: *Correlation between the Gini coefficient and homicide rates in Mexico*

Like for the US, we examined whether the data about Mexico conforms to a normal distribution. In cases where the data exhibited skewness, we addressed this issue by applying the natural logarithm to ensure that the fundamental assumptions of OLS regression would not be compromised. In the context of Mexico, our findings are as follows in Figure 6.

Table 7: *Mexico data summary*

Variable	Obs	Mean	Std. Dev.	Min	Max
GiniIndex	18	48.292	1.379	45.4	50.1
RobberyRate	18	411.087	197.311	128.833	649.766
HomicideRate	18	18.293	7.276	8.122	29.071
Urbanization	18	.782	.016	.757	.807
AverageSchooling	18	8.246	.588	7.132	9.221
GDPGrowth	18	.362	3.138	-8.655	3.733
GNIperCapita	18	9146.963	412.863	8488.411	9841.469
Year	18	2011.5	5.339	2003	2020

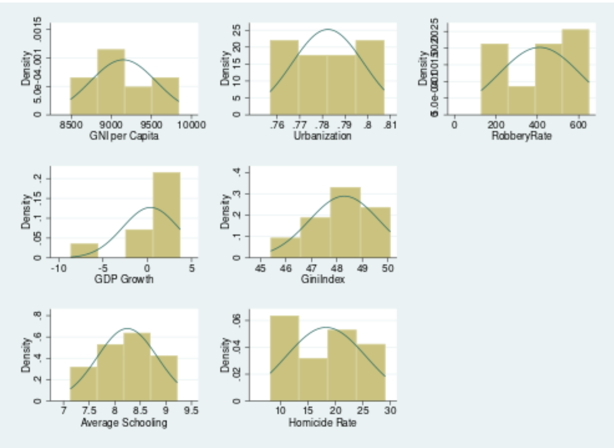

Figure 6: *Correlation between the Gini coefficient and homicide rates in Mexico*

We observe that the GiniIndex is the only skewed variable. Thus, we applied a logarithmic transformation to it for our modeling. The regression models for the Mexico dataset are outlined below:

Model 1:

$$\log(\text{HomicideRate}) = \beta_0 + \beta_1 \log(\text{GiniIndex})$$

Model 2:

$$\log(\text{HomicideRate}) = \beta_0 + \beta_1 \log(\text{GiniIndex}) + \beta_2 \,\text{GNIperCapita} + \beta_3 \,\text{GDPGrowth}$$

Model 3:

$$\log(\text{HomicideRate}) = \beta_0 + \beta_1 \log(\text{GiniIndex}) + \beta_2 \,\text{GNIperCapita} + \beta_3 \,\text{GDPGrowth} + \beta_4 \,\text{AverageSchooling} + \beta_5 \,\text{Urbanization}$$

The results of these regression models are available in Table 8.

Table 8: *Regression of the homicide rate in Mexico*

VARIABLES	Model One	Model Two	Model Three
logGiniIndex	-230.0**	-213.2***	-128.1*
	(26.34)	(39.73)	(61.97)
GNIperCapita		0.0024	-0.0047
		(0.0027)	(0.0049)
GDPGrowth		0.0013	0.479
		(0.331)	(0.432)
AverageSchooling			2.231
			(6.790)
Urbanization			239.4
			(264.0)
Constant	909.8***	823.0***	352.2
	(102.1)	(171.7)	(337.0)
Observations	18	18	18
R-squared	0.827	0.840	0.873

Note: *p<0.1; **p<0.05; ***p<0.01

All three models have successfully explained the variations in the dependent variable, demonstrating their effectiveness in capturing the underlying patterns in the data. However, it is essential to highlight that the third model has performed better than the other two, providing a more accurate and robust explanation of these variations.

Furthermore, it is crucial to point out that only the logGiniIndex variable has a statistically significant coefficient among all the variables. Interestingly, this negative coefficient suggests an inverse relationship between the logGiniIndex and the dependent variable. This finding may have important implications for interpreting the results and understanding the dynamics captured by the models.

Similar analyses were also conducted for RobberyRate Using the following models. The results of these are shown in Table 9.

Model 1:

$$\log(\text{RobberyRate}) = \beta_0 + \beta_1 \log(\text{GiniIndex})$$

Model 2:

$$\log(\text{RobberyRate}) = \beta_0 + \beta_1 \log(\text{GiniIndex}) + \beta_2\,\text{GNIperCapita} + \beta_3\,\text{GDPGrowth}$$

Model 3:

$$\log(\text{RobberyRate}) = \beta_0 + \beta_1 \log(\text{GiniIndex}) + \beta_2\,\text{GNIperCapita} + \beta_3\,\text{GDPGrowth} + \beta_4\,\text{AverageSchooling} + \beta_5\,\text{Urbanization}$$

Table 9: *Regression of the robbery rate in Mexico*

VARIABLES	Model One	Model Two	Model Three
logGiniIndex	2,102	-1,471	-4,616*
	(1.632)	(1,999)	(2,290)
GNIperCapita		-0.396**	-0.174
		(0.135)	(0.179)
GDPGrowth		18.87	12.26
		(16.66)	(15.95)
AverageSchooling			770.3***
			(250.9)
Urbanization			-38,876***
			(9,755)
Constant	-7,740	9,723	43,956***
	(6,329)	(8,637)	(12,450)
Observations	18	18	18
R-squared	0.094	0.450	0.764

Note: *p<0.1; **p<0.05; ***p<0.01

As in the previous analysis, the third model clearly offers a superior explanation of changes in the dependent variable compared to the other two models. This improvement is due to its effective control of distortions from omitted variables. Notably, the coefficient associated with the logGiniIndex variable in the third model approaches significance and demonstrates an unexpected negative relationship, contradicting both our prior evidence and expectations.

Overall, the data reveals an intriguing and contentious pattern in Mexico. There appears to be an almost inverse correlation between the income inequality index and the incidence of crime and criminal activities. This observation is both captivating and contentious, warranting further, more precise investigation in future studies.

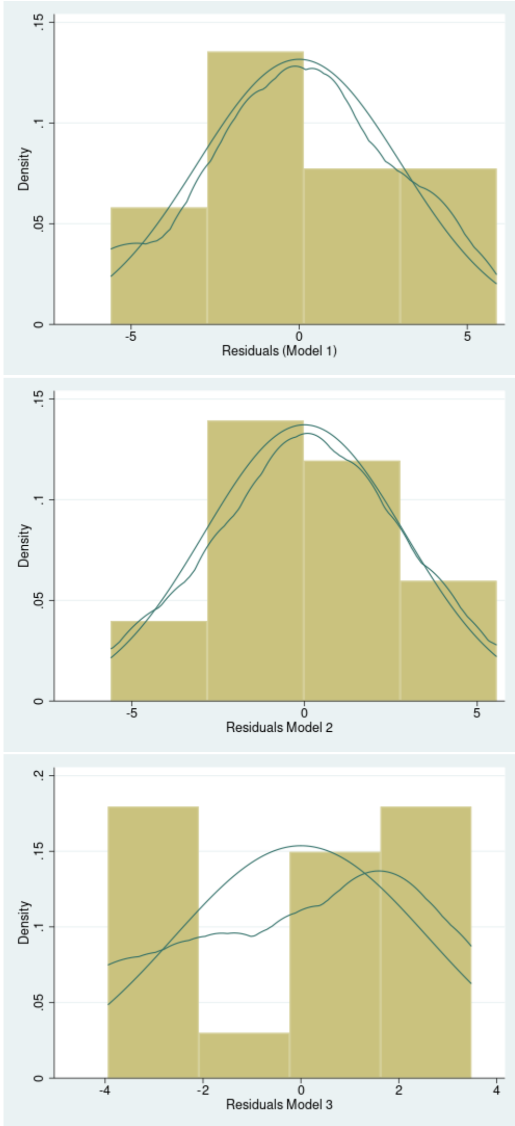

Figure 7: *Mexico error distributions*

Finally, Figure 7 shows density plots of the residuals for the three different models. In the first model, the residuals are roughly symmetric and centered around zero, indicating a reasonable model fit. However, the density plot shows a slight skew to the left, with heavier tails than a normal distribution. Model 2's residuals appear more normally distributed, with a smoother density curve. This suggests that Model 2 might provide a slightly better fit than Model 1, as the residuals are more normally distributed, and the tails have less deviation.

In contrast, Model 3 shows a more irregular distribution with two distinct peaks, suggesting issues like model misspecification, non-linearity, or outliers. The residuals for Model 3 deviate significantly from normality, indicating a poor model fit and the need for further investigation or adjustment.

3.4 Germany

The exact same procedures were also used for Germany, starting with assessing its data to determine how much empty data it had. The result of this study can be seen in Table 10.

Table 10: *Germany data information*

Variable	Missing	Total	Percent Missing
Country	0	18	0.000
Year	0	18	0.000
HomocideRate	0	18	0.000
RobberyRate	1	18	5.560
GiniIndex	2	18	11.110
GNIperCapita	0	18	0.000
GDPGrowth	0	18	0.000
AverageSch g	0	18	0.000
Urbanization	0	18	0.000

We can observe that all variables, except for RobberyRate and GiniIndex, are complete and free of any issues. These two variables have also been completed and are now prepared for processing through interpolation. As we did for the US and Mexico, getting a glimpse at Germany's data summary information is a good idea before delving into a detailed dataset analysis.

As evident in Table 11, we observe that Germany experiences consistently positive economic growth, which is noteworthy. Additionally, the incidence of robbery and homicide in the country is relatively low on average.

We then explored the correlation between the homicide rate and the Gini coefficient in Germany. Interestingly, we again found a negative relationship between these two variables when other factors were not considered (Figure 8).

Subsequently, we assessed whether this country's data adheres to a normal distribution. This step was taken to minimize potential disruptions to the fundamental assumptions of OLS regressions, as discussed in previous subsections. In the context of Germany, we have Figure 9.

We observe that the Urbanization and Average Schooling variables exhibit skewness. Consequently, we applied a logarithmic transformation to these variables and used their logarithmic values in our modeling.

Based on these actions, the regression models for the German dataset are as follows:

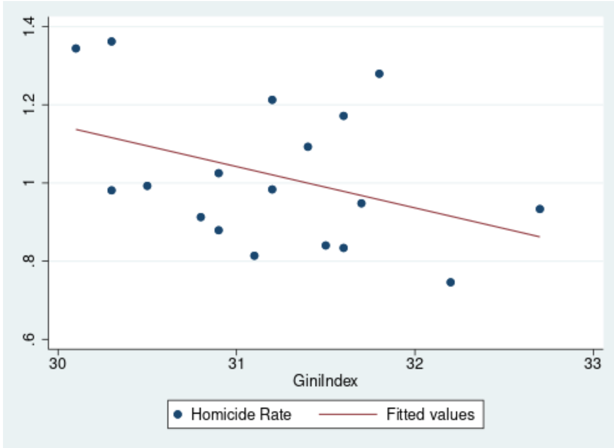

Figure 8: *Correlation between the Gini coefficient and homicide rates in Germany*

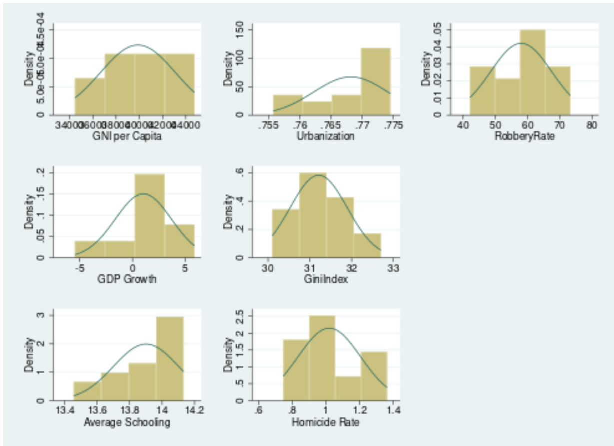

Figure 9: *Correlation between the Gini coefficient and homicide rates in Germany*

Model 1:

$$\log(\text{HomicideRate}) = \beta_0 + \beta_1\,\text{GiniIndex}$$

Model 2:

$$\log(\text{HomicideRate}) = \beta_0 + \beta_1\,\text{GiniIndex} + \beta_2\,\text{GNIperCapita} + \beta_3\,\text{GDPGrowth}$$

Model 3:

$$\log(\text{HomicideRate}) = \beta_0 + \beta_1\,\text{GiniIndex} + \beta_2\,\text{GNIperCapita} + \beta_3\,\text{GDPGrowth} + \beta_4\log(\text{AverageSchooling}) + \beta_5\log(\text{Urbanization})$$

The results of Germany's regression models are in Table 12.

Table 11: *Germany data summary*

Variable	Obs	Mean	Std. Dev.	Min	Max
GiniIndex	18	31.211	.685	30.1	32.7
RobberyRate	18	57.989	9.465	42.116	73.249
HomicideRate	18	1.02	.186	.746	1.362
Urbanization	18	.768	.006	.756	.775
AverageSchooling	18	13.901	.2	13.455	14.132
GDPGrowth	18	1.023	2.653	-5.455	5.87
GNIperCapita	18	39881.133	3234.015	34457.043	44758.513
Year	18	2011.5	5.339	2003	2020

Table 12: *Regression of the homicide rate in Germany*

VARIABLES	Model One	Model Two	Model Three
GiniIndex	-0.106	0.0488	0.00220
	(0.0624)	(0.0671)	(0.0496)
GNIperCapita		-4.8e-05***	1.02e-05
		(1.39e-05)	(2.37e-05)
GDPGrowth		0.00651	0.00228
		(0.0131)	(0.00963)
logAverageSchooling			5.409
			(4.489)
logUrbanization			-34.02***
			(8.497)
Constant	4.319**	1.419	-22.67
	(1.949)	(1.776)	(12.83)
Observations	18	18	18
R-squared	0.152	0.546	0.808

Note: *p<0.1; **p<0.05; ***p<0.01

The third model, with more controlled variables, explains nearly 81% of the dependent variable's variation, which is favorable for our analysis. The positive coefficient for logGiniIndex indicates a slight but non-significant relationship with the homicide rate. While these findings are insightful, interpreting the remaining variables and their implications requires further discussion and analysis beyond our current scope.

Similar models have been developed for RobberyRate as well, resulting in Table 13:

Model 1:

$$\log(\text{RobberyRate}) = \beta_0 + \beta_1 \,\text{GiniIndex}$$

Model 2:

$$\log(\text{RobberyRate}) = \beta_0 + \beta_1 \,\text{GiniIndex} + \beta_2 \,\text{GNIperCapita} + \beta_3 \,\text{GDPGrowth}$$

Model 3:

$$\log(\text{RobberyRate}) = \beta_0 + \beta_1 \,\text{GiniIndex} + \beta_2 \,\text{GNIperCapita} + \beta_3 \,\text{GDPGrowth} + \beta_4 \log(\text{AverageSchooling}) + \beta_5 \log(\text{Urbanization})$$

Table 13: *Regression of the robbery rate in Germany*

VARIABLES	Model One	Model Two	Model Three
GiniIndex	-9.687***	-1.173	-1.231
	(2.463)	(1.278)	(1.336)
GNIperCapita		-0.0026***	-0.0026***
		(0.00027)	(0.00064)
GDPGrowth		0.530*	0.570**
		(0.250)	(0.259)
logAverageSchooling			155.5
			(121.0)
logUrbanization			-279.5
			(229.0)
Constant	360.3***	198.7***	-282.1
	(76.89)	(33.83)	(345.7)
Observations	18	18	18
R-squared	0.492	0.936	0.946

Note: *p<0.1; **p<0.05; ***p<0.01

This regression analysis shows that the third model outperformed the other two models in explaining changes in the dependent variable. This superior performance can be attributed to its ability to control and eliminate distortions caused by omitted variables. Notably, in this case, the GiniIndex coefficient lacks statistical significance and even exhibits a negative relationship, which contradicts the expected evidence.

Furthermore, it's worth noting that the error distribution in these regressions closely approximates a normal distribution, as evident in Figure 10, which is a favorable characteristic for regression analysis.

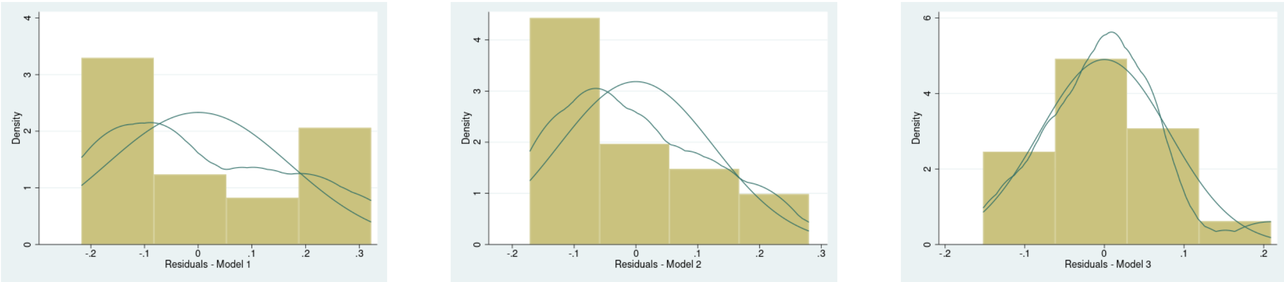

Figure 10: *Germany error distributions*

In summary, when examining the relationship between the index of income inequality and the incidence of crimes in Germany, it becomes evident that a precise and robust trend is not discernible.

4 Panel Data Analysis

4.1 Overview

In the upcoming section, we will introduce the regression model for panel data analysis. Before delving into the model itself, it's beneficial to examine the summary of its information (Table 14).

This summary provides critical descriptive statistics for the panel data used in our following regression analysis, covering 11 countries over 18 years (2003–2020). The variables include economic and social indicators such as GDP growth, average years of schooling, urbanization rate, homicide rate, robbery rate, Gini index (income inequality), and GNI per capita. The "overall," "between," and "within" variations show how much of the variation in each variable is explained across countries (between) and over time within countries (within). For example, GDP growth has a mean of 0.72%, with significant variation between countries (standard deviation = 0.93) and even larger fluctuations over time within countries (standard deviation = 2.85). This suggests that economic growth differs across countries but also changes considerably from year to year.

Another important observation is the variation in social variables such as the homicide and robbery rates. The homicide rate shows an overall mean of 8.37, but the high standard deviation (12.17) indicates vast differences, with substantial variation between countries (12.02) and relatively lower variation within countries over time (3.50). Similarly, the Gini index, which measures inequality, shows notable between-country differences, suggesting varying levels of inequality among the nations, while within-country changes over time are much more minor. These patterns are crucial for understanding how both country-specific characteristics and temporal dynamics influence the regression outcomes, and they highlight the importance of distinguishing between cross-country and within-country effects in the analysis.

4.2 Panel Regressions

It's worth highlighting that in section 3, we analyzed a sample of three countries from three different continents. Still, all 11 countries were included in the calculations and regression for the panel data analysis. The outcome of the regression analysis for the homicide rate and its explanatory variables is presented in Table 15.

In panel mode, the regression results indicate that approximately 81% of the variation in the dependent variable, the homicide rate, can be explained by the independent variables, as reflected by the R-squared value of 0.813. This suggests a robust model fit, meaning that the selected predictors—such as income inequality, GDP growth, education, urbanization, and GNI per capita—significantly influence homicide rates across countries and over time.

A closer look at the individual variables reveals that the Gini coefficient, which measures income inequality, is highly significant (p-value < 0.01) and has a positive coefficient (0.898). This implies that an increase in income inequality is associated with a higher homicide rate. Specifically, for every one-unit increase in the Gini coefficient, controlling for other factors, the homicide rate increases by approximately 0.9 homicides per 100,000 people. This relationship underscores the social cost of economic inequality, suggesting that disparities in income distribution may exacerbate social tensions, leading to higher levels of violent crime.

Additionally, other variables like urbanization and GDP growth are also significant. Urbanization has a considerable positive impact on the homicide rate, with a coefficient of 21.512, indicating that a higher urbanization rate is associated with an increase in homicide rates. This may reflect challenges in managing crime in rapidly urbanizing areas. On the other hand, average schooling years has a negative and significant effect, with a coefficient of -0.885, suggest-

Table 14: *Summary of panel data information*

Variable	Mean	Std. dev.	Min	Max	Observations
Countr~e overall	6	3.170294	1	11	N = 198
between		3.316625	1	11	n = 11
within		0	6	6	T = 18
Year overall	2011.5	5.201279	2003	2020	N = 198
between		0	2011.5	2011.5	n = 11
within		5.201279	2003	2020	T = 18
GDPGro~h overall	0.720527	2.984468	-10.01628	6.522816	N = 198
between		0.9274805	-0.6818621	2.209211	n = 11
within		2.847947	-9.818751	6.491138	T = 18
Averag~g overall	10.69311	2.219893	5.804	14.132	N = 198
between		2.270063	7.071833	13.90067	n = 11
within		0.4684339	9.425273	11.78555	T = 18
Urbaniz~n overall	0.7819927	0.056451	0.65489	0.92236	N = 198
between		0.0559142	0.6891517	0.8746194	n = 11
within		0.0181848	0.7107833	0.8297333	T = 18
Homic~ve overall	8.365383	12.16919	-46.15538	56.70396	N = 198
between		12.01912	-7.174444	34.17747	n = 11
within		3.502582	-3.171863	30.89187	T = 18
Robber~e overall	154.4513	180.7944	5.90248	806.491	N = 198
between		172.3402	22.77389	556.77394	n = 11
within		74.50891	-127.8025	404.2029	T = 18
GiniIn~x overall	37.65985	9.457186	25.1	57.6	N = 198
between		9.794737	27.61471	53.64111	n = 11
within		1.334037	29.79846	42.72906	T = 18
GNIPer~a overall	30742.5	21639.6	4008.583	81611.71	N = 198
between		22516.15	5319.622	75930.25	n = 11
within		2245.945	18302.06	37313.46	T = 18

Table 15: *Panel regression on the homicide rate*

	Coef.	St.Err.	t-value	p-value	[95% Conf.	Interval]	Sig
GiniIndex	0.898	0.072	12.43	0.000	0.756	1.041	***
GNIPerCapita	0.000	0.000	-1.14	0.255	-0.000	0.000	
GDPGrowth	0.267	0.129	2.08	0.039	0.014	0.521	**
AverageSchooling	-0.885	0.333	-2.65	0.009	-1.542	-0.227	***
Urbanization	21.512	7.869	2.73	0.007	5.991	37.033	***
Constant	-31.997	6.933	-4.61	0.000	-45.672	-18.322	***
Mean dependent var		8.365	SD dependent var			12.169	
R-squared		0.813	Number of obs			198	
F-test		166.403	Prob > F			0.000	
Akaike crit. (AIC)		1231.014	Bayesian crit. (BIC)			1250.744	

Note: * p<0.1; ** p<0.05; *** p<0.01

ing that higher education levels are associated with lower homicide rates. This highlights the potential of education as a mitigating factor against crime, emphasizing the importance of policies that increase access to education as a crime prevention strategy.

The results show that income inequality, urbanization, and education are key drivers of homicide rates, with the model providing meaningful insights into the socioeconomic factors that contribute to violent crime.

Table 16: *Panel regression on the robbery rate*

	Coef.	St.Err.	t-value	p-value	[95% Conf.	Interval]	Sig
GiniIndex	6.979	1.584	4.41	0.000	3.855	10.103	***
GNIPerCapita	0.000	0.001	-0.22	0.826	-0.001	0.001	
GDPGrowth	-3.460	2.819	-1.23	0.221	-9.021	2.100	
AverageSchooling	-30.176	7.303	-4.13	0.000	-44.580	-15.772	***
Urbanization	767.986	172.448	4.45	0.000	427.85	1108.121	***
Constant	-379.459	151.936	-2.50	0.013	-679.138	-79.781	**
Mean dependent var		154.451	SD dependent var			180.794	
R-squared		0.592	Number of obs			198	
F-test		55.730	Prob > F			0.000	
Akaike crit. (AIC)		2453.521	Bayesian crit. (BIC)			2473.251	

Note: * p<0.1; ** p<0.05; *** p<0.01

We also do this panel regression for robbery rates to further examine these factors. Table 16 depicts the results of this analysis.

The same holds for robbery rates! The regression results for the robbery rate reveal that approximately 60% of the variation in the dependent variable (robbery rate) can be explained by the independent variables in the model, as reflected by the R-squared value of 0.592. While this indicates a moderately strong fit, it suggests that other unobserved factors may also contribute significantly to robbery rates. Despite this, the selected variables—income inequality, GDP growth, education, urbanization, and GNI per capita—provide valuable insights into the socioeconomic drivers of robbery.

The Gini coefficient is again highly significant (p-value < 0.01) and positively correlated with robbery rates. Its coefficient of 6.979 suggests that for every one-unit increase in the Gini coefficient, the robbery rate increases by approximately 6.98 per 100,000 people, holding all other factors constant. This finding underscores the link between income inequality and property crimes like robbery, as more significant economic disparities may increase the propensity for crime driven by economic desperation or social tension.

The urbanization rate also significantly positively affects robbery rates, with a significant coefficient of 767.986. This implies that a higher urbanization rate is associated with a substantial increase in robbery rates. Urban areas, especially those experiencing rapid growth, may face challenges in maintaining public safety, which can lead to higher crime rates. The relationship suggests that urban management and policing are critical factors in addressing robbery in cities.

On the other hand, the average schooling years have a significant negative effect on the robbery rate, with a coefficient of -30.176 (p-value < 0.01). This indicates that higher levels of education contribute to a lower robbery rate, highlighting education's role as a preventive measure against crime. Increased access to education may provide more opportunities for economic advancement, reducing the incentives for engaging in criminal activities like robbery.

Other variables, such as GDP growth and GNI per capita, do not appear to have a statistically significant effect on robbery rates, suggesting that short-term economic fluctuations or overall national wealth may not directly influence robbery rates. The results emphasize the importance of focusing on structural factors like income inequality, urbanization, and education when addressing robbery rates through public policy and social interventions.

5 Conclusion

In this research, we aimed to shed light on the intricate relationship between income inequality and crime levels, a long-debated issue among economists, politicians, and society. We aimed to explore the factors influencing this relationship by meticulously analyzing data gathered from various sources. Although our dataset spanned 11 countries from different regions of the world, we focused on three representative countries: the United States, Mexico, and Germany. This approach allowed us to assess income inequality and crime across distinct contexts without the redundancy that could arise from analyzing each country individually. Our primary focus was to determine whether a robust correlation between income inequality, as measured by the Gini coefficient, and crime rates, with homicide rates and robbery rates as its proxies, exists across these regions rather than to delve deeply into the unique sociopolitical or economic conditions of each country. As a future direction, scholars can conduct more granular, country-specific analyses, delving into the socioeconomic dynamics of each country to better understand the nuances behind the correlation and to theorize why the relationship between inequality and crime manifests differently across various contexts.

While our country-specific analyses occasionally diverged from our initial expectations and, in some cases, contradicted them, several plausible reasons could account for these discrepancies. Firstly, individual country analyses might be influenced by short-term fluctuations or anomalies in local economic conditions, governance, or public safety measures. For example, abrupt changes in economic policies, shifts in law enforcement strategies, or social unrest could disrupt crime patterns in ways that a single data snapshot might not fully capture. Additionally, the complexity of crime itself, which can be driven by myriad factors such as education, unemployment, cultural norms, and even urban planning, might dilute the apparent relationship with income inequality when viewed at the country level. Another hypothesis is that crime, particularly violent crime, might be more sensitive to regional or local disparities within a country rather than national income inequality measures. These factors could obscure the overall relationship between the Gini coefficient and crime in specific countries. Future research could address these issues by focusing on more localized analyses, utilizing extended time frames, or incorporating additional variables to understand country-specific dynamics better.

In contrast, our panel data analysis, which combined data from multiple countries over the period from 2003 to 2020, revealed a robust and statistically significant correlation between the Gini coefficient and crime rates. By leveraging data from across different regions, the panel approach allowed us to capture both between-country and within-country variations, offering a more comprehensive picture of the income inequality-crime relationship. The model explained a substantial portion of the variation in crime rates (with R-squared values of 0.813 for homicide and 0.592 for robbery), suggesting that income inequality plays a key role in influencing crime. The results indicated that as income inequality increases, crime rates—particularly homicide and robbery—rise significantly, supporting the hypothesis that economic disparities may foster social tensions or desperation, ultimately leading to higher crime rates. The significance of other factors like education and urbanization further highlighted the complex interplay between socioeconomic variables and crime. This panel analysis underscores the value of cross-country comparisons in revealing patterns that might not be as apparent at the individual country level, thus providing substantial support for the theories discussed in the introduction.

While our findings point to a robust overall link between income inequality and crime, future research could benefit from diving deeper into country-specific contexts to better understand the localized drivers of crime. Additionally, exploring the role of regional inequalities or conducting longer-term studies may yield further insights into this critical and complex issue.

References

[1] Pablo Fajnzylber, Daniel Lederman, and Norman Loayza. "Inequality and violent crime". In: *The journal of Law and Economics* 45.1 (2002), pp. 1–39.

[2] Gary S Becker. "Crime and punishment: An economic approach". In: *Journal of political economy* 76.2 (1968), pp. 169–217.

[3] The Economist. *The stark relationship between income inequality and crime*. 2018. URL: `https://www.economist.com/graphic-detail/2018/06/07/the-stark-relationship-between-income-inequality-and-crime` (visited on 08/11/2024).

[4] Luiz Guilherme Scorzafave and Milena Karla Soares. "Income inequality and pecuniary crimes". In: *Economics Letters* 104.1 (2009), pp. 40–42.

[5] Marvin D Krohn. "Inequality, unemployment and crime: A cross-national analysis". In: *The Sociological Quarterly* 17.3 (1976), pp. 303–313.

[6] Ngozi Bosede Adeleye. "The determinants of income inequality and the relationship to crime". In: *Yayınlanmış Yüksek Lisans Tezi, University of Sussex, United Kingdom* (2014).

[7] Stefan Harrendorf, Markku Heiskanen, and Steven Malby. *International statistics on crime and justice*. European Institute for Crime Prevention and Control, affiliated with the ..., 2010.

[8] Maia Szalavitz. "Income inequality and homicide". In: *Scientific American* 319.5 (2018), pp. 9–9.

[9] Daniel L Hicks and Joan Hamory Hicks. "Jealous of the Joneses: conspicuous consumption, inequality, and crime". In: *Oxford Economic Papers* 66.4 (2014), pp. 1090–1120.