On measuring dependence of income distribution and crime

Juhani Mutikainen

School of Science

Bachelor's thesis
Espoo 7.10.2015

Thesis supervisor and advisor:

Asst. Prof. Pauliina Ilmonen

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
BACHELOR'S THESIS

Author: Juhani Mutikainen

Title: On measuring dependence of income distribution and crime

Date: 7.10.2015 Language: English Number of pages: 4+35

Degree programme: Engineering physics and mathematics

Supervisor and advisor: Asst. Prof. Pauliina Ilmonen

The causes and effects of income and wealth distribution are a current and politically important topic. For example, this year OECD has published the book "In It Together: Why Less Inequality Benefits All", which argues that the widening income and wealth gap in western countries leads to lower economic growth.
This thesis seeks to discover the strength of the relationship between income distribution and violent crime in western countries using statistical analysis from correlations to multiple linear regression. The results suggest there is a robust relationship between the two and unequal income distribution is likely one causing macro-level factor for violent crime.

Keywords: Income distribution, criminality, multiple regression, social policy

Preface

I want to thank Assistant Professor Pauliina Ilmonen for her excellent guidance.

Otaniemi, 7.10.2015

Juhani Mutikainen

Contents

1 Introduction

Economic inequality is an important topic that has many dimensions. In his popular book "Capital in the the Twenty-First Century", economist Thomas Piketty argues that the current western economic model leads to concentration of wealth, which has negative social and economic implications [1]. This year, the book "In It Together: Why Less Inequality Benefits All" published by OECD argues that the widening income and wealth gap in western countries leads to lower economic growth [7].

While often we are more interested in wealth distribution rather than income distribution, the lack of reliable data usually limits studies to observe the effects of income distribution. This thesis studies the dependence of income distibution and violent crime in western countries in the 21st century. In practice, we seek to find further evidence for the results of the previous study by Fajnzylber et al. [2]. While this thesis is more limited in scope, focusing merely on western countries allows us to ignore possible problems arising from differing societal structures.

First we study the dependence with correlations and graphical analysis for three different time periods. Several different indices are used to study the effect of income distribution, including Gini coefficent and percentile ratios. There are two studied crime types, homicides and robberies, and an index approximating overall crime rate combined from the two.

To account for other possible factors affecting both income distribution and crime, we also study the dependence with multiple linear regression. In multiple regression analysis, we limit the study to year 2012 and OECD countries. Other macro-level factors studied in the multiple regressions include gross national product, change in gross domestic product and education. For the effects of education we study both the share of adults with tertiary education and the enrolment rate of teenagers.

2 Background on income distribution and crime

2.1 Income distribution

To study the effects of income distribution, we need to have an index for the distribution that shows the relevant parts of the distribution as well as possible. The income discussed later is equivalised disposable income, that is total income of a household after taxes divided by the square root of household size. A household of four persons is assumed to have needs twice as large as that of one person. All members of the household then have the same equivalised disposable income. For more information on income distribution, see for example [4] and the references therein.

2.1.1 Lorenz curve and Gini coefficent

Perhaps the best-known and most used index for income distribution is the Gini coefficent. Gini coefficent is based on the Lorenz curve that is the graphical representation of the cumulative distribution of the probability function of income. Point x on the curve shows the share of income earned by bottom x % of population (when ranked from lowest to highest incomes). Gini coefficent describes how close the Lorenz curve is to the line of equality where everyone has the same income, essentially trying to capture the amount of inequality in the distribution. If A is the area between the Lorenz curve and the line of equality and B is the area below the Lorenz curve, Gini coefficent is defined as [13]

$$Gini = \frac{A}{A+B}. \tag{1}$$

Gini coefficent of 0 indicates total equality and coefficent of 1 results from one person earning all of total income. Lorenz curve and Gini coefficent are illustrated in figure 1.

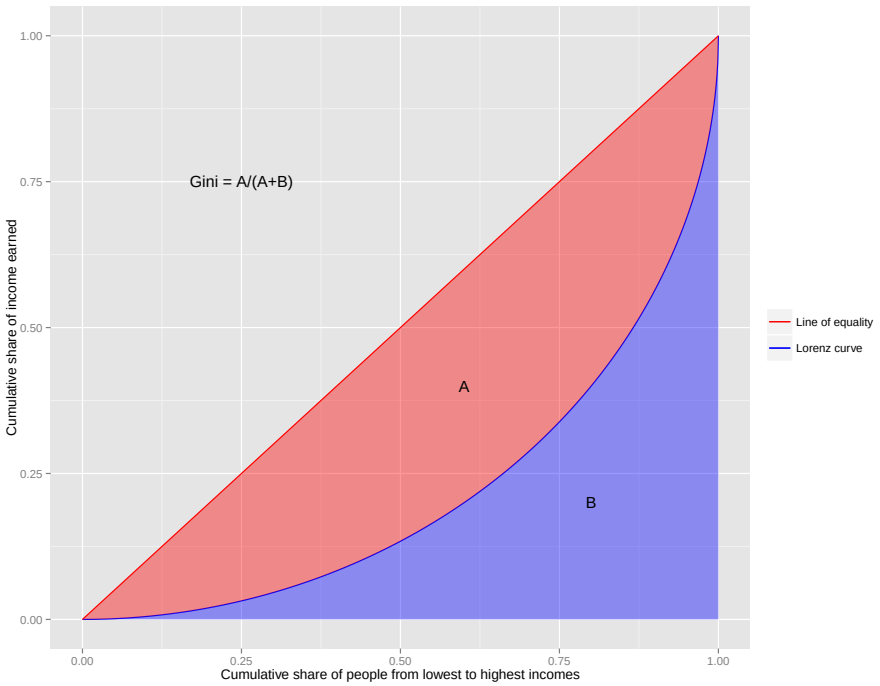

Figure 1: Lorenz curve and Gini coefficent

Despite the straightforward graphical definition of Gini, it can't be seen as a perfect or neutral indicator for the inequality of an income distribution. Different Lorenz curves may intersect and have the same Gini coefficent, so Gini coefficent can't tell whether the inequalities lie in the bottom or top of the income spectrum. The definition of Gini as an area also makes it highly sensitive to inequalities in the middle of the income spectrum. However, these qualities may be justified if they make it a better explanation for an effect (such as crime in this study).

2.1.2 Atkinson index

The Atkinson index is a measure of income distribution developed by and named after Anthony Barnes Atkinson, who thought that that inequality "cannot, in general, be measured without introducing social judgements. Measures such as the Gini coefficient are not purely 'statistical' and they embody implicit judgements about the weight to be attached to inequality at different points on the income scale". There is a parameter ε associated with the index. The Atkinson parameter can be thought of as inequality aversion, the amount of social utility that is assumed to be gained from complete redistribution of resources. When no social utility is to be gained by complete redistribution ($\varepsilon = 0$), Atkinson index $A_\varepsilon = 0$. On the other hand, when the social utility is infinite ($\varepsilon = \infty$), $A_\varepsilon = 1$. Usually values of 0.5, 1, 1.5 or 2 are used. The higher the value, the more sensitive the index becomes to inequalities at the bottom of the income spectrum.

The Atkinson index for a population of N persons is defined mathematically as

$$A_\varepsilon(y_1, ..., y_N) = \begin{cases} 1 - \frac{1}{\mu}(\frac{1}{N}\sum_{i=1}^{N} y_i^{1-\varepsilon})^{\frac{1}{(1-\varepsilon)}} & 0 \leq \varepsilon \neq 1 \\ 1 - \frac{1}{\mu}(\prod_{i=1}^{N} y_i)^{\frac{1}{N}} & \varepsilon = 1, \end{cases} \quad (2)$$

where y_i is an individual income and μ is the mean income [13]. After ε has been chosen Atkinson index behaves like Gini coefficent: values range from 0 to 1 and lower values indicate more equal distribution. Unlike Gini coefficent, Atkinson index is subgroup decomposable meaning that the index of the whole population can be computed as the sum of the indices of its subgroups. The index can be thought to describe the proportion of total income required to have the same social welfare as with the current income distribution if the incomes were equally distributed. Atkinson index of 0.4 indicates that only 60 % of current total income would be required to achieve the same level of social welfare as currently.

2.1.3 Distribution as ratio

The inequalities of income distribution can also be measured as ratios of different sorts. One method is a percentile ratio. For example, P90/P10 refers to the upper bound value of income of the ninth decile (of people ranked according to their incomes) divided by the income of the first decile. Other percentile ratios used often are P90/P50 and P80/P20.

Another way to measure inequalities are share ratios. S90/S10 is the share of income by the richest (i.e. people with the highest income) 10 % divided by the share of the poorest 10 % [8]. In contrast to P90/P10, the measure of S90/S10 is also affected by the incomes of the richest and poorest people, for example the top and bottom 1 %.

2.2 Crime

2.2.1 Crime explanations

The purpose of this study is to see empirically if income distribution is one of many possible macro-level determinants for crime. This study does not seek to discover the mechanism that causes crime, but rather tries to find if there actually is any, which is far more relevant for political decisions concerning income distribution.

Theoretically crime can be explained by inequality of income distribution either economically or sociologically. Economically crime rate can be seen as dependent of the net gains of committing crime, which in turn are (partially) caused by wealth differences. Wealth distribution would likely be a better explanator for crime than income distribution, but lack of reliable data forces researchers to usually limit themselves to study the effects of income distribution.

Sociological theories for crime arise from observations that crime tend to be committed by lower-class people. According to the theory of relative deprivation, economic inequality causes social tensions between the less-well off and wealthier

people. These tensions lead to the poor seeking compensation and satisfaction by different means, including crime. For further information, see [2] and the references therein.

2.2.2 Crime statistics

The reliability of crime data is major challenge for this study. The previous study by Fajnzylber et al. [2] tries to maximize the number of observed countries at the expense of data reliability. In this study we limit the observed countries first to mostly western countries and then only to OECD countries to maximize reliability. We also use only the relatively reliably gathered crime data from homicides and robberies with similar definitions in western countries. According to the United Nations Office on Drugs and Crime [10], "intentional homicide is defined as lawful death inflicted upon a person with the intent to cause death or serious injury". Robbery is defined as "the theft of property from a person, overcoming resistance by force or threat of force". As a third crime type, we study a crime index (CI) that tries to capture overall violent crime rate from both homicides and robberies. Mathematically it can be calculated for country i as

$$CI = \frac{AHR_i}{AHR} + \frac{ARR_i}{ARR}, \tag{3}$$

where AHR_i is the average homicide rate in country i and AHR the average of the average homicide rates in all observed countries (ARR for robberies).

3 Statistical methods

3.1 Cross-national analysis

There are many ways to study the effects of income distribution on crime. This study observes the effects on a national level and thus compares the crime rates of different countries. No dynamic analysis is done due to data limitations. It would also require great effort for a dynamic model to be sufficiently accurate, for there can be many sorts of lags and unknown predictors affecting changes in crime rate. The relationships between income distribution and crime are studied with correlation, simple linear regression (graphically) and multiple linear regression.

3.2 Simple linear regression

Simple linear regression studies the relationship between two quantities (X and Y). The relationship is described by a straight line, that is

$$Y = \alpha + \beta X, \tag{4}$$

where α is the value of Y when X equals zero. We may need to transform the scales of X and/or Y for linearity to hold. In this study, crime rates (Y) are transformed to their logarithms in order to obtain a more fitting linear model.

However, real data usually has statistical errors for various reasons. These errors account for the failure of a model to provide an exact linear fit. If we have observed values (x_i, y_i) of X and Y for i = 1,2, ..., n and e_i is the statistical error, then simple linear regression model can be defined as

$$\begin{cases} y_i = \alpha + \beta x_i + e_i \quad i = 1, 2, ..., n \\ \text{with } E(e_i) = 0 \\ \text{var}(e_i) = \sigma^2 \\ \text{cov}(e_i, e_j) = 0 \qquad\qquad i \neq j \end{cases} \tag{5}$$

3.3 Least squares estimation

To obtain estimates (denoted with a "hat") of the parameters in a model, a method called ordinary least squares is used in this study. This method seeks to minimize a quantity called the residual sum of squares. In case of a simple linear model, the residual for the ith observation is

$$\hat{e} = y_i - (\hat{\alpha} + \hat{\beta}) \quad i = 1, 2, ..., n. \tag{6}$$

Thus, the least squares estimators are those values that minimize the residual sum of squares

$$RSS(\alpha, \beta) = \sum_{n=1}^{N} [y_i - (\hat{\alpha} + \hat{\beta} x_i)]^2. \tag{7}$$

3.4 The coefficent of determination

The coefficent of determination, R^2, is a popular statistic that summarizes the strength of the relationship between observed data. First we define the corrected sum of squares for the y_i's, that is

$$SYY = \sum_{n=1}^{N}(y_i - \overline{y})^2 = \sum_{n=1}^{N} y_i^2 - (\sum_{n=1}^{N} y_i)^2/n \tag{8}$$

The coefficent of determination is then defined as

$$R^2 = \frac{SYY - RSS}{SYY} = 1 - \frac{RSS}{SYY}. \tag{9}$$

For a simple linear model, the coefficent of determination is the same as the square of the sample correlation. Thus,

$$R^2 = r_{XY}^2. \tag{10}$$

3.5 Multiple linear regression

Multiple linear regression is an extension of simple linear regression. Instead of one independent variable X, we investigate the effect of several simultaneously. While this study focuses on the effects of income distribution alone, the use of multiple linear regression allows us to eliminate possible bias due to an ignored variable that affects both X and Y. In other words, with multiple linear regression we can determine if there is likely to be causation behind observed correlation. The model (for p predictors) is defined as

$$y_i = \beta_0 + \sum_{j=1}^{p}(\beta_j x_i) + e_i \quad i = 1, 2, ..., n. \tag{11}$$

In matrix form this can be rewritten as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}. \tag{12}$$

The residual sum of squares is

$$RSS(\boldsymbol{\beta}) = \sum_{n=1}^{N}(y_i - x_i^T\boldsymbol{\beta})^2 = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{13}$$

The least squares estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ can be calculated from the formula

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}. \tag{14}$$

The statistical significance of these estimators can then be interpretated from the p-values of t-tests. The coefficent of determination for multiple linear regression is

the same as for simple linear regression (9). However, R^2 increases with the addition of extra predictors. To account for this phenomenon, adjusted R^2 is defined as

$$\overline{R}^2 = R^2 - \frac{(1 - R^2)p}{n - p - 1}, \tag{15}$$

where p is the number of predictors and n the number of observations.

When choosing the predictors for the model, highly correlated independent variables should be avoided. This is called the problem of multicollinearity. Collinear predictors will lead to large variances for estimated coefficients. To avoid multicollinearity, we can study variance inflation factor (VIF) for each predictor, that is

$$VIF = \frac{1}{1 - R_i^2}, \tag{16}$$

where R_i^2 is the coefficent of the determination for the ith predictor as a regression of the other predictors. VIF of more than 10 indicates high collinearity. For more on multiple linear regression and statistical methods, see for example [11] and [12].

4 Results

4.1 Correlation between income inequality and crime

In this section we consider three different time periods where we study the correlations and regressions. First wave refers to years 2003-2005, second wave to 2006-2009 and third wave to 2010-2013. Income distribution data comes from LIS data center [3] and crime data from UNODC [10]. While the observed countries were determined by the available data, most of the countries studied in this section can be labelled as western. The observed crime types are homicides, robberies and a crime index made to approximate overall violent crime rate.

Correlation matrices of income distribution indices and crime rates for third wave are presented in table 2, others are in appendix A. Crime rates (number of incidents per 100,000 people) are averages of the period and are expressed in logarithms to linearize the relationships. Regression lines and data points for the third wave are presented in figures 3, 4 and 5, others can be found in appendix B.

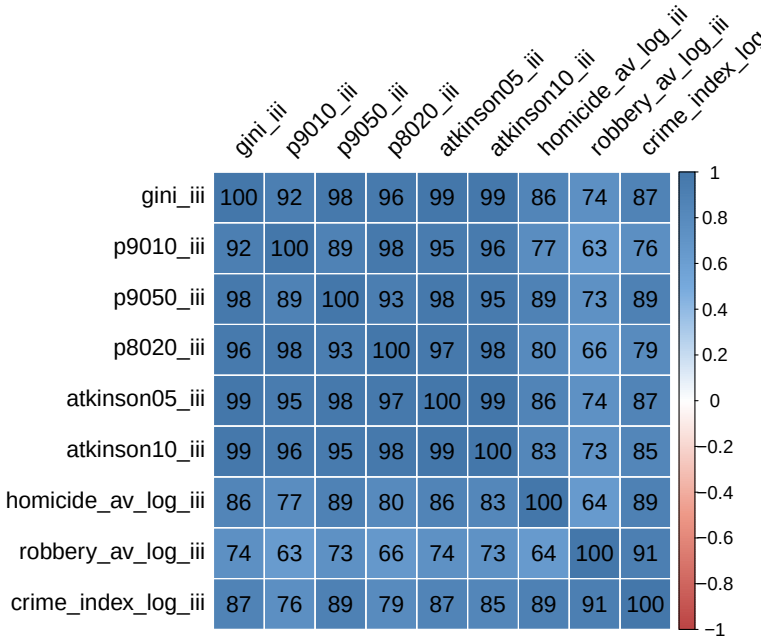

Figure 2: Correlation matrix from third wave (2010-2013)

Clearly there is a strong relationship between crime rates and income distribution, but this could be due to a third factor correlating with both of them (e.g. GDP). Robberies don't correlate with inequality of income distribution as much as homicides do. This could be caused by lesser reliability of crime data or possibly robberies are simply more random or explained by other macro-level factors. Interestingly the crime index combining both homicides and robberies has correlations on the same level as homicides. This could mean that different crime types are more common in certain countries, reducing their observed correlation with economic inequality. Different measures for income distribution perform similarly. Although this means that no index can be named superior, the similar results indicate robustness of the findings.

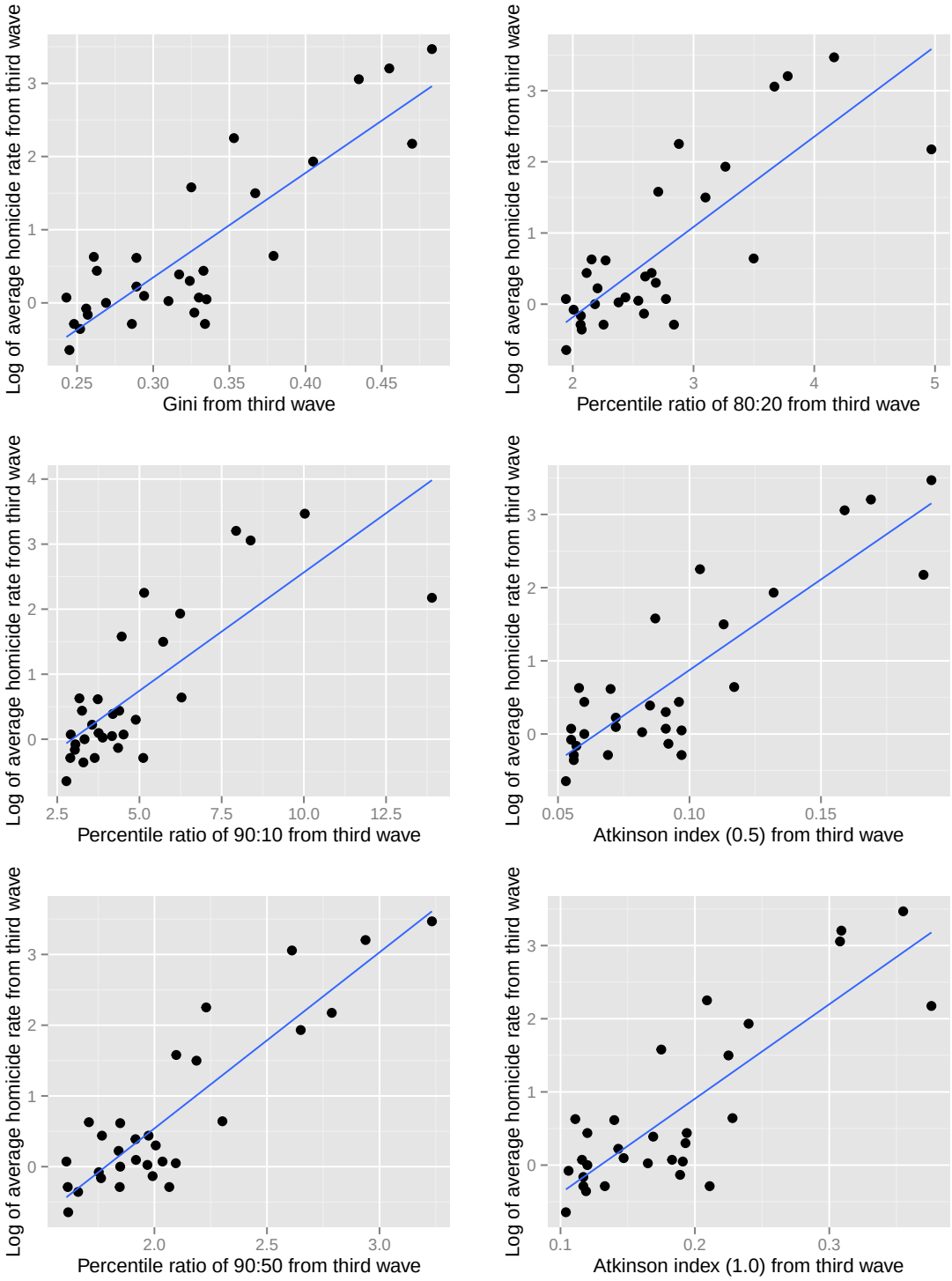

Figure 3: Income distribution and homicide rates from third wave (2010-2013)

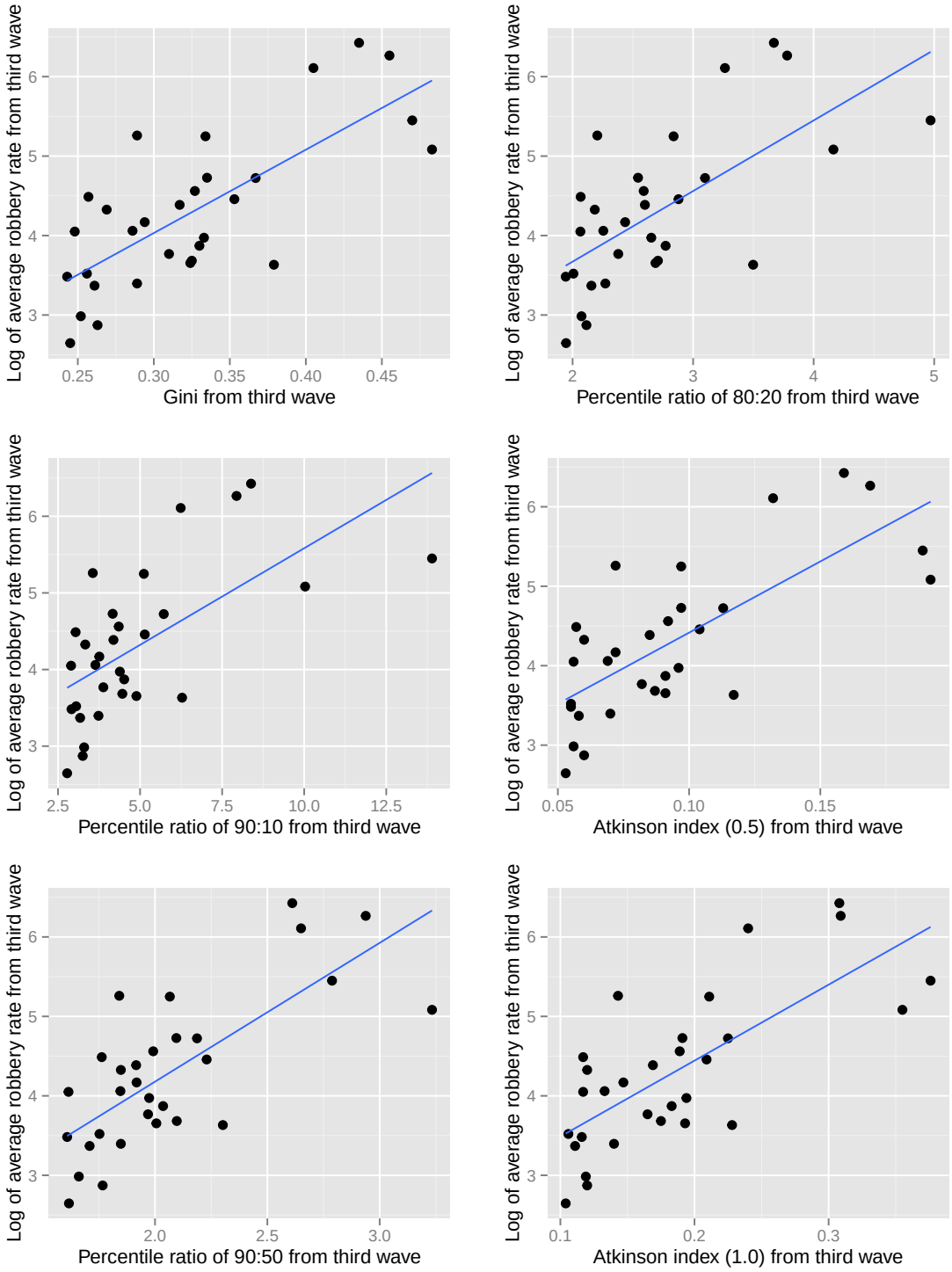

Figure 4: Income distribution and robbery rates from third wave (2010-2013)

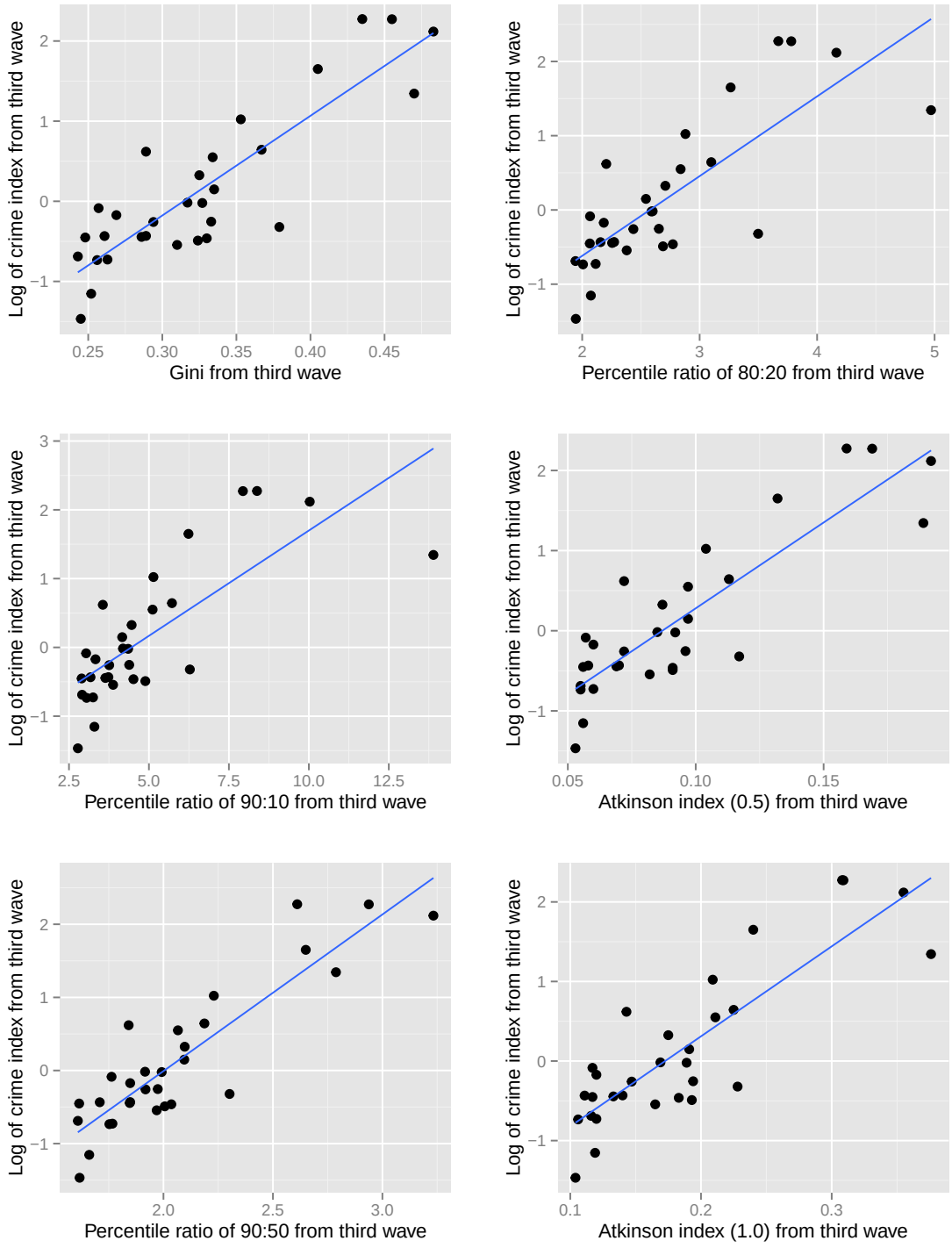

Figure 5: Income distribution and crime index from third wave (2010-2013)

4.2 Multiple linear regression in OECD countries

In this section we study the possible causation of inequality of income distribution on crime rates by multiple linear regression. We limit the study to all OECD countries with sufficient data available to unify the societal structure of the observed countries. The income distribution data comes from OECD [8] in this section. Therefore we have to leave out Atkinson index out from this section, but as the previous section suggests, the index wouldn't perform much better or worse than other indices for income distribution. The models are built for year 2012 to maximize the number of countries. The studied crime types are the same as in previous section. Crime rates are averages of years 2012 and 2013 to smooth possible crime peaks out. This also leaves some room for actual causation to happen between observations. However, the real lags are likely to be much longer.

Apart from income distribution, we also study the macro-level effects of gross domestic product [6], change in gross domestic product (2000-2007, 2007-2012 and 2000-2012) and education [9], [5]. The parameter choices have been influenced by the previous study of Fajnzylber et al. [2]. For education we have two indices: the educational enrolment rate of 15-19 year olds is an alternative measure for income distribution due to the high correlation between them. The share of adults (25-34 year olds) with tertiary education tries to capture the level of education in general.

Correlation matrix of the observed parameters is presented in figure 6. Simple linear regressions of interest are presented in appendix C. Results of multiple linear regressions are presented in tables 1 and 2 for the crime index and in appendix C for homicides and robberies. Appendix C also includes the variance inflation factors of the regression models in table 7.

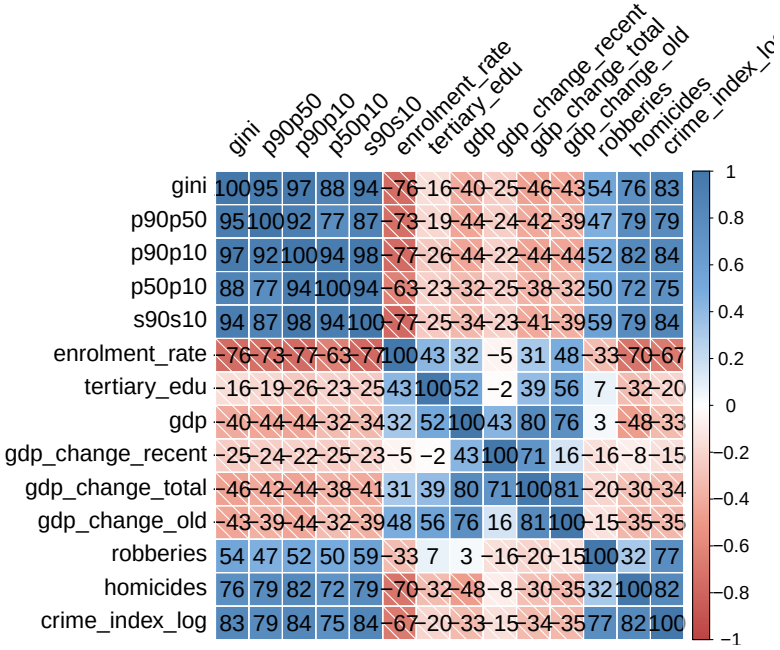

Figure 6: Correlation matrix for OECD countries (2012)

Unexpectedly, the correlation results for income distribution and crime are highly similar to those of the previous section. Of particular interest is that the crime index again has an unintuitively high correlation with income distribution in spite of the low correlation between homicides and robberies. The relatively low level of correlation between the two education indices is also noteworthy.

The regression models indicate that inequality of income distribution is a statistically significant contributor to crime. Educational enrolment rate of young people is linked with both income distribution and crime but cannot explain their dependence alone. However, due to their high complexity, these linked relationships cannot be further analyzed with the results of this thesis.

Table 1: Crime index regression results

	Dependent variable:		
	crime_index_log		
	(1)	(2)	(3)
gini	12.006***		
	(1.874)		
p90p50		2.094***	
		(0.376)	
p90p10			0.525***
			(0.081)
gdp	−0.00000	0.00001	0.00000
	(0.00001)	(0.00001)	(0.00001)
gdp_change_total	0.00002	−0.00001	0.00000
	(0.00003)	(0.00003)	(0.00003)
tertiary_edu	−0.008	−0.006	0.00001
	(0.012)	(0.013)	(0.012)
Constant	−3.260***	−3.639***	−1.933***
	(0.797)	(0.967)	(0.618)
Observations	27	27	27
R^2	0.695	0.637	0.700
Adjusted R^2	0.639	0.571	0.645
Residual Std. Error (df = 22)	0.438	0.478	0.435
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2: Crime index regression results

	Dependent variable:		
	crime_index_log		
	(1)	(2)	(3)
p50p10	1.732***		
	(0.362)		
s90s10		0.140***	
		(0.020)	
enrolment_rate			−0.055***
			(0.013)
gdp	−0.00001	−0.00001	−0.00001
	(0.00002)	(0.00001)	(0.00002)
gdp_change_total	0.00001	0.00002	−0.00001
	(0.00004)	(0.00003)	(0.00004)
tertiary_edu	0.002	0.004	0.017
	(0.014)	(0.011)	(0.016)
Constant	−3.103***	−0.989*	4.802***
	(1.011)	(0.488)	(1.018)
Observations	27	27	27
R^2	0.571	0.723	0.500
Adjusted R^2	0.493	0.672	0.409
Residual Std. Error (df = 22)	0.519	0.418	0.561
Note:			*p<0.1; **p<0.05; ***p<0.01

5 Conclusions

This thesis studied the effects of income distribution on violent crime in western countries in the 21st century. Correlations and graphical analysis showed that indices measuring inequality of income distribution seem to explain both homicides and robberies. Multiple linear regression analysis suggested that the dependence between income distribution and crime is not due to a third factor affecting both, but rather to causation from inequality of income distibution.

It was harder to explain robbery rates by macro-level factors than homicide rates. However, income distribution was still the main and statistically significant contributor to both crime types. There were no major differences between different income distribution indices on explaining crime rates, which suggests robustness of the findings.

Both crime types were combined to a crime index that measured overall violent crime prevalence. This crime index was affected by income distribution at least as much as homicide rates. This can be explained with the presumption of different crime types being more common in certain countries.

The share of adults with tertiary education seems to have little to do with crime rates. On the other hand, the educational enrolment rate of teenagers (15-19 year olds) is highly linked with both income distribution and crime. However, education of young people cannot alone explain the dependence of income distribution and crime.

The results of this thesis have notable political implications. In countries with relatively high levels of economical inequality and criminality, reducing the inequality of income distribution will likely lead to lower crime rates. One possible method to achieve this could be increasing the enrolment rate of teenagers. On the other hand, countries with relatively low levels of economic inequality and criminality can expect a rise in crime rates if economic inequality is let to increase.

References

[1] J. Cassidy. *Forces of Divergence.* 2014. http://www.newyorker.com/magazine/2014/03/31/forces-of-divergence (Accessed on 05 October 2015)

[2] P. Fajnzylber, D. Lederman, N. Loayza. *Inequality and Violent Crime* 2002. Journal of Law and Economics, vol. XLV. http://siteresources.worldbank.org/DEC/Resources/Crime%26Inequality.pdf (Accessed on 05 October 2015)

[3] LIS Cross-National Data Center in Luxembourg. Key figures. http://www.lisdatacenter.org/ (Accessed on 13 August 2015)

[4] F. Maio. *Income inequality measures.* 2007. Journal of Epidemiology and Community Health, 61(10), 849–852. http://doi.org/10.1136/jech.2006.052969 (Accessed on 05 October 2015)

[5] OECD (2015), Enrolment rate (indicator). doi: http://10.1787/1d7e7216-en (Accessed on 13 August 2015)

[6] OECD (2015), Gross domestic product (GDP) (indicator). doi: http://10.1787/dc2f7aec-en (Accessed on 13 August 2015)

[7] OECD (2015), *In It Together: Why Less Inequality Benefits All.* OECD Publishing, Paris. doi: http://10.1787/9789264235120-en (Accessed on 05 October 2015)

[8] OECD (2015), Income inequality (indicator). doi: http://dx.doi.org/10.1787/459aa7f1-en (Accessed on 13 August 2015)

[9] OECD (2015), Population with tertiary education (indicator). doi: http://10.1787/0b8f90e9-en (Accessed on 13 August 2015)

[10] UNODC - United Nations Office on Drugs and Crime. Statistics. https://data.unodc.org/ (Accessed on 13 August 2015)

[11] S. Weisberg. *Applied Linear Regression.* Second Edition. New York: John Wiley & Sons, 1985.

[12] T. Wonnacott, R. Wonnacott. *Introductory Statistics.* Third edition. New York: John Wiley & Sons, 1977.

[13] World Bank. Poverty Manual. 2005. http://siteresources.worldbank.org/PGLP/Resources/PMch6.pdf (Accessed on 05 October 2015)

A Additional correlation matrices

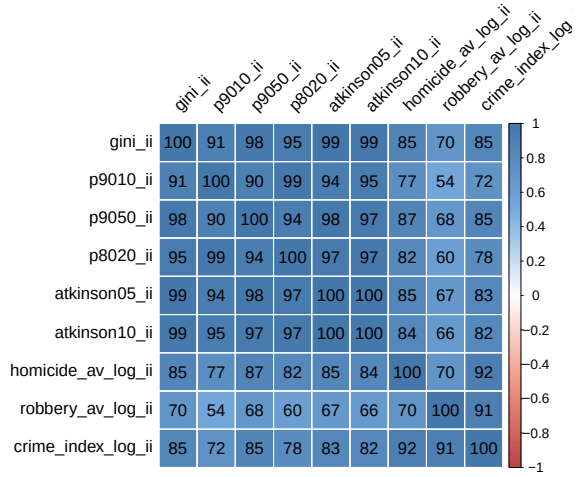

Figure 7: Correlation matrix from second wave (2006-2009)

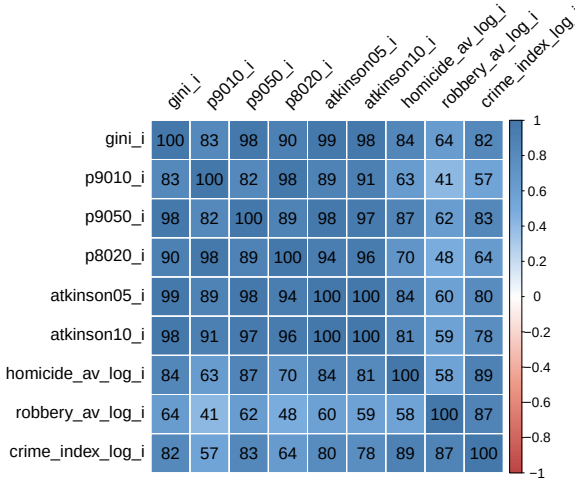

Figure 8: Correlation matrix from first wave (2003-2005)

B Additional graphical results from LIS data

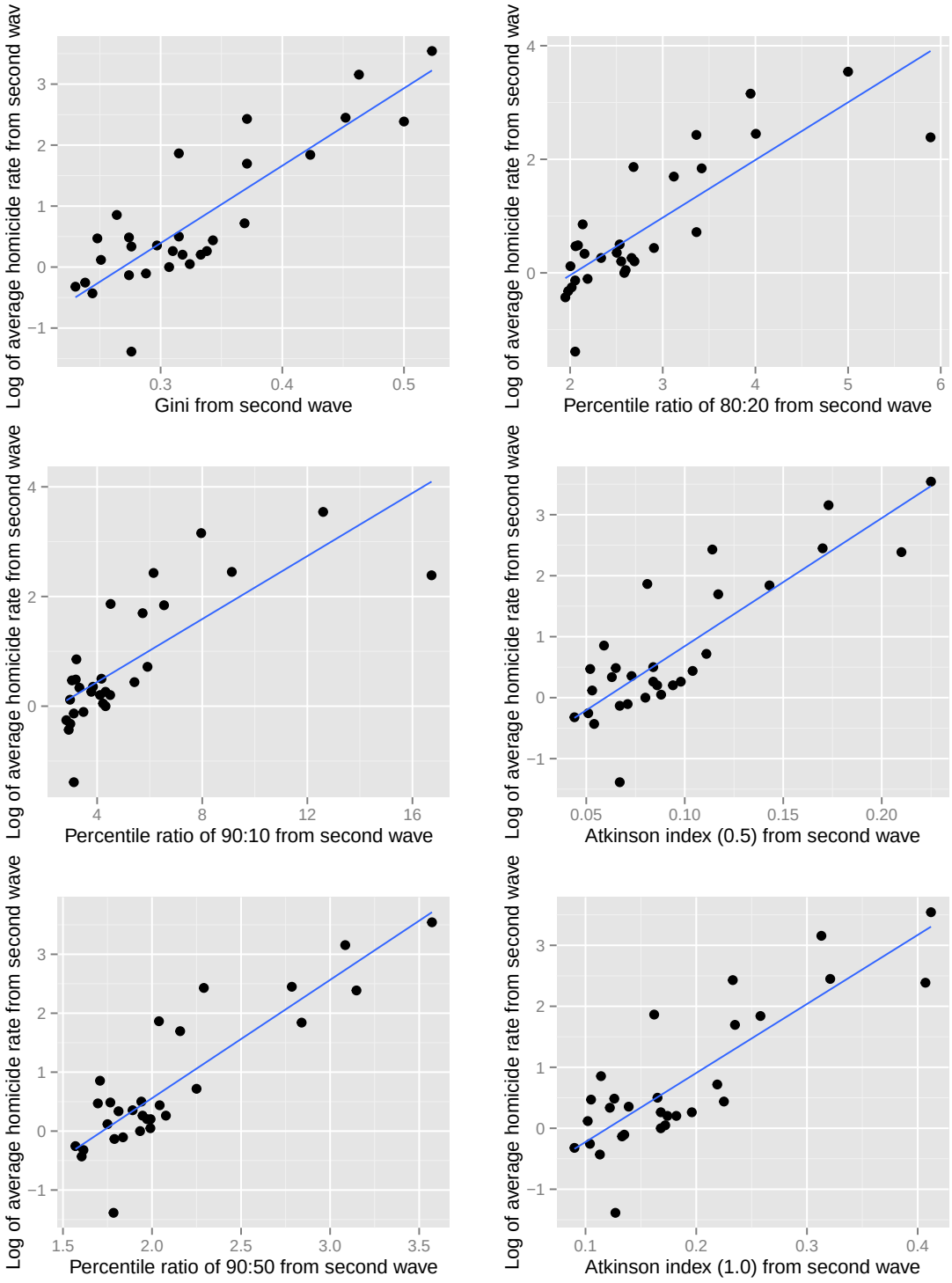

Figure 9: Income distribution and homicide rates from second wave (2006-2009)

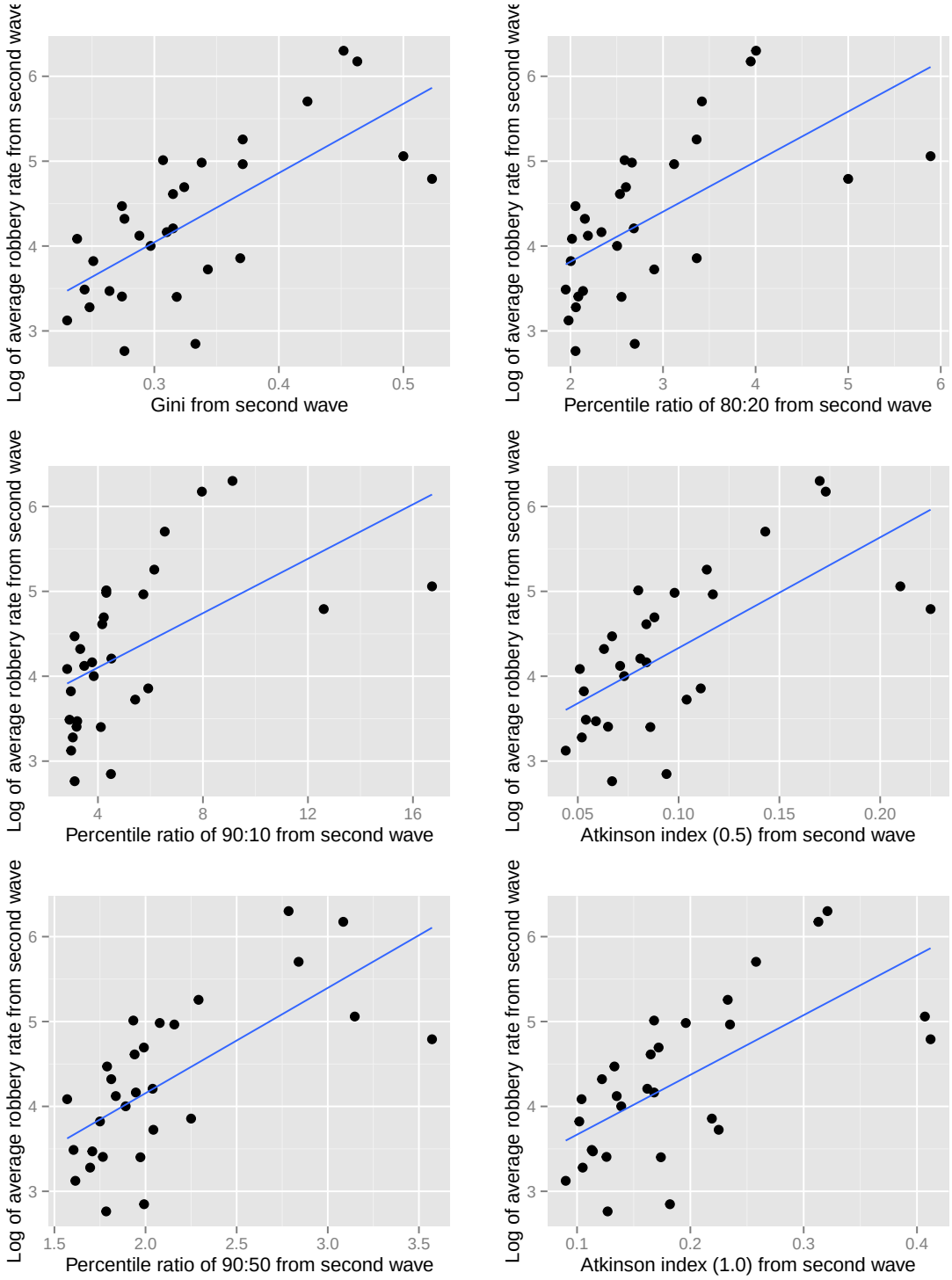

Figure 10: Income distribution and robbery rates from second wave (2006-2009)

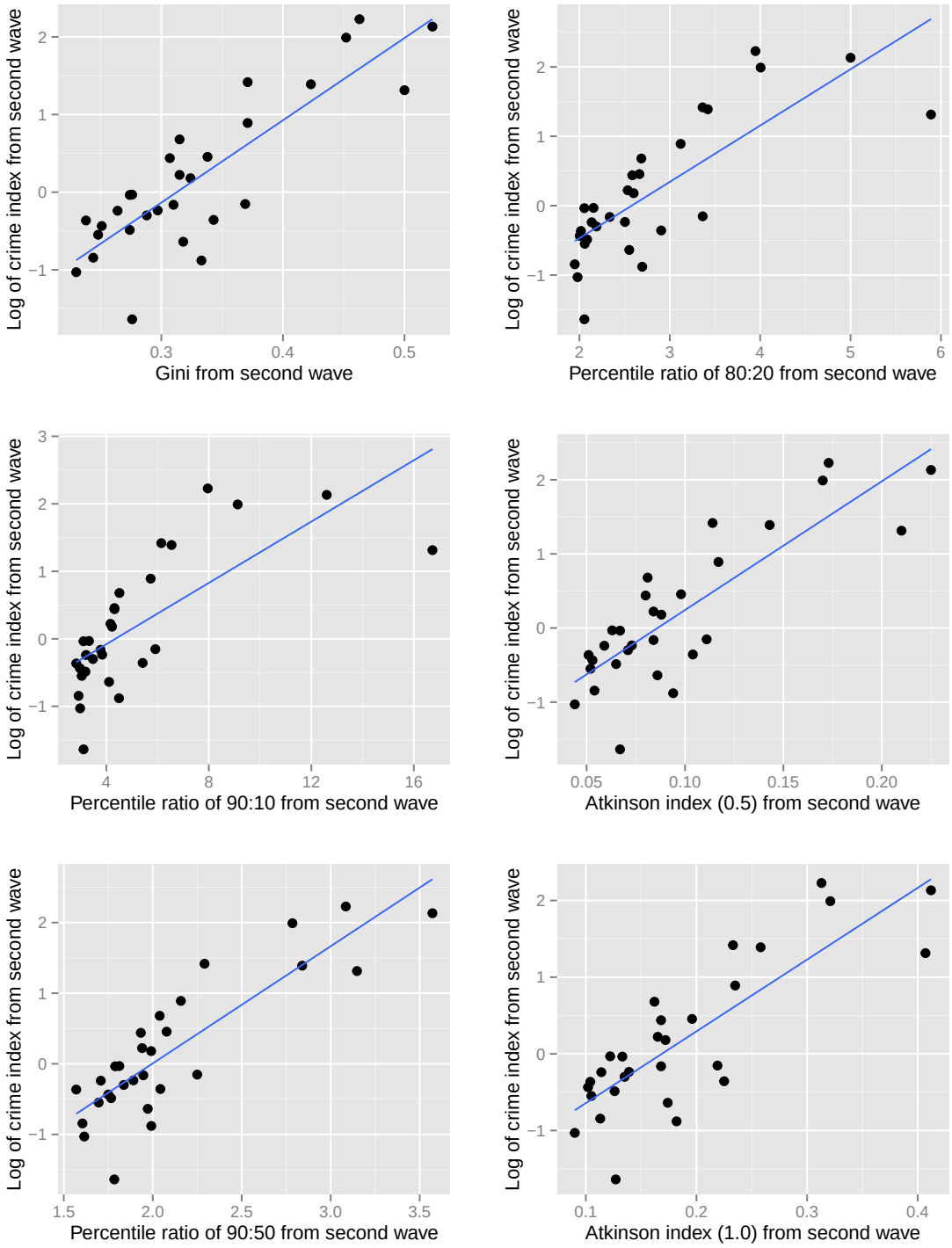

Figure 11: Income distribution and crime index from second wave (2006-2009)

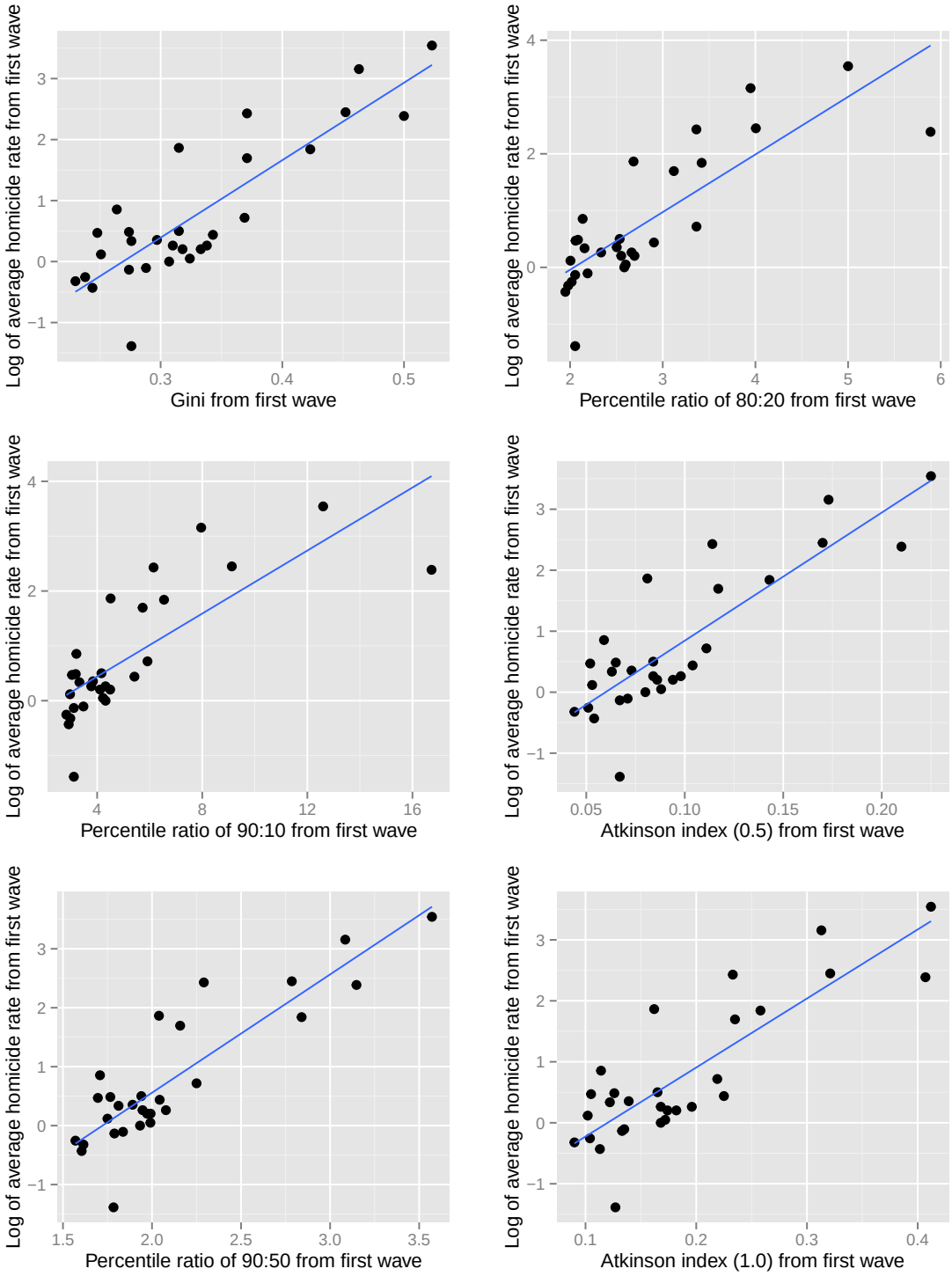

Figure 12: Income distribution and homicide rates from first wave (2003-2005)

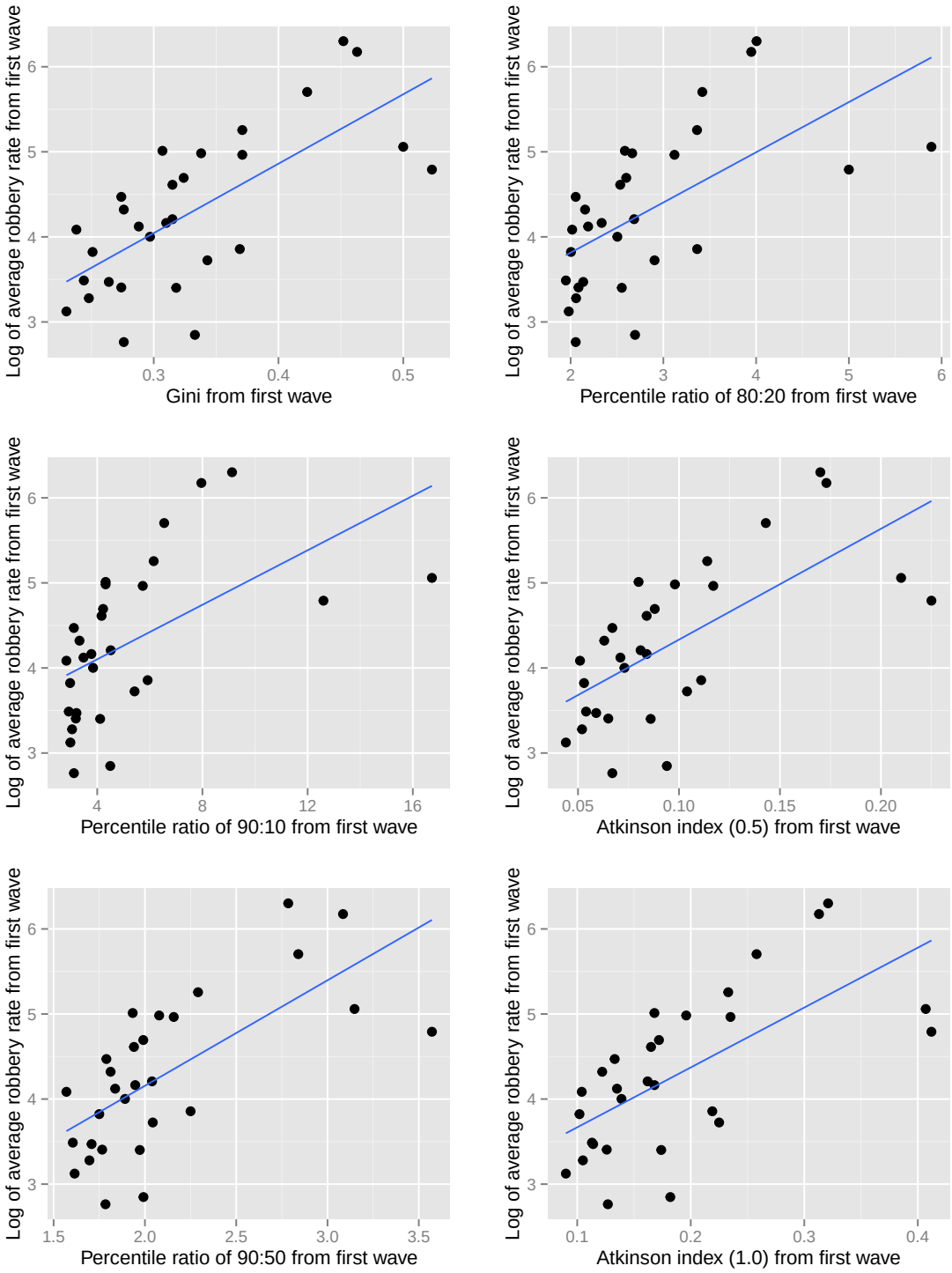

Figure 13: Income distribution and robbery rates from first wave (2003-2005)

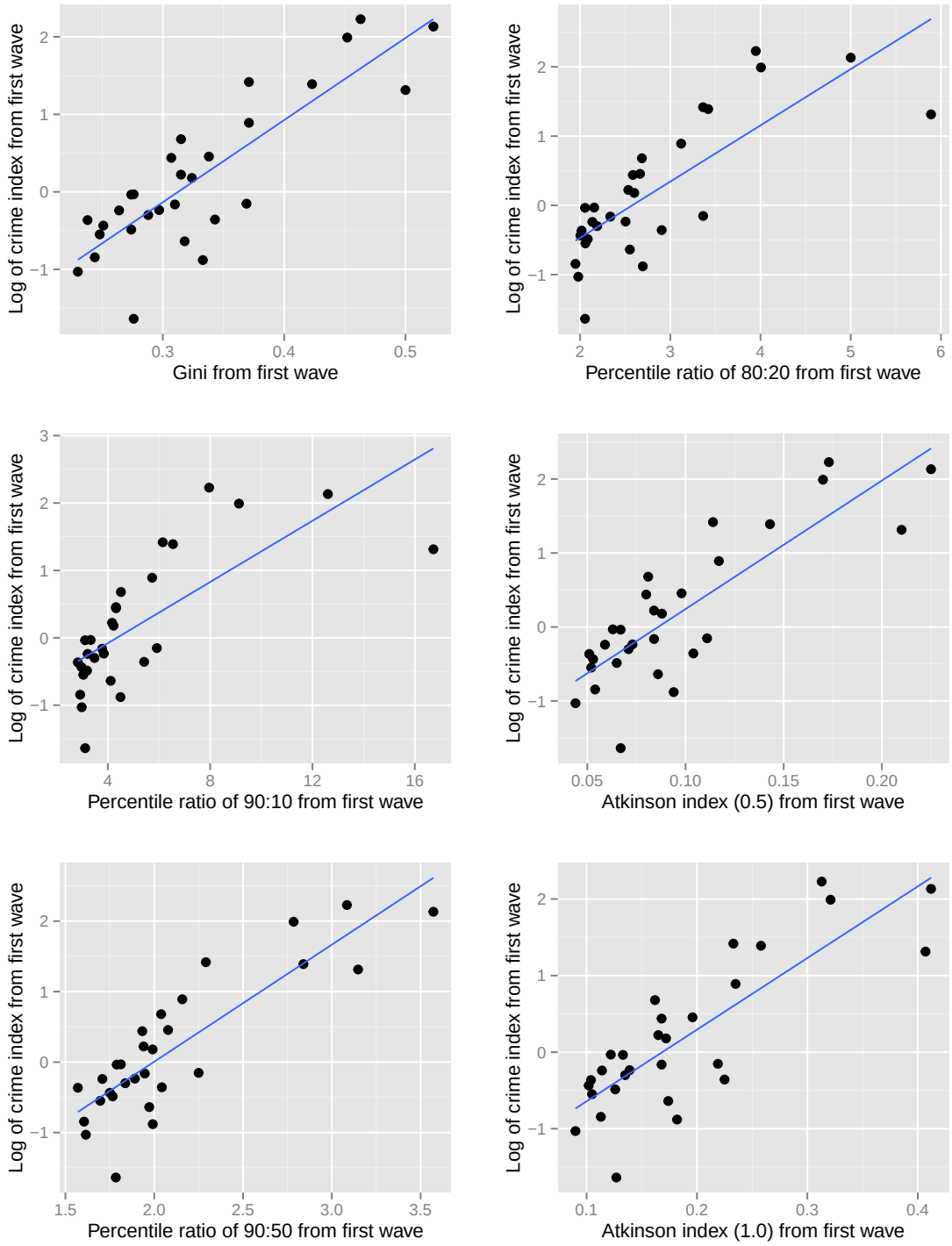

Figure 14: Income distribution and crime index from first wave (2003-2005)

C Additional results from OECD data

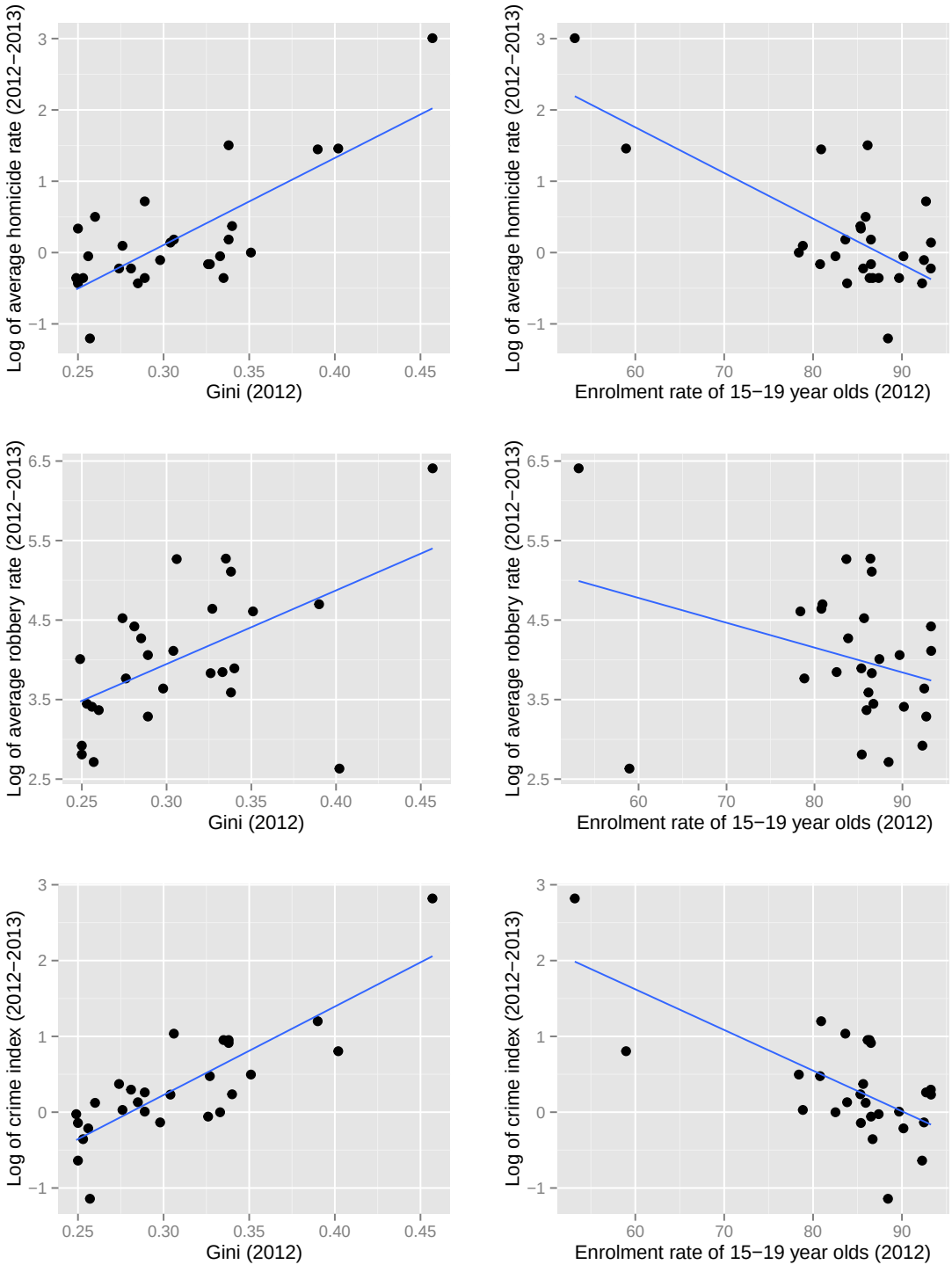

Figure 15: Gini coefficent, enrolment rate of 15-19 year olds and crime rates in OECD countries (2012)

Table 3: Homicide regression results

	Dependent variable:		
	homicides		
	(1)	(2)	(3)
gini	12.329***		
	(2.071)		
p90p50		2.237***	
		(0.390)	
p90p10			0.560***
			(0.084)
gdp	−0.00004**	−0.00003*	−0.00003**
	(0.00001)	(0.00001)	(0.00001)
gdp_change_total	0.0001**	0.0001*	0.0001**
	(0.00004)	(0.00003)	(0.00003)
tertiary_edu	−0.012	−0.011	−0.005
	(0.013)	(0.013)	(0.012)
Constant	−3.035***	−3.624***	−1.797***
	(0.881)	(1.002)	(0.638)
Observations	27	27	27
R^2	0.714	0.701	0.754
Adjusted R^2	0.662	0.646	0.710
Residual Std. Error (df = 22)	0.484	0.495	0.448
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Homicide regression results

	Dependent variable:		
	homicides		
	(1)	(2)	(3)
p50p10	1.870***		
	(0.373)		
s90s10		0.150***	
		(0.021)	
enrolment_rate			−0.062***
			(0.013)
gdp	−0.00004**	−0.00004***	−0.00004**
	(0.00002)	(0.00001)	(0.00002)
gdp_change_total	0.0001**	0.0001***	0.0001
	(0.00004)	(0.00003)	(0.00004)
tertiary_edu	−0.002	0.0002	0.015
	(0.014)	(0.012)	(0.016)
Constant	−3.100***	−0.792	5.625***
	(1.040)	(0.502)	(1.010)
Observations	27	27	27
R^2	0.651	0.775	0.622
Adjusted R^2	0.588	0.734	0.553
Residual Std. Error (df = 22)	0.534	0.429	0.557
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5: Robbery regression results

	Dependent variable:		
	robberies		
	(1)	(2)	(3)
gini	10.134***		
	(3.099)		
p90p50		1.719***	
		(0.591)	
p90p10			0.448***
			(0.134)
gdp	0.00004*	0.00005**	0.00005**
	(0.00002)	(0.00002)	(0.00002)
gdp_change_total	−0.0001	−0.0001*	−0.0001*
	(0.0001)	(0.0001)	(0.0001)
tertiary_edu	0.003	0.004	0.009
	(0.019)	(0.020)	(0.019)
Constant	0.248	0.039	1.339
	(1.318)	(1.519)	(1.022)
Observations	27	27	27
R^2	0.427	0.385	0.435
Adjusted R^2	0.323	0.273	0.333
Residual Std. Error (df = 22)	0.724	0.750	0.719
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6: Robbery regression results

	Dependent variable:		
	robberies		
	(1)	(2)	(3)
p50p10	1.479**		
	(0.530)		
s90s10		0.125***	
		(0.034)	
enrolment_rate			−0.038*
			(0.020)
gdp	0.00004	0.00004*	0.00004
	(0.00002)	(0.00002)	(0.00002)
gdp_change_total	−0.0001	−0.0001	−0.0001*
	(0.0001)	(0.0001)	(0.0001)
tertiary_edu	0.011	0.013	0.021
	(0.020)	(0.019)	(0.023)
Constant	0.338	2.060**	6.529***
	(1.477)	(0.811)	(1.480)
Observations	27	27	27
R^2	0.371	0.474	0.274
Adjusted R^2	0.257	0.379	0.142
Residual Std. Error (df = 22)	0.759	0.694	0.815
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7: Variance inflation factors (VIF) of regression models

Regression model	1	2	3	4	5	6
Gini	1.27					
P90/P50		1.26				
P90/P10			1.28			
P50/P10				1.18		
S90/S10					1.22	
Enrolment rate						1.26
GDP	3.27	3.36	3.30	3.25	3.25	3.25
Total GDP change	2.97	2.84	2.88	2.93	2.98	2.83
Tertiary education	1.38	1.38	1.38	1.39	1.40	1.55

D Yhteenveto (in Finnish)

Taloudellinen epätasa-arvo on yhteiskunnallisesti merkittävä aihe, jolla on monia ulottuvuuksia. Aihetta on tutkittu paljon niin syiden kuin seurausten näkökulmasta, mutta lisätutkimuksille on runsaasti tarvetta. Vaikka usein varsinaisena kiinnostuksen kohteena on varallisuuden jakautuminen, yleensä taloudellista epätasa-arvoa ja sen vaikutuksia tutkitaan tulonjaon avulla sen laajemman ja luotettavamman tilastoinnin vuoksi. Tulonjaon mittaamiseen on kehitetty useita eri indeksejä, joista tunnetuin on Gini-kerroin, joka ilmaisee oikean tulonjaon eron hypoteettisesta tilanteesta, jossa kaikilla olisi samat tulot. Muita indeksejä ovat mm. erilaiset tulonjaon persentiilisuhteet, jotka vertaavat hyvätuloisten tuloja köyhempien kansalaisten tuloihin. Tutkimuksessa käytetyt tulonjaon indeksit ovat peräisin "Luxembourg Income Study Database":n ja Taloudellisen yhteistyön ja kehityksen järjestön (OECD) tilastoista.

Kandidaatintyössä tutkittiin tilastollisilla menetelmillä epätasaisen tulonjaon vaikutusta väkivaltarikollisuuteen länsimaissa. Rikollisuuden tilastointi on harvoin erityisen luotettavaa, eikä rikoksia ole määritelty yhteneväisesti kaikissa maissa. Tästä syystä tutkimus keskittyikin vain kahteen suhteellisen luotettavasti tilastoituun rikostyyppiin, henkirikokset ja ryöstöt, joiden maakohtaiset määrittelyt eivät vaihtele merkittävästi. Näiden lisäksi tarkasteltiin tutkimusta varten kehitettyä rikosindeksiä, joka mittaa väkivaltarikollisuuden yleisyyttä henkirikos- ja ryöstötilastojen avulla. Tutkimuksessa käytettiin Yhdistyneiden kansakuntien huumausaine- ja rikosasioiden toimiston (UNODC) rikostilastoja.

Tutkimuksen ensimmäisessä osassa tarkasteltiin tulonjaon ja rikollisuuden yhteyttä lähinnä länsimaista koostuvassa joukossa maita 2000-luvulla. Tarkastelussa käytettiin kolmella eri ajanjaksolla maita, joista oli saatavilla riittävästi aineistoa. Rikostilastoista (rikoksia sataatuhatta henkeä kohden) otettiin kyseisen ajanjakson keskiarvo. Riippuvuuksien linearisoimiseksi näille keskiarvoille tehtiin logaritmimuunnos. Tutkimuksessa havaittiin, että tulonjaon indeksit korreloivat hyvin vahvasti väkivaltarikollisuuden kanssa. Graafisessa tarkastelussa huomattiin, että korrelaatio johtui melko selvästä lineaarisesta riippuvuudesta. Korrelaatio oli henkirikoksissa vahvempaa kuin ryöstöissä ja tuoreemmissa tilastoissa vahvempaa kuin vanhemmissa, mihin puutteellinen tilastointi voi olla yksi selittävä tekijä. Väkivaltarikollisuuden yleisyyttä kuvaava indeksi korreloi vähintään yhtä vahvasti tulonjaon epätasaisuuden kanssa kuin henkirikokset, mikä voi selittyä sillä, että erilaiset rikostyypit ovat yleisiä eri maissa. Käytetyllä tulonjaon indeksillä ei havaittu olevan merkittävää vaikutusta tuloksiin. Eri indeksien samankaltaiset tulokset kuitenkin vahvistavat tulkintaa tulonjaon ja rikollisuuden merkittävästä riippuvuudesta.

Koska tulonjaon ja rikollisuuden korrelaatio voisi selittyä kolmannella tekijällä, josta molemmat olisivat riippuvaisia, työn toisessa osassa tutkittiin mahdollista syyseuraussuhdetta moniuloitteisilla lineaarisilla regressiomalleilla. Moniuloitteisessa regressiossa tarkastellaan selitettävää muuttujaa (rikollisuus) summana useista selittävistä muuttujista. Regressiomallista voidaan tulkita selittävän tekijän (tulonjaon epätasaisuus) vahvuus ja tilastollinen merkittävyys, mihin pelkkä korrelaatiokerroin ei riitä. Tutkimuksen regressiomallit tehtiin vuodelle 2012, mutta rikollisuutta tarkasteltiin keskiarvona vuosista 2012 ja 2013. Tarkastellut maat rajattiin OECD:n

jäsenvaltioihin, jotta yhteiskuntarakenteet olisivat mahdollisimman yhteneväisiä maiden välillä. Selittävinä tekijöinä käytettiin tulonjaon lisäksi bruttokansantuotetta, bruttokansantuotteen muutosta eli talouskasvua ja koulutusta. Talouskasvu 2000-luvulla jaettiin kolmeen indeksiin: talouskasvu ennen finanssikriisiä eli 2000-2007, finanssikriisin jälkeen eli 2007-2012 ja kokonaiskasvu eli 2000-2012. Rikollisuuden kannalta ei kuitenkaan ollut juurikaan merkitystä, mitä näistä indekseistä käytettiin tutkimuksessa. Koulutuksen vaikutusta tutkittiin kahdella eri indeksillä. Kolmannen asteen koulutuksen suorittaneiden osuus aikuisväestöstä oli mittarina yleiselle koulutustasolle. Nuorten (15-19-vuotiaiden) kirjoillaolo-osuus oppilaitoksissa taas oli vaihtoehtoinen mittari tulonjaolle korkean negatiivisen korrelaation vuoksi.

Tulonjaon epätasaisuuden ja rikollisuuden korrelaatio oli vahvaa tutkimuksen toisessakin osassa, vaikka tarkasteltujen rikostyyppien keskenäinen korrelaatio oli melko alhainen. Ryöstöjä oli vaikeinta selittää makrotason tekijöillä, kun taas tutkimuksessa käytetylle rikosindeksille se oli helpointa. Mielenkiintoisesti yleinen koulutustaso ei tuntunut vaikuttavan rikollisuuteen merkittävästi. Nuorten koulutus puolestaan selitti rikollisuutta lähes yhtä vahvasti kuin tulonjako, mutta tutkimuksen perusteella ei voida sanoa, että tulonjaon yhteys rikollisuuteen johtuisi vain nuorten koulutuksesta. Regressiomalleista havaittiin, että tulonjaolla on selkeä ja tilastollisesti merkitsevä vaikutus väkivaltarikollisuuteen länsimaissa.

Työn tuloksilla on merkittäviä yhteiskunnallisia tulkintoja. Länsimaissa, joissa on suhteellisen korkeat tuloerot ja rikollisuus yleistä, rikoslukuja voi todennäköisesti alentaa kaventamalla tuloeroja. Tutkimuksen perusteella nuorten koulutukseen panostaminen ja erityisesti koulutuksellisen syrjäytymisen ehkäiseminen on yksi mahdollinen tapa vähentää tuloeroja ja rikollisuutta. Toisaalta alhaisten tuloerojen maissa, kuten pohjoismaissa, rikollisuus luultavasti yleistyy, jos tuloerojen annetaan kasvaa.