

Basic Examination of the Correlation between Crime Rates and Income Inequality

By

J. Travis Parsley

RESEARCH PAPER 2001-19

2000-2001 Undergraduate Research Fellow

Regional Research Institute

West Virginia University

Morgantown, WV 26506-6825

Mentor: William Trumbull, Ph.D.

Associate Professor of Economics

West Virginia University

August 2001

Abstract: This paper analyzes the significance of income inequality in determining crime rate. Researchers have frequently investigated this topic, albeit with inconsistent results. Guiding myself with previous research, I have attempted to conduct my analysis with simple, unambiguous techniques. I used county-level crime data and various census data to estimate the correlation between crime rates and several explanatory variables including income inequality. My results show that income inequality, as measured by the Gini coefficient, has a statistically significant effect on property crime rates as well as violent offense rates.

Background

Income inequality has been steadily trending upwards since the 1960s in the United States, making it an especially consequential issue. This trend will likely continue. As reason suggests, if everyone invested an equal percentage of their wealth and earned the same rate of return, the wealthiest individuals would outstrip the earnings of others by virtue of proportion alone. Other factors also bear influence. At least one scientist has hypothesized that as the proportion of women permanently entering the work force increases, income inequality will rapidly increase (Rivlin, 1975). If increasing income inequality leads to more crime, the United States cannot afford to ignore equity during the current rapid economic expansion. To address the problem, policy makers need to know the precise effect of inequality on crime. Government may need to reallocate its resources to efficiently quell participation in illegitimate activity. Welfare and taxation policies in particular may have unexpected consequences depending on the strength of the correlation between income inequality and crime rate. Even increased educational spending could have an adverse affect. Consider this finding: "...ceteris paribus, higher per capita educational expenditures tend to be associated with states that have income inequality which is greater than the U.S. average" (Bishop, et al., 1992). This implies that if income inequality significantly affects crime rates, increased educational spending may cause crime rate elevation.

What is income inequality? An income inequality measurement assesses the income distribution of a given population. If all measured units of population earn the same income, zero inequality exists. Conversely, a large earnings gap between high and low wage earners translates into high income inequality. How does this relate to crime? Fundamental economic

crime theory, specifically the work of Ehrlich (1973), suggests that low-income residents in disparate areas have greater incentive to commit crimes. Succinctly, potential criminals are increasingly likely to pursue illegitimate activities when surrounded by greater wealth. Thus, I will empirically test the hypothesis that income inequality significantly correlates with crime rate: populations exhibiting high income inequality will have higher crime rates than populations with less disparate income distributions.

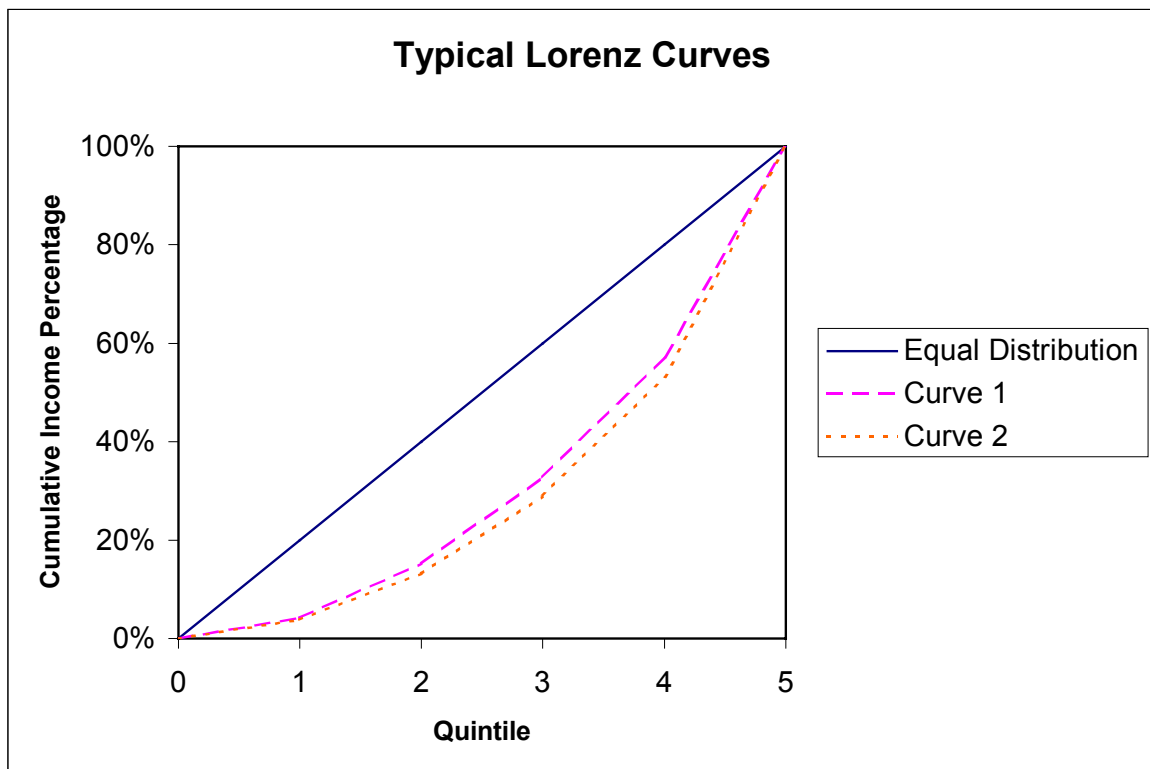
Previous Research

Many researchers have studied the relationship between income inequality and crime. Although scientists generally agree that an increase in income inequality will result in more crime, measurement inconsistencies have rendered research about the correlation inconclusive. For example, in his 1973 investigation of crime, Ehrlich found a “strong positive correlation between income inequality and crimes against property.” Ehrlich, however, used a nonstandard measurement of inequality, “the percentage of families below one-half of the median income” (1973). Clearly, Ehrlich’s measurement does convey income inequality, but it represents an arbitrary definition that hampers efforts to objectively compare his results with others.

According to Schutz, “Equality of income distribution is found when every income-receiving unit receives its proportional share of the total income” (1951). Researchers agree that when this linear distribution does not occur, income inequality exists. The main source of discrepancy lies within how scientists measure income distributions and define the “income receiving unit.”

Various types of income inequality measurements have gained acceptance with scientists. Economists most frequently use the Lorenz curve, and its derivative, the Gini index. The Lorenz

curve graphically represents income inequality. To create a Lorenz curve, the researcher first ranks each unit in the study population from poorest to richest. He then divides the ordered population into equally sized groups, usually quintiles, and determines the share of total income that each group earns. The quintiles are then plotted against the cumulative percentage of income. Ideally, each quintile would have the same income, and the resulting Lorenz curve would follow a line of slope 1. In actuality, however, moderate inequality exists and Lorenz curves deviate from the 45-degree line of equality. The magnitude of this deviation is proportional to the degree of income inequality in the study population.



Curve 2 represents a more unequal income distribution than Curve 1; it is farther from the 45-degree line
Source: (Gini says: measuring income inequality)

The Gini index is a coefficient that quantifies the gap between the 45-degree line and the actual Lorenz curve. A Lorenz curve mimicking the egalitarian 45-degree line would translate to

a Gini index of 0. In a perfectly unequal society, where one unit earned 100% of the total income, the Lorenz curve would resemble a backward L, spiking to 100% cumulative income at the far end of the x-axis. This would translate to a Gini index of 1.

In addition to selecting an appropriate measurement of income inequality, researchers also must make an important decision about which income-receiving unit to study. Possible study units include individuals, households, or family units. Researchers generally consider the individual level an illogical study choice. Rivlin suggests using either households or family units: “The distribution of income among individuals is meaningless because people live in groups and pool their incomes...” (1975). Unfortunately, no similar argument promotes the use of either family units or households over the other.

In my test of the effect of income inequality on crime rates, I utilize the Gini index to measure inequality. Several factors have influenced my decision. Economic studies have traditionally used the Gini inequality coefficient. Although it is tedious to calculate, Gini data is readily available and will provide a quantitative input for regression calculations, unlike the Lorenz curve. Also, because the United States Census Bureau has kept track of the national Gini index for the past fifty years, using the Gini coefficient in my study will allow relative comparisons of my study populations with the entire United States. My research utilized family unit Gini data for availability reasons.

Implementation

The crime data used in my analysis come from the FBI Uniform Crime Reporting Program. Specifically, I use reported Class I offenses as well as Class I offense arrests from the ninety most populous counties in the United States for 1990. By performing my analysis with

county-level data, I reach a low aggregation level while maintaining a sufficient population size. City or metropolitan-level data would likely yield inaccurate results due to the effect of anomalous crime rates during a specific study period. In a city of 50,000, a random killing spree might increase murders to fifteen, tripling the typical annual number of five. I have utilized data from very populous counties in order to help minimize the problem of abnormal yearly crime rates. Anomalous data are equally probable, but larger populations should lead to proportionally smaller errors.

The UCR data set includes the reported offenses and number of arrests for eight Class I crimes: murder, rape, robbery, assault, burglary, larceny, motor vehicle theft, and arson. Ehrlich found that income inequality had a greater effect on crimes against property than on violent crimes, presumably because income differences primarily affect the offender's benefits from property crime. Thus, I have divided the eight Class I crimes into property crimes (burglary, larceny, motor vehicle theft, and arson) and crimes against the person (murder, rape, robbery, and assault). For each county in the study, I have divided the crime statistics by county population to obtain per capita crime rates. This results in four categories per county: property crime reports, property crime arrests, violent crime reports, and violent crime arrests. This grouping will allow me to compare the correlation between income inequality and the two main offense types. The division between reported offenses and actual arrests is only meant to provide an extra set of data to analyze in case one set as a whole provides a better model. I do not make cross-category comparisons between violent crime arrests and property crime reports or vice versa.

I collected the 1990 Gini coefficient for each of the ninety subject counties. I then calculated each county's offense rates by dividing the categorized reported offenses by the

county's population. To determine which offenses are most affected by income inequality, and to what magnitude, I performed an ordinary least squares regression analysis on each county's crime rates. In addition to the Gini coefficient, I have used several other explanatory variables in an attempt to create a simple, yet viable model:

$$R_i = \beta_0 + \beta_1 APROB_i + \beta_2 G_i + \beta_3 U_i + \beta_4 M_i + \beta_5 NW_i + \beta_6 SP_i + \beta_7 ND_i + \beta_8 W_i + \varepsilon_i$$

<p>R = VARATE, VRRATE, PARATE, PRRATE</p> <ol style="list-style-type: none"> 1. VARATE = violent crime arrest rate 2. VRRATE = violent crime report rate 3. PARATE = property crime arrest rate 4. PRRATE = property crime report rate <p>APROB = VAPROB, PAPROB</p> <ol style="list-style-type: none"> 1. VAPROB = violent crime arrest probability 2. PAPROB = property crime arrest probability 	<p>G = Gini coefficient</p> <p>U = % urban population</p> <p>M = % male population</p> <p>NW = % nonwhite population</p> <p>SP = % population married with spouse present</p> <p>ND = % population without high school diploma</p> <p>W = % households on income assistance</p>
--	--

Generic regression model used in analysis. Four models are used, with R taking on each value and APROB taking on its appropriate value for each R.

VAPROB and PAPROB are popular explanatory variables, as evidenced by their inclusion in many past economic crime analyses (Cornwell, Trumbull 362). I used a simple approach for calculating their values. Each is simply the quotient of arrests per reported offense of the respective type, violent or property. (*The arrest rate for Jefferson County, Kentucky is an anomaly in that more arrests than offense reports were recorded.*) I leave out any variable measuring severity of punishment because of the difficulty in obtaining such data on a county level. I include U to account for the fact that opportunities for crime may present themselves more often in a densely populated urban environment than otherwise. M tests the hypothesis that males are more likely to commit crimes than females. Ehrlich placed importance on this possibility, as he included gender in several of his regression variables (544). Ehrlich also

included a variable measuring the percentage of nonwhites in the population. In his regressions, NW frequently has a positive coefficient. Ehrlich reasons that this reflects the “inferior legitimate market opportunities” available to minorities in the 1960s (548). I have included NW to see if possible prejudice persists in the job market thirty years later. The final two variables are included without the benefit of previous study. I have included SP with the presumption that its coefficient will be negative. I reason that if a person is married with a spouse present, he or she is under less financial obligation than otherwise. Such a situation eliminates several likely causes for financial stress such as single parenthood and alimony. While being married with a spouse present may present financial difficulties of its own, the number of potential household earners is doubled. Lastly, I intend W to represent financial need. Although a crude proxy, I expect households on assisted income to be more likely to desire an additional source of income than others.

Results and Implications

I include following account of modifications made while performing my regressions so the reader can have insight into my thought processes and understand that I have not attempted to “massage” my results to fit a certain hypothesis.

As I performed my first regressions on the data, I quickly realized which dependent variables were more feasible. When using the model to try to explain arrest rates, I came up with initially surprising results. The estimated coefficients for VAPROB and PAPROB were significant and positive. Upon reflection, I realized the fallacy in my regression. Arrest rates are directly correlated with the probability of arrest. Of course VAPROB and PAPROB would show a positive relationship with crime rates if using arrest data. Theory states that arrest probabilities

are inversely proportional to *offense rates*. Thus, the equations explaining offense rates are most significant.

The results of the offense rates regressions evidenced another change necessary to optimize my model. The t-statistics for W were low and insignificant at the 95% confidence level. Because I added this variable without theoretical support, I chose to drop it from the final regressions. The results of the optimized regressions are as follows:

Violent Offense Rate

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	0.022785944	0.017138275	1.329535	0.187404	-0.01131	0.056886	-0.01131	0.056886
VAPROB	-0.008582063	0.002470896	-3.47326	0.000828	-0.0135	-0.00367	-0.0135	-0.00367
G	0.025663263	0.010401016	2.46738	0.015718	0.004968	0.046358	0.004968	0.046358
U	0.000459313	0.008305502	0.055302	0.956034	-0.01607	0.016985	-0.01607	0.016985
M	-0.023650859	0.020164345	-1.1729	0.244272	-0.06377	0.01647	-0.06377	0.01647
NW	0.013346522	0.004309861	3.096741	0.002688	0.004771	0.021922	0.004771	0.021922
SP	-0.024791488	0.00831604	-2.98117	0.003791	-0.04134	-0.00825	-0.04134	-0.00825
ND	0.016012456	0.007371459	2.172223	0.032763	0.001346	0.030679	0.001346	0.030679

Regression Statistics

Multiple R	0.861389733
R Square	0.741992271
Adjusted R Square	0.719695307
Standard Error	0.003482462
Observations	89

Property Offense Rate

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-0.12188406	0.090172276	-1.35168	0.18024	-0.3013	0.057531	-0.3013	0.057531
PAPROB	-0.170308831	0.06154053	-2.76743	0.006999	-0.29276	-0.04786	-0.29276	-0.04786
G	0.198097156	0.052551235	3.7696	0.000309	0.093537	0.302658	0.093537	0.302658
U	0.051224552	0.041919805	1.221965	0.225265	-0.03218	0.134632	-0.03218	0.134632
M	0.170698709	0.103498712	1.649283	0.102963	-0.03523	0.376629	-0.03523	0.376629
NW	0.035189454	0.022074345	1.594134	0.114801	-0.00873	0.079111	-0.00873	0.079111
SP	0.035455332	0.042601962	0.832246	0.407717	-0.04931	0.12022	-0.04931	0.12022
ND	0.02127086	0.037201479	0.571775	0.569058	-0.05275	0.09529	-0.05275	0.09529

Regression Statistics

Multiple R	0.658181891
R Square	0.433203402
Adjusted R Square	0.38422098
Standard Error	0.017625368
Observations	89

In the violent crime regression, the coefficients of five variables are statistically significant at the 95% confidence level: VAPROB, G, NW, SP, and ND. The negative coefficient of VAPROB and the positive coefficients on NW and ND are supported by past research and economic theory. Criminals are less likely to participate in illegitimate activities as the probability of arrest increases. The fact that violent crime increases with the percentage of minority population may be attributable to racially motivated violence. Reduced legitimate market opportunities for those with less education probably contribute to an increase in crime as well. I speculate that the negative coefficient of SP relates to domestic violence. People should be less likely to wed and consider themselves “together” after incidents of domestic violence. Therefore, violent crime reports should be inversely proportional to the number of people who are happily married.

The positive coefficient of G, the Gini index, is the notable result of this regression. It supports the assertion that crime will increase with income inequality. This model suggests that as a county’s Gini coefficient of income inequality increases by 0.1, the number of reported violent offenses will increase by approximately 0.00257 per person. In a county of 100,000, this would imply an increase of 257 reported violent crimes.

Supporting economic theory, the property crime regression shows that income inequality has a more pronounced effect on property offense rates. The statistically significant coefficient of G is 0.198, an order of magnitude higher than its coefficient in the violent crime regression. This implies that reports of property crime would increase by 1,980 if a county of 100,000 people had a 0.1 increase in its Gini coefficient. It is also interesting to note that based on the estimated coefficients, probability of arrest acts as a better deterrent against property crime than

violent crime. Violent crimes committed in the “heat of the moment” without rational thinking may explain this finding.

In the property crime regression, neither NW, SP, nor ND has a significant coefficient at the 95% confidence level. This may imply that these factors do not play a role in determining property crime offense rates. Removing them however, did not significantly affect the somewhat low R^2 value of the regression. Also, since they were significant in the violent crime rate regression, I decided not to remove them. In both regressions, U and M have insignificant estimated coefficients, in spite of their theoretical importance. I believe that the structure of my data set hindered regression on these variables. Because my data set consisted of the ninety most populous counties, I presume that there wasn't enough variation in urban population to accentuate any effect that U might have on crime rates. Likewise, I doubt that there was significant variation in male population between the studied counties.

Conclusion

As national income inequality continues to increase, so will the number of scholarly attempts to understand its consequences. This research will constitute another exploration into the relationship between income distribution and criminal activity. Hopefully, my findings will reinforce the discoveries of others and national policy makers will recognize the role income inequality plays in aforementioned issues such as welfare and taxation policy, law enforcement, and educational spending.

In conducting this research, I have attempted to eliminate as many sources of potential error as practical. I believe that its simplicity is a virtue. However, several factors limit the productivity of this research. My limited statistical background presents an obstacle in

determining the precise relationship between crime and income inequality. Optimally, I would have investigated simultaneity and different regression methods more advanced than ordinary least squares. Also, analyzing more data could further my research. While available county crime data shows offense totals for many years, robust calculations would also require several years' worth of income data on the ninety examined counties. Despite its shortcomings, however, my cursory examination of the income/crime relationship will add to the library of research showing that income inequality does influence crime rates.

Bibliography

- Becker, G. S. 1968. "Crime and Punishment: An Economic Approach." The Journal of Political Economy, 76(2). 169-217.
- Bishop, J. A., J. P. Formby, and P. D. Thistle. 1992. "Explaining Interstate Variation in Income Inequality." The Review of Economics and Statistics, 74(3). 553-557.
- Cornwell, C., and W. N. Trumbull. 1994. "Estimating the Economic Model of Crime with Panel Data." The Review of Economics and Statistics, 76(2). 360-366.
- Ehrlich, I. 1973. "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation." The Journal of Political Economy, 81(3). 521-565.
- Rivlin, A. M. 1975. "Income Distribution—Can Economists Help?" The American Economic Review, 65(2). 1-15.
- Schutz, R. R. 1951. "On the Measurement of Income Inequality." The American Economic Review, 41(1). 107-122.
- "Gini says: measuring income inequality;"
<http://www.panix.com/~dhenwood/Gini_supplement.html>; (accessed 8 January 2001).