The Long Shadow of Slavery: Land-Holding Inequality and Its Impact on Crime in the United States

ECON 191 Topics in Economics, UC Berkeley

Jacob Sharadin

Abstract

This study examines the causal impact of land inequality on violent crime in the United States, leveraging historical slavery as an instrumental variable to address endogeneity. Using county-level data from the 1860 U.S. Census and modern FBI crime statistics, I find evidence of a significant positive relationship between land inequality and violent crime. Robustness checks, including Generalized Method of Moments (GMM) and weighted bootstrap regression, support these findings. However, no significant relationship is observed between inequality and property crime. These results highlight the enduring influence of slavery in shaping inequality and its broader social consequences, with implications for contemporary policies addressing crime and systemic inequality.

Content

1 Introduction

1.1 Background

Economic theory suggests a positive relationship between inequality and crime, a concept first explored in the 1960s. Since then, researchers have sought to empirically verify this connection, though the evidence has often been mixed and inconclusive. At the same time, historical U.S. slavery created profound economic disparities that persisted for many African Americans long after the Civil War. Between 1774 and 1860, the Gini coefficient, a measure of inequality, increased from 0.441 to 0.529 (Shanahan, 2000). This legacy endures, as regions historically affected by slavery continue to exhibit significant land inequality and elevated crime rates. In 2023, for instance, violent crime rates in states like Mississippi and Louisiana were 25% higher than the national average.

There is extensive literature on the relationship between crime and inequality. Notable work by Kelly (2000) found that "inequality has no effect on property crime but a strong and robust impact on violent crime." Kelly implemented a Poisson regression model with GMM as a robustness check, as well as three separate models with instruments to address endogeneity concerns. The instruments used were capita income, expenditure share, and non-democratic vote share. Kelly's IV-GMM estimates, while weaker than their Poisson estimates, aligned with their conclusion. Similarly, Buonanno (2017) used slavery as an instrument to explore the effects of inequality and crime in Colombia, highlighting the relevance of historical institutions in shaping contemporary socioeconomic outcomes.

1.2 Research purpose and organization

This paper seeks to answer: How does the legacy of slavery influence contemporary land inequality and crime rates in the southern United States. Inspired by the work of Kelly and Buonanno, this paper aims to estimate the effect of legacy slavery on land inequality and crime rates. By focusing on land inequality, the study seeks to capture the deep-rooted economic imbalances that stem from slavery and influence various social outcomes, including crime, if they exist. The underlying assumption is that, in the absence of slavery, land distribution would have been more equitable, and contemporary crime rates would be lower. Since northern states had largely abolished slavery before 1860, this paper examines the regional differences in these effects.

This study employs an IV-Generalized Method of Moments (GMM) estimator to address potential endogeneity and uses historical slavery as an instrumental variable for land inequality. The identification strategy hinges on the strength of slavery as an exogenous instrument for land inequality and measuring the difference in crime rates in regions with varying levels of historical slavery.

Endogeneity is addressed due to concerns about reverse causality, omitted variable bias, and measurement error. Finally, potential threats to identification, such as violations of the exclusion restriction, are acknowledged and discussed.

The remainder of this paper is organized as follows. Section 2 provides an overview of the data and variables used in the analysis. Section 3 outlines the empirical strategy and methodology, including the IV-GMM framework and robustness checks. Section 4 presents the results, discussing the findings for both violent and property crime rates. Section 5 addresses limitations and potential threats to identification, such as measurement errors and omitted variables. Section 6 concludes by summarizing the findings and their implications for understanding the long-term socioeconomic consequences of slavery in the United States.

From a policy perspective, these findings can guide efforts to address systemic inequality, inform reparative justice initiatives, and develop evidence-based strategies to reduce crime in historically disadvantaged regions.

2 Literature Review

2.1 Related to slavery, inequality, and crime

Several studies have demonstrated how historical factors, including slavery, have long-lasting consequences on economic inequality. For instance, Acemoglu et al. (2001) show that colonial institutions designed to extract resources contributed to persistent inequality, while Nunn (2008) highlights how the legacy of slavery in the Americas continues to shape economic outcomes. These studies emphasize the enduring impact of institutional factors on wealth distribution and social structures.

Scholars such as Engerman and Sokoloff (2002) have specifically examined the relationship between slavery and land inequality. They argue that regions with a history of slavery tend to exhibit higher levels of land inequality, a trend that has persisted into the modern era. This relationship stems from the historical concentration of land ownership in the hands of a few, particularly in plantation economies. Consequently, regions heavily reliant on slave labor experienced greater disparities in land ownership, perpetuating broader forms of economic inequality.

Additionally, Buonanno and Vargas (2018) explored the relationship between economic inequality and crime in Colombian municipalities. They utilized the proportion of slaves in each municipality before abolition as an instrument for economic inequality, documenting a strong association between inequality and both violent and property crime. This framework demonstrates how historical variables, like slavery, can be used to study the long-term consequences of inequality. The study also highlights concerns about exogeneity, noting that slavery might have left behind cultural or institutional legacies that influence crime independently of inequality.

A relevant body of research examines the relationship between inequality and crime. For example, Fajnzylber et al. (2002) provide empirical evidence of a positive correlation between income inequality and crime rates. However, empirical studies on this topic often yield mixed results, as establishing causality is challenging. A recent meta-analysis by Pazzona (2024) revisits the inequality-crime puzzle, analyzing 1,341 estimates from 43 studies. It concludes that while there is a statistically significant relationship, the effect sizes are economically insignificant, ranging from 0.007 to 0.123. This suggests that inequality may not be the primary driver of crime and underscores the need to address potential biases in empirical studies.

Furthermore, Glaeser et al. (1996) emphasize that confounding variables, such as regional differences in economic conditions, social policies, and law enforcement, can obscure the true effects of inequality on crime. These structural factors make it difficult to draw causal inferences without addressing endogeneity and omitted variable bias.

2.2 Related to econometric methods

My research builds on this existing literature by employing an instrumental variables (IV) approach, using the proportion of enslaved individuals in the 19th century as an instrument for land inequality. This approach isolates exogenous variation in land inequality, allowing for a more accurate estimation of the causal effect of inequality on crime.

Additionally, I use the Generalized Method of Moments (GMM-IV) technique to address endogeneity concerns and ensure the robustness of my findings. GMM-IV is particularly effective in dealing with weak instruments and unobserved heterogeneity, strengthening the validity of causal inferences. This method allows me to account for potential biases and inconsistencies that might arise from omitted variable bias, reverse causality, and measurement error.

By focusing on the U.S. context, my research extends the existing analysis of historical inequality and its long-term effects. Using U.S. census data, I explore how slavery's legacy shapes contemporary inequality and crime, contributing to the gap in studies examining the long-term impact of historical factors on modern socioeconomic outcomes.

3 Empirical Strategy

3.1 IV regression

I start by outlining the model and methodology I intend to use for studying the effect of historical slavery as an instrument for land inequality on crime. My analysis focuses on U.S. counties $i \in \{1, \ldots, n\}$, where the land inequality and crime rates are observed.

I define $Slavery_i$ as the proportion of enslaved individuals in the 19th century in county i, which I use as an instrument for land inequality LI_i. The key assumption is that slavery shaped long-term land ownership patterns and wealth distribution, creating disparities that persisted into the modern era. Crime rates $crime_i$ are observed for each county, and I model these as a function of predicted inequality $\hat{LI}_i$ derived from the instrument $Slavery_i$, as well as additional control variables.

My first equation for land inequality is:

$$LI_i = \phi_1 Slavery_i + \lambda^T X_i + u_i \tag{1}$$

Where LI_i is land inequality, X_i is a vector of control variables including population density, young males, and economic activity, and u_i represents the unobserved error term. For crime, I use the second-stage equation:

$$crime_i = \alpha + \beta_1 \hat{LI}_i + \delta^T X_i + \epsilon_i \tag{2}$$

Where $crime_i$ is the crime rate (both violent and property crimes), $\hat{LI}_i$ is the predicted inequality from the first-stage regression, and ϵ_i is the error term. The coefficient of interest β_1 captures the causal impact of land inequality on crime.

As a measure of land inequality I calculate the Gini coefficient:

$$LI_i = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|}{2n^2 \bar{y}} \tag{3}$$

Where:

$$y_i, y_j : \text{Cash value of farms for different observations,}$$
$$n : \text{Number of observations,}$$
$$\bar{y} : \text{Mean cash value of farms in the county.}$$

The Gini coefficient is calculated in the same manner in which it is done in the previous study by

Buonanno and Vargas (2018). It is calculated as the current distribution of the value of landholdings in our counties.

To account for potential heteroskedasticity and autocorrelation in the error terms, I employ the Generalized Method of Moments (GMM) technique. This method allows us to efficiently estimate parameters while accounting for potential endogeneity. The GMM estimator solves the following minimization problem:

$$\hat{\beta_1} = \min_{\theta} \left(g(\theta)' W g(\theta) \right) \tag{4}$$

Where:

- θ represents the parameters to be estimated, which include ϕ_1 and β_1,
- $g(\theta)$ is the vector of moment conditions, typically the instruments (such as $Slavery_i$) multiplied by the residuals from the model,

$$g(\theta) = \frac{1}{n} \sum_{i=1}^{n} Z_i (Y_i - \beta_0 - \beta_1 X_i) \tag{5}$$

- W is a weighting matrix, chosen as the inverse of the covariance matrix of the moment conditions.

The moment conditions for the GMM-IV estimator are derived from the orthogonality between the instruments Z (in this case, slavery data) and the error terms. These conditions can be written as:

$$E[Z_i(\epsilon_i)] = 0 \tag{6}$$

This ensures that the instruments Z are uncorrelated with the error term ϵ, allowing us to consistently estimate the parameters of interest.

3.2 Robustness checks

One of the critical issues in this analysis is ensuring the validity of the exclusion restriction. The proportion of slaves in each county may influence contemporary inequality, but I assume that its direct effect on crime operates only through its impact on land inequality. This addresses concerns about confounding factors, as slavery's effect on crime is mediated by economic disparities that have persisted over time. However, I acknowledge the potential violation of this exclusion restriction, as some historical institutional legacies may affect crime directly, and plan to test the robustness of my results by relaxing this assumption.

To address potential biases arising from differences in county sizes or characteristics, we assign weights w_i to each county i based on a relevant factor, such as population or land area. The weights are calculated as follows:

$$w_i = \frac{x_i}{\sum_{i=1}^{N} x_i} \tag{7}$$

Where:

- x_i represents the factor of interest for county i (e.g., population size),
- N is the total number of counties.

This ensures that the weights sum to 1:

$$\sum_{i=1}^{N} w_i = 1 \tag{8}$$

The weighted regression model is then estimated as:

$$\hat{\beta} = \left(X^T W X\right)^{-1} X^T W Y \tag{9}$$

Where:

- W is a diagonal matrix with weights w_i on the diagonal,
- X is the matrix of independent variables,
- Y is the vector of dependent variables.

To assess the robustness of parameter estimates, we perform bootstrapping by generating B bootstrap samples, where B is a large number (e.g., 1,000). Each bootstrap sample is drawn with replacement from the original dataset, ensuring the sample size remains $N = 130$. The procedure is as follows:

(a) Let the original dataset be $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$.

(b) For each bootstrap iteration $b = 1, 2, \ldots, B$:

   i. Randomly sample N observations with replacement to create a bootstrap sample $\{(x_1^b, y_1^b), \ldots, (x_N^b, y_N^b)\}$.

   ii. Estimate the model (e.g., regression) on the bootstrap sample and store the parameter estimates $\hat{\beta}^b$.

After B iterations, the parameter estimates are aggregated as follows:

- The mean estimate:
$$\bar{\beta} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}^b \tag{10}$$

- The standard error:
$$\mathrm{SE}(\beta) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\beta}^b - \bar{\beta}\right)^2} \tag{11}$$

- The confidence intervals, calculated using the percentile method (e.g., 2.5th and 97.5th percentiles of $\hat{\beta}^b$) or standard normal approximation.

To combine reweighting and bootstrapping, we adjust the probabilities of selecting each county during resampling based on the weights w_i. The probability of selecting county i is proportional to its weight:

$$P(i) = w_i \tag{12}$$

The regression model is then estimated on each weighted bootstrap sample, and the variability of estimates is analyzed as described above.

4 Data

4.1 Data sources

I am interested in data on land inequality, crime rates, and historical slavery in the U.S. to assess the relationship between inequality and crime. For historical slavery data, I utilize census-based records available through the Integrated Public Use Microdata Series (IPUMS), which provides detailed information on the proportion of enslaved individuals in each municipality prior to the abolition of slavery. To measure land inequality, I draw from the U.S. Department of Agriculture (USDA) records on land ownership distribution, particularly focusing on regions historically affected by slavery. Crime data, including violent and property crime rates, was be sourced from the Federal Bureau of Investigation's Uniform Crime Reporting (UCR) program, specifically using data from 2010. The 2010 crime data was chosen for its completeness, comparability across counties, and relevance in capturing the contemporary consequences of historical inequalities. The 2010 data offers a comprehensive snapshot of modern crime rates, enabling analysis of the long-term socioeconomic impacts of historical institutions such as slavery. To control for potential confounding factors, I incorporate economic indicators such as GDP and income inequality, both accessible through the U.S. Bureau of Economic Analysis (BEA) and the U.S. Census Bureau. Additionally, data on demographic factors, including population distribution and urbanization, was be sourced from the U.S. Census Bureau to ensure a comprehensive analysis of the relationship between land inequality and crime.

4.2 Data limitations & assumptions

There are several challenges and limitations in the data required for this study. The key measure of inequality I use is the land Gini coefficient, as I want to avoid switching measures of inequality between time periods. While income inequality data became available in U.S. Census data starting in 1940, it is not possible to use income quintiles as a measure for earlier periods. This makes land inequality the most consistent metric for assessing the relationship between historical slavery and contemporary crime. For 1860, land inequality is measured using data from the Census of Agriculture, which provides detailed information on land ownership, farm acreage, and farm values. However, using this historical data introduces challenges, including missing or inconsistent data for certain states or territories. Additionally, changes in state boundaries and definitions between the 19th century and today complicate data merging. Nevertheless, I calculate the land Gini coefficient for each state and territory based on the 1860 census and maintain consistency by using the same measure in modern periods.

The dataset from the 1860 Census of Agriculture includes approximately 2,000 observations from counties across the U.S. The primary variables include land value, total population, and the number of enslaved individuals. These are the three key variables of interest for the study. In addition, there is 3 control variables, such as population density, the proportion of young males aged 15-29, and GDP per county. These variables are crucial to control for factors that may also influence crime and inequality, ensuring a more accurate estimation of the relationship between slavery and land inequality. This gives us a total of around nine variables of interest for the analysis.

I merged the datasets, each containing information on their respective variable at the county level. This allowed me to examine the relationship between slavery and land inequality across regions. The merging process posed challenges, particularly in ensuring consistency across datasets due to changing names, boundaries, and administrative definitions of counties between the 19th century and modern periods. However, using counties as the unit of analysis allowed for a larger sample size and potentially more granular insights.

The data is generally up to the task of exploring the long-term relationship between slavery, land inequality, and crime, but requires careful handling. While the large number of observations increases the statistical power of the analysis, I still expect measurement errors, particularly in the land value data from the 1860 Census of Agriculture, where land-recorded value may have been under reported. Additionally, inconsistent reporting between regions may create some gaps in the dataset.

The small sample size also presents a challenge. I use a weighted bootstrap regression to address potential biases and ensure robust inference when running regressions. This approach helps to mitigate issues arising from the small sample size by resampling the data with weights to account for variability and outliers. Further refinement and sensitivity checks are essential. I also perform robustness checks using GMM (Generalized Method of Moments) to ensure that the model is well-specified and to account for potential endogeneity in the regression model.

Once the land Gini coefficients are calculated, I regress land inequality on the number of slaves per county in 1860. I expect the regression to produce a positive relationship, meaning counties with higher numbers of enslaved individuals exhibit greater land inequality. This aligns with the historical understanding that slavery contributed to the concentration of land ownership in the hands of a few, creating enduring economic disparities.

Despite the dataset's size, potential challenges with statistical significance persist, as outliers may disproportionately influence the estimates. While weighted bootstrap helps address some of these concerns, careful interpretation of the results remains necessary. Exploring alternative models or incorporating additional control variables may further enhance the robustness of the findings. Given the dataset's constraints and possible model limitations, ongoing refinements are crucial to ensure the reliability of the conclusions.

In this research, several challenges and limitations have emerged, particularly regarding data consistency and measurement accuracy. Calculating key measures like the Gini coefficient for land inequality revealed inconsistencies. Specifically, calculations using different identifiers—county codes versus county names—produced varying results, highlighting potential errors in how the data was grouped or merged. These issues likely stem from mismatches between county names and codes or duplicated or improperly aggregated data entries.

Additionally, some basic regressions did not consistently yield statistically significant results for expected relationships, such as the positive correlation between slavery and land inequality. This may be attributed to missing data, incorrectly merged variables, or outliers in the dataset. For example, certain counties reported extreme land values or population counts, disproportionately influencing the Gini coefficient calculations and regression outputs.

Measurement error presents another limitation, particularly with historical data from the 19th century. Land values recorded in the Census of Agriculture, for instance, may have been underreported due to tax evasion or misreported due to incomplete record-keeping practices. These inaccuracies could introduce bias into the Gini coefficient calculations, affecting the validity of the regression models.

To address these challenges, I standardized county codes and names across all datasets to ensure alignment and corrected discrepancies to ensure accurate representation of all counties. Duplicate records and improperly aggregated data were removed, and unique observations were retained for analysis. Outliers in the land value data were identified and handled appropriately to minimize their influence on the results.

4.3 Data overview

The dataset preparation process began with a meticulous preprocessing phase. Figure 1 highlights some of the key variables later used in my OLS, IV and IV-GMM regressions, with the results displayed in Tables 1 and 2. To create the finalized dataset, summarized in Figure 1, I filtered and combined data from several sources: the 1860 Schedule 1 and Schedule 2 Census, the 1860 Agricultural Census, the 2010 FBI UCR crime data, and the 2010 US Census data.

To measure land inequality, I calculated the Gini coefficient of the total cash value of farms at the county level using Stata's ineqdeco function, a widely recognized method for computing inequality indices. This function also enables the decomposition of inequality into within-group and between-group components, which was particularly useful for analyzing variations in inequality across counties. I leveraged the within-group decomposition to generate county-specific Gini coefficients, which serve as a key independent variable in my analysis.

During this process, I observed that many counties had a Gini coefficient value of zero. Suspecting errors due to missing data, I decided to exclude these counties. I further ensured data consistency by removing observations where the cash value of farms was zero, as these counties often had a population of zero as well. However, there was one anomaly: a county with a population exceeding 60,000 but a cash value of zero. To avoid skewing the results, I excluded this county from the analysis. Ultimately, I only included counties with positive population counts and non-zero cash values to maintain the integrity of the Gini coefficient calculations. This initial Gini-slavery dataset contained 282 counties, which I later refined.

Next, I incorporated population and geographic data from the 2010 Census to calculate population density for both 1860 and 2010. Since detailed historical county boundary data was unavailable, I used 2010 county area measurements (total square miles of land and water) as a proxy for the 1860 county areas. Additionally, I used 2010 population data to compute the number of individuals aged 15–29.

At this stage, I was left with a refined dataset, ready for exploratory analysis.

Table 1: Summary Statistics of Key Variables

Statistic	log_slave	gini_stata	log_murder_rate	log_mtv_theft_rate	log_crime_rate
Count	126.000	126.000	126.000	126.000	126.000
Mean	6.780	0.469	1.095	2.732	8.159
Std	1.194	0.144	0.987	1.048	0.580
Min	2.565	0.019	0.000	0.000	5.669
25%	6.057	0.418	0.000	2.468	7.912
50%	7.104	0.502	1.307	2.904	8.213
75%	7.661	0.568	1.815	3.364	8.513
Max	8.502	0.716	3.579	4.717	9.549

Note: This table provides summary statistics for selected variables in the dataset.

Figure 4 below illustrates the relationship between the Gini coefficient and the log of the murder rate for an initial inspection. Many values for the log murder rate are zero. At this stage, it is unclear whether these zeros represent true values or result from measurement errors. A similar issue arises for the log motor vehicle theft and overall log crime rates. Cross-referencing with the FBI UCR data suggests that for some counties, no data is available, leading me to lean towards interpreting these zeros as measurement errors.

During my exploratory data analysis (EDA), I examined the distributions and relationships within the dataset, focusing on land inequality, slavery, and crime statistics. To analyze wealth distribution across counties, I calculated a Lorenz curve based on the cumulative cash value of farms. The curve revealed that roughly 20% of the population controlled approximately 54% of the total cash value of farms—deviating from the conventional "80/20" Pareto Principle. While this principle often applies in various economic contexts, the findings here indicate a somewhat more balanced, though still substantial, concentration of wealth. This suggests that historical systems of inequality, such as slavery, have likely contributed to enduring disparities in wealth and land ownership.

I then explored the Gini coefficients across counties by plotting their regional distribution (Figure 3). The data showed notable variations: counties indexed above 100 generally exhibited lower Gini coefficients, indicating less inequality. To gain deeper insights, I applied k-means clustering to the Gini values, which revealed distinct clusters of counties with varying levels of inequality. This clustering provided a more nuanced understanding of regional disparities in land inequality.

5 Results

5.1 Baseline analysis: OLS regression

Ordinary Least Squares (OLS) regression was used to establish a baseline relationship between economic inequality and different crime types, focusing on log murder rates, log motor vehicle theft rates, and log overall crime rates as dependent variables. The models included control variables such as population density and the proportion of young males aged 15–29 to account for their association with crime. The results revealed a strong positive and statistically significant relationship between economic inequality and murder rates, with a 1-unit increase in economic inequality associated with a 1.2026-unit increase in the log of the murder rate (p ¡ 0.05). This relationship remained robust after controlling for other factors.

In contrast, the coefficients for economic inequality were positive but not statistically significant for motor vehicle theft (e.g., coefficient = 0.3534, SE = 0.697) and overall crime rates (e.g., coefficient = 0.5815, SE = 1.066). These findings suggest that economic inequality has a more pronounced effect on violent crimes like murder compared to property crimes or overall crime rates. However, it is important to note that OLS does not address potential endogeneity, such as omitted variables or reverse causality, which could lead to biased estimates and limit the causal interpretation of the results.

Table 2: Effect of Land Inequality on Crime, Ordinary Least Squares

	(1)	(2)	(3)	(4)
Panel A. Dependent variable: Log murder rate				
Ec. Inequality	1.2282**	1.2870**	1.3693**	1.3696**
	(0.604)	(0.596)	(0.610)	(0.612)
Panel B. Dependent variable: Log vehicle theft rate				
Ec. Inequality	0.0813	0.1132	0.2756	0.2756
	(0.651)	(0.652)	(0.664)	(0.667)
Panel C. Dependent variable: Log crime rate				
Ec. Inequality	-0.0307	-0.0365	0.1050	0.1069
	(0.361)	(0.362)	(0.366)	(0.365)
Controls:				
Pop. density		Y	Y	Y
% males aged 15-29			Y	Y
Economic activity				Y
Observations	126	126	126	126

Notes: Ordinary Least Squares regression. Standard errors in parentheses. Regression sample given by counties as existing in 1860, before the federal abolition of slavery. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

5.2 Main Analysis: IV Regression

Instrumental Variables (IV) regression was employed to address potential endogeneity, using historical slavery as an instrument for economic inequality. This approach aimed to isolate exogenous variation in inequality, and the first-stage regression confirmed the instrument's strength in explaining land inequality. The models included the same control variables as in the OLS analysis, such as population density and the proportion of young males aged 15–29.

To better understand the real-world implications of the coefficient, we back-transformed the results to express the impact of land inequality on murder rates in percentage terms. The coefficient for the log murder rate in the IV regression was 3.94. This implies that a 1-unit increase in land inequality is associated with a 294% increase in the murder rate ($\exp 3.94 - 1$).

To provide a more intuitive interpretation, I scaled down the coefficient to examine the effect of a smaller, more realistic change. A 0.1-unit increase in land inequality corresponds to an increase in the log murder rate of 0.394. Back-transforming this change, we find that a 0.1-unit increase in land inequality is associated with a 48.3% increase in the murder rate ($\exp 0.394 - 1$).

These findings highlight the substantial impact of even small changes in land inequality on violent crime, underscoring the role of inequality as a significant driver of social outcomes.

Despite these findings, the analysis is subject to limitations. The exclusion restriction assumption could be violated if historical slavery affects crime rates through pathways other than inequality, such as cultural or institutional legacies, which would weaken the causal interpretation of the results.

Table 3: Effect of Land Inequality on Crime, Instrumental Variables

	(1)	(2)	(3)
Second Stage			
Panel A. Dependent variable: Log murder rate			
Ec. Inequality	3.4827***	3.9461***	3.9356***
	(0.9542)	(1.0179)	(1.0112)
Panel B. Dependent variable: Log vehicle theft rate			
Ec. Inequality	0.9179	1.4388	1.4355
	(1.3533)	(1.4477)	(1.4366)
Panel C. Dependent variable: Log crime rate			
Ec. Inequality	-0.4371	-0.1059	-0.1619
	(0.6743)	(0.6630)	(0.6652)
First Stage			
Panel D. Dependent variable: Land Inequality			
Log num. slaves 19th cent.	0.0634***	0.0587***	0.0594***
	(0.010)	(0.010)	(0.013)
Controls:			
Pop. density	Y	Y	Y
% males aged 15-29		Y	Y
Economic activity			Y
Observations	126	126	126

Notes: Ordinary Least Squares regression. Standard errors in parentheses. Regression sample given by counties as existing in 1860, before the federal abolition of slavery. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

5.3 Robustness Checks

Generalized Method of Moments (GMM) was applied to provide additional robustness to the IV regression results, particularly addressing the potential heteroskedasticity in the error terms. Given the just identified nature of the model, where the number of instruments matches the number of endogenous variables, GMM produced coefficient estimates identical to the standard IV regressions. This outcome aligns with econometric theory, as the GMM reduces to IV in just-identified settings, regardless of the weighting matrix used.

While the coefficients were identical, the standard errors differed slightly due to the robust weighting matrix employed in the GMM framework. Both methods confirmed a strong positive effect of economic inequality on murder rates, while coefficients for motor vehicle theft and overall crime rates remained positive but not statistically significant. These results reaffirm the robustness of the findings and the strength of the instrument in addressing endogeneity.

By applying GMM, the analysis shows that the results hold across different estimation techniques, further strengthening the causal interpretation of the relationship between inequality and crime.

Table 4: Effect of Land Inequality on Crime, Instrumental Variables - GMM

	(1)	(2)	(3)
Second Stage			
Panel A. Dependent variable: Log murder rate			
Ec. Inequality	3.4827***	3.9461***	3.9356***
	(0.9542)	(1.0179)	(1.0112)
Panel B. Dependent variable: Log vehicle theft rate			
Ec. Inequality	0.9179	1.4388	1.4355
	(1.3533)	(1.4477)	(1.4366)
Panel C. Dependent variable: Log crime rate			
Ec. Inequality	-0.4371	-0.1059	-0.1619
	(0.6743)	(0.6630)	(0.6652)
First Stage			
Panel D. Dependent variable: Economic Land Inequality			
Log num. slaves 19th cent.	0.0634***	0.0587***	0.0594***
	(0.010)	(0.010)	(0.013)
Controls:			
Pop. density	Y	Y	Y
% males aged 15-29		Y	Y
Economic activity			Y
Observations	126	126	126

Notes: Standard errors in parentheses. Regression sample given by counties as existing in 1860, before the federal abolition of slavery. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

A weighted bootstrap regression method was applied to assess the robustness of the IV regression results, particularly under finite-sample conditions. This approach aimed to provide more reliable inference by addressing potential biases in small samples through weighted resampling of the data.

The bootstrap results confirmed the significant positive relationship between economic inequality and murder rates, consistent with the findings from both IV and GMM regressions. However, as with the other methods, the coefficients for motor vehicle theft and overall crime rates remained statistically insignificant, reinforcing the conclusion that inequality has a stronger effect on violent crimes like murder than on property or overall crime rates.

Despite its strengths, weighted bootstrap regression does not fully address limitations related to measurement error in historical data or potential violations of the exclusion restriction. These persistent challenges highlight areas for further refinement and validation in the analysis.

Table 5: Effect of Land Inequality on Crime, Instrumental Variables - Weighted Bootstrap

	(1)	(2)	(3)
Second Stage			
Panel A. Dependent variable: Log murder rate			
Ec. Inequality	3.6641	4.2014	4.2024
	(1.1645)	(1.2634)	(1.2634)
Panel B. Dependent variable: Log vehicle theft rate			
Ec. Inequality	1.0674	1.6576	1.6750
	(1.4846)	(1.6012)	(1.6028)
Panel C. Dependent variable: Log crime rate			
Ec. Inequality	-0.4113	-0.0403	-0.0653
	(0.7569)	(0.7461)	(0.7470)
First Stage			
Panel D. Dependent variable: Land Inequality			
Log num. slaves 19th cent.	0.0638	0.0589	0.0597
	(0.0103)	(0.0105)	(0.0162)
Controls:			
Pop. density	Y	Y	Y
% males aged 15-29		Y	Y
Economic activity			Y
Iterations	1000	1000	1000

Notes: Standard errors in parentheses. Regression sample given by counties as existing in 1860, before the federal abolition of slavery. Significance codes excluded. Estimates provides for comparison.

6 Conclusion

This study provides evidence of a strong relationship between economic inequality and violent crime, particularly murder rates, using historical slavery as an instrument for inequality. While the findings are robust across various methods, the validity of the exclusion restriction remains a key concern. The exclusion restriction assumes that slavery affects crime solely through its impact on economic inequality. However, historical evidence suggests that slavery's legacy influenced crime through additional mechanisms, potentially violating this assumption.

For instance, the Ku Klux Klan (KKK), which emerged in the aftermath of slavery's abolition, serves as a notable example. Founded to maintain racial hierarchies and intimidate African Americans, the KKK directly contributed to violence and unrest in the South. These actions were not mediated by economic inequality but rather by the cultural and institutional legacies of slavery, such as systemic racism and weak governance. Similarly, practices like lynching, racially motivated vigilantism, and the development of a culture of violence in the South represent other pathways through which slavery may have influenced crime rates independently of inequality. These mechanisms challenge the assumption that the instrument exclusively affects the outcome through the endogenous variable.

Research by historians like Eric Foner and economists such as Shari Eli and Trevon Logan, and Boriana Miloucheva has highlighted how the abolition of slavery and the subsequent Reconstruction period gave rise to institutions like the KKK that perpetuated violence and oppression beyond economic inequality. For example, Logan and Miloucheva (2019) document how the KKK and related forms of violence significantly impacted voter suppression and law enforcement practices, likely shaping crime rates through channels unrelated to inequality. These findings suggest that the legacy of slavery fostered direct cultural and institutional mechanisms that could confound the instrument's validity.

Another potential limitation lies in the magnitude of the observed relationship. The coefficient for land inequality on violent crime, expressed as a percentage increase in murder rates, appears exceptionally large compared to results from similar studies. This may reflect, in part, the effect of log-transforming the murder rate, which can exaggerate the magnitude of the relationship when underlying values are small. While log transformation enhances interpretability and addresses skewness, it may overstate the practical implications of inequality on crime, particularly when interpreting results in real-world terms.

If the exclusion restriction is violated or if the magnitude of the relationship is overestimated due to methodological choices, the coefficient estimates may be biased, limiting the study's ability to isolate the true causal effect of inequality on crime. Consequently, the findings might overstate the role of inequality in driving crime rates, and additional research is necessary to address these concerns.

Future studies could explore alternative instruments that better isolate the causal pathway of economic inequality on crime or use different transformations—or no transformations—on the crime data to verify the robustness of the findings. Incorporating additional robustness checks to address potential confounding mechanisms and improving the measurement of historical data could also provide greater clarity. By acknowledging and addressing these limitations, subsequent research can strengthen the causal interpretation of these findings and contribute to a deeper understanding of the long-term impacts of historical institutions on contemporary social outcomes.

7 Appendix

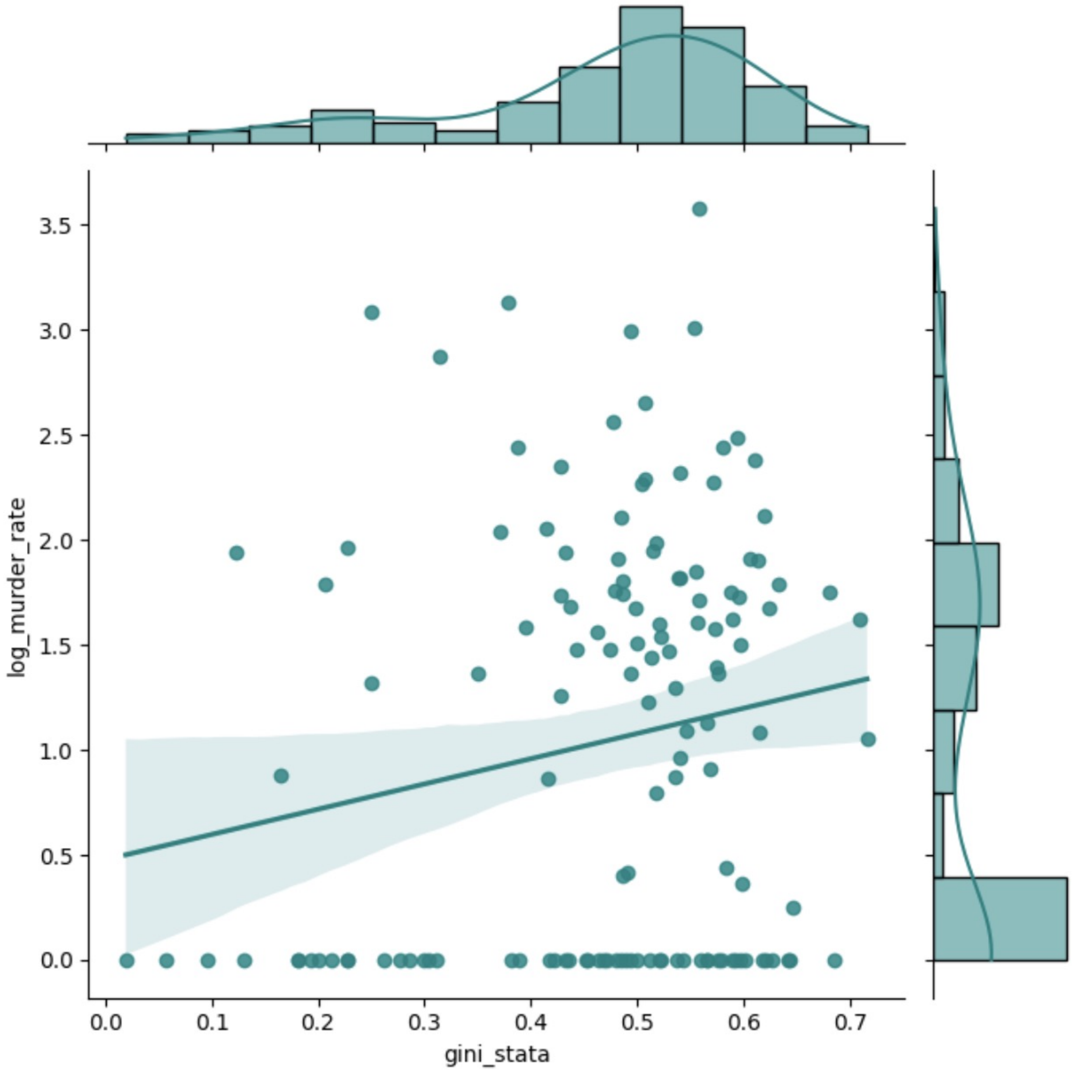

Figure 1:

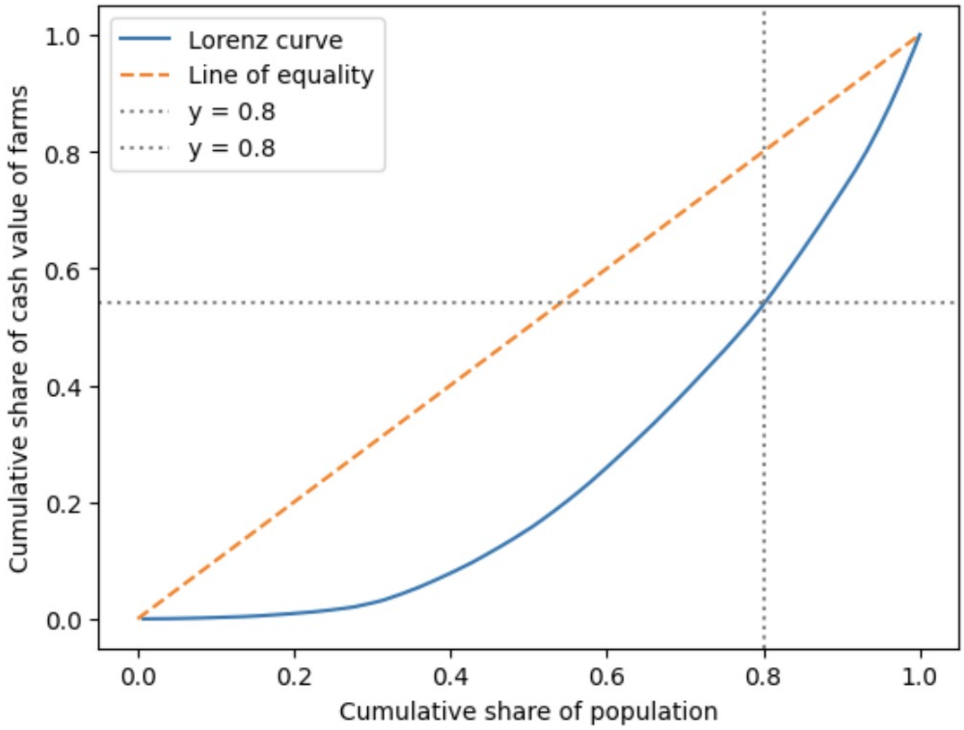

Figure 2:

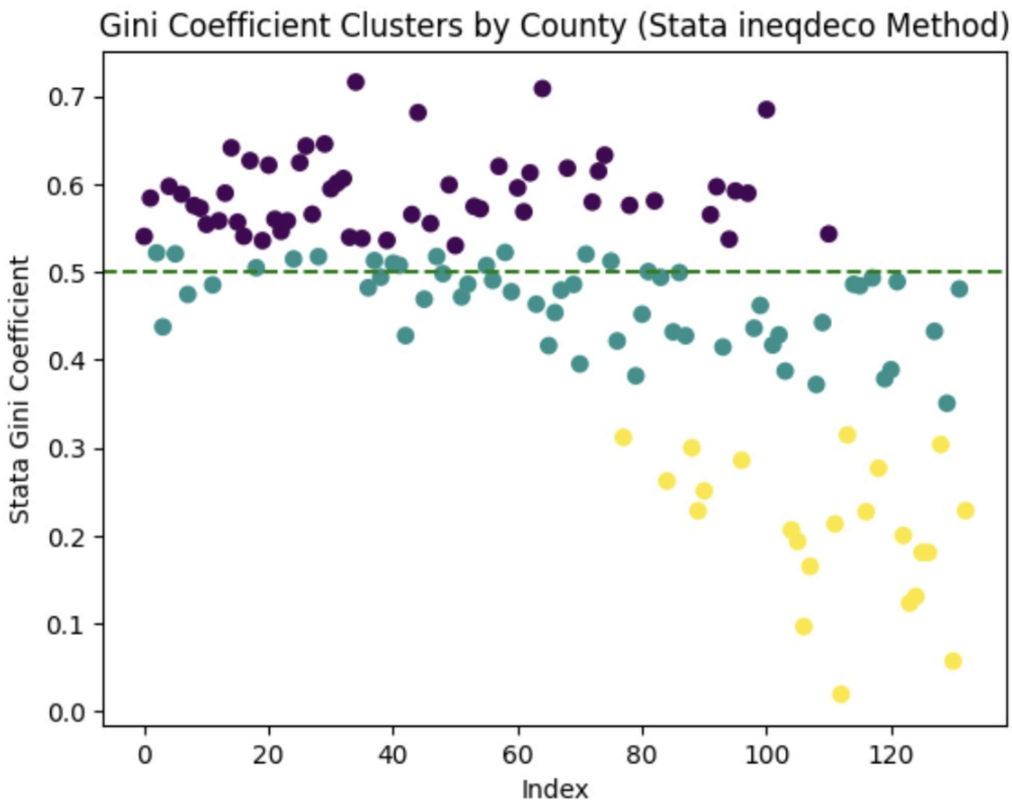

Figure 3: Figure 3

8 Bibliography

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). *The Colonial Origins of Comparative Development: An Empirical Investigation.* American Economic Review, *91*(5), 1369–1401.

Becker, G. S. (1974). *Crime and Punishment: An Economic Approach.* In G. S. Becker & W. M. Landes (Eds.), *Essays in the Economics of Crime and Punishment* (pp. 1–54). NBER. Retrieved from `http://www.nber.org/chapters/c3625`

Buonanno, P., & Vargas, J. F. (2019). *Inequality, crime, and the long run legacy of slavery.* Journal of Economic Behavior & Organization, *159*, 539–552. Retrieved from `https://doi.org/10.1016/j.jebo.2017.12.022`

Engerman, S. L., & Sokoloff, K. L. (2002). *Factor Endowments, Inequality, and Paths of Development Among New World Economies.* National Bureau of Economic Research Working Paper Series (No. 9259). Retrieved from `http://www.nber.org/papers/w9259`

Fajnzylber, P., Lederman, D., & Loayza, N. (2002). *Inequality and Violent Crime.* The Journal of Law and Economics, *45*(1), 1–39. Retrieved from `https://doi.org/10.1086/338347`

Kelly, M. (2000). *Inequality and Crime.* The Review of Economics and Statistics, *82*(4), 530–539. Retrieved from `http://www.jstor.org/stable/2646649`

Nunn, N. (2008). *The Long-term Effects of Africa's Slave Trades.* The Quarterly Journal of Economics, *123*(1), 139–176. Retrieved from `https://doi.org/10.1162/qjec.2008.123.1.139`

Pazzona, M. (2024). *Revisiting the Income Inequality-Crime Puzzle.* World Development, *176*, 106520. Retrieved from `https://doi.org/10.1016/j.worlddev.2023.106520`

Shanahan, M., Corell, M. (2000). How much more unequal? Consistent estimates of the distribution of wealth in the United States between 1774 and 1860. Journal of Income Distribution, 9(1), 3.