Relationship between crime index and years of education

Bernardo Vedor, David Issá, Edgar Pon, João Ribeiro[1]

Nova School of Business and Economics

May 27, 2019

Abstract

Using a sample of 60 countries from 2013, we analyzed the impact of several key variables on crime index. Unemployment rate, median age of the population, Gini index and population density were found to be statistically significant at the 10% level and our model had an R-squared of 0.55. However, our hypothesis that education had a significant negative impact on crime was falsified, after controlling for several factors. For policy makers, it seems that focusing their efforts on decreasing income inequality, measured by the Gini index, could go a long way in reducing crime, especially if that can be achieved with a lower unemployment rate.

Keywords: education, crime

[1] Student numbers 32373, 32135, 32243, 32399

INTRODUCTION

It is well established that one of the central goals that all countries have in common is the reduction of the crime rate. Crime is a big problem for society, originating strong negative consequences both economically and socially.

Detotto and Otranto (2010) have found that higher crime rates have a negative impact on the growth rate of GDP by discouraging domestic and foreign investments, reducing firms' competitiveness and reallocation resources which generates uncertainty and inefficiency. In particular, Goulas (2012) concluded that in countries with higher-than-average macroeconomic uncertainty, a 10% increase in crime rate can reduce annual per-capita GDP growth by between 0.49% and 0.62%. Finally, Andersson (1999) estimated the annual burden of crime to be close to $1 trillion for the USA.

Furthermore, Whitley and Prince (2005) concluded that fear of crime caused negative mood and low self-esteem by preventing individuals from attending certain places at specific times, especially for low-income mothers. Another consequence of fear of crime was studied by Yin (1982) who found that it led to neighborhood dissatisfaction, low morale, and involuntary isolation in elderly people.

Due to crime's high social cost, it's critical for policymakers to understand the main drivers of crime rate and to subsequently implement measures that successfully tackle the most important variables.

In this work, we follow the example of Rathie, Sipos, and Ahuja (2016) and explore the relationship between a country's level of education and its crime rate index. There is evidence that other things being equal, educated people generally have better jobs and higher income than those with less schooling (Vila, 2000), which increases their quality of life.

Our hypothesis is that countries with a higher level of education have a lower crime rate index. Since education increases individuals' expected income raising the opportunity cost of crimes such as theft (indeed, crime is associated with lower wealth according to Muroi and Baumann (2009)) and that education can have positive mental impacts on people. We will test this under a simple linear regression model.

Afterward, we include other variables in our analysis to control for factors that may be correlated with education and crime rate, which would invalidate the previous SLR analysis, and test if there are other variables that can help explain differences in crime rate across countries so that policymakers can make informed decisions. As such, we will test the hypotheses that income inequality, population density, share of drug use and unemployment rate will have a significant positive impact on a country's crime rate index and that GDP per capita and median age will have a negative impact.

LITERATURE REVIEW

Most of the literature on the relationship between crime and education has been done on a country-by-country basis, especially in the US and other countries such as England, Italy and the Netherlands. For example, Machin, Marie, and Vujić (2011) have provided strong support for the theory that increased educational attainment leads to a reduction in property crime and produces large social benefits over time. However, this study focused only on England and Wales. Furthermore, Lochter and Moretti (2004) found supporting evidence in favor of the theory that increased high school graduation rates in the United States do correspond to significant decreases in the crime rate.

Besides focusing on specific nations, research has largely focused on the impact of increased high school education on crime rates, whereas we were interested in studying the impact of education in general across several countries. With that in mind, we found two key studies relating our variables of interest in different countries.

Rathie, Sipos, and Ahuja (2016), using data from the World Bank and Numbeo, obtained a total of 64 observations (each observation represents a country) in 2013. Initially, by using a simple linear regression model, the authors studied the relationship between crime index and a country's average years of education (including primary, secondary, tertiary school and above), estimating that for every unitary increase in the average years of education, the country's crime rate index will drop 2.44 points. This result was significant at 1% confidence, however, the authors decided to include more variables in the model due to the low R-squared obtained (0.1664).

For the study mentioned above, the final model ended up being a multiple linear regression with average years of education, Gini index, natural logarithm on GDP per capita and of population density, although the number of countries observed declined to 60 with the inclusion of more variables, due to lack of data. Only Gini index and population density turned out to be significant at 1% significance level and education was only statistically significant at 12.7%. The model had an R-squared of 0.358 and an F-statistic of 7.67 (the variables are jointly significant). The impact of an additional year of average education on a country's crime rate index turned out to be -1.58 points, compared to -2.44 on the simple linear regression model. Interestingly, the authors found that an increase in population density actually led to a decrease in crime rate, contrasting with their initial theory that the relationship between the two variables was positive.

We believe that the authors of the previously mentioned study left out some important variables that help explain variations in crime across countries. As such, we decided to include the median age of the country's population, unemployment rate, share of drug use in order to control for more factors and get unbiased estimators for the impact of Gini index, GDP per capita and population density on crime. Another reason to include the additional variables is to test whether they are statistically significant at explaining differences in crime index between countries.

Britt (1997) tested the unemployment-crime relationship by testing simultaneously for variation in the unemployment-crime relationship by age group and over time, using time-series data for the United States from 1958 to 1995. He found a weak causal relationship between higher unemployment rate and higher crime, especially on property crime among youth and young adults, hence the inclusion of this variable in our model.

Additionally, McBride and McCoy (1993) suggested the existence of strong empirical evidence of the statistical overlap between drug-using and criminal behavior: indeed, drug use is seen as increasing and sustaining criminal behavior. With this in mind, we decided to add this variable to our model to test its significance in explaining variations in crime.

Farrington (1986) studied the relationship between age and crime which tends to peak during teenage years and decrease gradually from then onwards. Our hypothesis then is that a country's median age is an important variable to explain its crime index.

Gonzales (2015) conducted research using data from 1998 to 2012 for 204 countries and estimated the following model:

Model 3: *College Graduation Rate and Youth Unemployment Rate (Lagged 1 Year)*

$$\textbf{Intentional Homicide Rate}_{i,t} = \beta_0 + \beta_1(\textbf{College Graduation Rate}) + \beta_2(\textbf{Youth Unemployment Rate, L(1)}) + \beta_3(\textbf{Population Density}) + \beta_4(\ln \textbf{Total Population}) + \beta_5(\ln \textbf{GDP per Capita}) + \beta_6(\textbf{Polity Score}) + \varepsilon_{i,t}$$

He concluded that college graduation rates had a statistically significant negative impact on homicides and that youth unemployment rate (lagged one period) was statistically significant at a 5% level. Similarly, to previous literature, the model includes population density and the natural logarithm of GDP per capita as explanatory variables and we follow the example in our own work.

However, we decided to use a crime index instead of intentional homicide rate to allow the inclusion of different types of crime, especially more common ones such as theft and assault. Additionally, we think using a broader measure of education (average years of education) could have more predictive capabilities than college graduation rate and perhaps be more useful to policymakers, although there is clearly a correlation between both measures. Finally, we did not include lagged variables since we only collected data for 2013, although it makes sense in this specific model since improvements in education can have effects only in later periods.

DATA OVERVIEW

Crime and education are variables that are usually perceived as having a negative correlation between them. That is, as the average number of years of education in a given country increases, we expect its crime rate to decrease.

The data set we use was specifically collected for the year 2013 as it was the year with a broader country representation, 60 countries in total, which we understood would give us a better overall picture of the worldwide relationship between crime and education.

The crime index in 2013 was collected from *NUMBEO*, a database widely used by credible business magazines such as Forbes and The Economist, as well as well-known information channels like BBC and The New York Times. At the website, there were 118 countries listed with the most recent data being 2019. However, after matching up the data with the data collected for average education, we were left with a sample of 60 countries. The year 2013 was selected as it was the most recent and complete set of data.

The average education was retrieved from *UNESCO's* database which is widely regarded as an impartial source. Having this, both variables are expected to reflect accurately how average country education affects crime rates, as long as the statistical analysis is performed correctly.

Median age of population data was collected from the *United Nations – Population Division,* a respected organization worldwide. At the website there was information about 241 countries for the year of 2013 from which we retrieved the 60 countries that already compose our data set.

Data for the GINI index was retrieved from *the World Bank*, an organization that aims to end world poverty and its database offers information about hundreds of countries worldwide. The countries listed matched the 60 we had information for. We expect countries with higher income inequalities to have higher crime rates.

Data for GDP per capita was also taken from the *World Bank* for the respective 60 countries of our data set. We expect to observe a negative correlation between GDP per capita and a country's crime rate.

Unemployment rate data was found on the *World Bank* database. We intend to test our hypothesis that a higher unemployment rate is positively correlated with an increase in a given country's crime rate as a more socially unstable environment can be conductive to an increase in crime.

Population density per square meter was collected from the *World Bank*. We aim to test the premise that crime rates are higher on more densely populated countries, that is, that there is a positive correlation between heavily populated areas and higher crime rates.

Finally, data for the share of drug use was gathered from *Our World in Data*, a website that was created to present and research data on the UN's Sustainable Development Goals (SDGs). We expect countries with a higher share of drug use to also exhibit a higher crime rate.

Label	Variable Description	Type of Measure	Type of Variable
CrimeIdx	*An index of crime by country based on surveys given to its citizens, that is based off the national standards of crime*	*[0;1] - Measure of a contry's crime relative to other coutries*	*Dependent*
EducIdx	*Index based on the combination of average adult years of schooling with expected years of schooling for children, each receiving 50% weighting.*	*[0;1] - Mean years of Schooling and Expected yaers of schooling*	*Independent*
MedianAge	*The median age of a country's population is an indicator of demographic makeup of the country and of its the population growth.*	*Median divides the population in two parts of equal size*	*Independent*
GiniIdx	*The GINI coefficient measures the distribution of income inequality of a given country*	*[0;100] - Measure of statistical dispersion*	*Independent*
lnGDPpc	*The natural log of the GDP per capita in a given country*	*Natural log of dollars*	*Independent*
PopDensity	*Population density is midyear population divided by land area in square kilometers. Land area is a country's total area, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones*	*Ratio between total population and land area of a given country*	*Independent*
ShareDrug	*Share of population with drug use disorders (Drug dependence is defined by the International Classification of Diseases as the presence of three or more indicators of dependence for at least a month within the previous year. Drug dependency includes all illicit drugs)*	*[0;1] - Measure of the share of drugs used relative to total population*	*Independent*
UnempRate	*The unemployment rate is the share of the labor force that is unemployed*	*[0;1] - Measure of unemployed workers against total labor force*	*Independent*

The sample of countries composing our data set was chosen randomly, and after a brief analysis, we can observe that they cover a diverse background, from rich to poor and undeveloped to developed. The list of countries used is stated in the appendix.

Descriptive statistics for all the variables used can be found in Figure 1, on the appendix.

From the 60 observed countries, the average value for crime index was 40.8 (out of 100). The crime index is a measure of overall crime within a given country. The interpretation of its scale is: a crime level below 20 is regarded as very low, from 20 to 40 is regarded as low, from 40 to 60 it is considered moderate and from 60 to 100 it is considered very high. The average crime of 40.8 overall is then considered moderate to low. It considers factors such as people's perceived safety at night, growth rate of crime in the last years, probability of being robbed, corruption, etc.

The average education index was 0.76 (out of 1) which indicates an overall high education index. An education index below 0.55 is considered low, from 0.55 to 0.7 it is considered medium, from 0.70 to 0.80 it is considered high, and from 0.80 to 1 it is considered very high.

The standard deviation for the crime index is relatively large which means that there is a lot of variability as one standard deviation covers both low and medium. Regarding the education index, the standard deviation is also fairly large since one standard deviation covers both the very high and the medium index consideration, assuming a normal distribution.

REGRESSION ANALYSIS CONSIDERATIONS

Functional Form:

- Simple Linear Regression Model: We regress crime index on education. The model is of the form.

$$CrimeIdx_i = \beta_0 + \beta_1 EducIdx_i + u_i$$

- Multilinear Regression Model: We regress crime index on education plus six more variables. The full model is of the form.

$$\begin{aligned}CrimeIdx_i = \beta_0 + \beta_1 EducIdx_i + \beta_2 Ln(GDPpcap)_i + \beta_3 UnempRate_i \\ + \beta_4 GiniIdx_i + \beta_5 PopAge_i + \beta_6 ShareofDrug_i + \beta_7 PopDensity_i \\ + u_i\end{aligned}$$

The variable GDP is logged for various reasons. Since there is a very wide range of GDP per capita observations across the sample (standard deviation of 23.7 thousand of dollars), logging this variable helps narrow its range and minimize any adverse effects of outliers, as in Wooldridge (2009). Furthermore, the use of the log for a positive monetary amount is consistent with most empirical work and makes sense in this context, as we are predicting a percentage point change in crime index, therefore, measuring the effect of a percentual change in GDP per capita is more intuitive.

REGRESSION ANALYSIS - ASSUMPTIONS

Following the first glance of what our main model is, it is of extreme importance to analyze each one of the six Gauss Markov Assumptions, in order to assure that we got the best linear unbiased estimators and also to be comfortable on making the usual inferences using the t and F statistics.

To start with, as our model can be written in the form of $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + u$, with the betas being our constant parameters of interest that capture the effect of each independent on the dependent variable, and u being an unobserved random error, we can, therefore, confirm the first assumption MLR.1: Linearity in

Parameters. Next, we can verify the MLR.2: Random Sampling, by the fact that the data contains information of each variable from countries all over the world, as we did not need to further calculate any method of picking and choosing. The model for the random sample is $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + u_i$. The following assumption is that none of the independent variables are constant and that there are not exact linear relationships among the independent variables, that is, MLR.3: No Perfect Collinearity. This can be easily confirmed when looking to the correlation matrix, in Table 2. It is important to note that assumption MLR.3 does allow the independent variables to be correlated, as they just cannot be perfectly correlated.

So far, so good, the first three assumptions hold with no big issues. However, the more delicate ones need to be carefully discussed. The first one is MLR.4: Zero Conditional Mean. This assumption states that the error u has an expected value of zero given any values of the independent variables. It is hard to believe that all relevant factors affecting crime index have been controlled for, and in fact, some of these uncontrolled factors may be correlated with our variables, so that MLR.4 does not hold and there is an omitted variable bias. One strong example may be how the literacy rate affects the crime index, which is also correlated with the education index. It is expected that a higher literacy rate would result in a lower crime index, and also that it might be positively correlated with education index. Thus, the omission of this rate may create a negative bias regarding the education index. Although this may be the case, the bias originated would not be that big and, on the other hand, efficiency would be lost in the OLS estimators if we considered a huge number of variables. Thus, for the sake of simplicity of the further tests and results, we assume that MLR.4 holds, so that the expected value of the error is zero, given any value of the independent variables.

Another complication that arises is whether the error u has the same variance given all the values of the independent variables, that is, whether or not assumption MLR.5: Homoskedasticity holds. In order to check for it, we will need to run a specific test. We can do it using two alternatives, first with the Breusch-Pagan Test, or second with the use of the White-Test (in this case we used the “lazy” one). In both, we have to regress the crime index on all of the previously stated independent variables, picking then the residuals $\hat{u}$, and computing $\hat{u}^2$. What differs is that in the former, we regress $\hat{u}^2$on all of the explanatory variables, while in the latter we regress the same squared error on $\hat{y}$ and on $\hat{y}^2$, thus capturing all the products within the variables. Despite the differences, when testing for homoskedasticity, the results will be the same as we can observe from Figure 2 and Figure 3. This can be stated from the fact that the F statistic is higher than a significance level of 5%, meaning that we fail to reject the null hypothesis of homoskedasticity. We, therefore, conclude that we are in the presence of a homoskedastic model.

Lastly, we will need to see if the error term, independent of all the explanatory variables, is normally distributed with mean zero and variance σ^2, which is the same as saying if MLR.6: Normality holds. It is pretty obvious that Crime Index, as we observe in Figure 3, is a variable that can only take positive values, thus, strictly speaking, it cannot have a normal distribution, breaking then the assumption. However, even though it does not follow a normal distribution, we can use the central limit theorem (assuming MLR.1-MLR.5 hold), to conclude that the OLS estimators are asymptotic normal, meaning these are approximately normally distributed, in large enough sample sizes. However, it is worth noting that a sample size of 60 observations is quite limited for a study of this nature, even though we can apply the central limit theorem and conduct inference asymptotically, the results would be more reliable had we included more observations. Concretely, an increase in sample size is likely to increase the variability in the independent variables, decreasing the variance of each OLS estimator.

We now have all the tools needed, and assumptions verified, in order to make inference and check if the model and its variables are significant or not.

REGRESSION ANALYSIS – RESULTS

Simple Regression Model

Education Index. Looking at the results, the Education Index is individually statistically significant (p-value of 0.4%). Its estimated coefficient is -48.58, however education is measured from 0 to 1 and crime from 0 to 100, so we scale up education in this analysis. This implies that an increase of one unit (from 0 to 100) in the level of education, is estimated to decrease crime rate by, on average, 0.4858 points, other things remaining the same. However, its scatter plot graph (figure 6) and R-squared are not that attractive as the representative dots are not concentrated around the mean and only 13.24% of the variation in the Crime Rate is explained by the Education Level (figure 5), so it makes sense to include more variables.

Multiple Linear Regression Model

Education Index. Looking at the results of the multiple linear regression, (figure 7) one can observe that education became statistically not significant given that the t-statistics because the -value test fails in all three percentage levels (1%, 5%, and 10%). Its coefficient was 19.02, implying that a unit increase (from 0 to 100) in education level is expected to decrease the crime index by approximately 0.19 points, on average, ceteris paribus. It is strange that this coefficient is positive, education is presumed to have a negative impact on crime, not positive. However, if we look at the 95% confidence interval of the coefficient, it includes values from -36 to 74, due to a very high standard error (probably related with the low sample size causing low variability in the dependent variable), so the results would vary drastically from sample to sample, meaning that our results are not very reliable.

Gini Index. In our regression model, the Gini Index is seen to be statistically significant at the 10% level (see figure 7), with a p-value of 6.3%. In fact, a percentage point increase in the Gini Index is predicted to increase the Crime Index by 0.415 points, on average, ceteris paribus (because the Index is measured from 0 to 100). Its coefficient is in accordance with its theoretical value (it has a positive impact on the Crime Rate Index). One of the reasons that may explain the impact is that higher levels of Gini Index, demonstrate higher disproportions of income and welfare distribution, meaning that high-income individuals receive a large percentage of the total income of the population. If income is concentrated in a small part of the population, it is expected that poor people have less disposable income, and the poorer they are, the higher the chance of committing a crime.

In our sample collection, countries/locations with higher disproportions in income distribution across the population tended to have the Crime Index relatively higher, with the exception being **Hong Kong**, which has one of the highest values of Gini Index, but the second lowest rate of Crime. Obviously, given other favorable situations and conditions of this city, it does make sense to have lower crime rates due to its huge development in all areas. Those are factors included in the error term.

GDP per Capita. Although the disposable income matters for the living conditions and, consequently, the Crime Rates, the results show that the logarithm of GDP per Capita is not statistically significant at any percentage level. Its t-statistic is extremely small and the P-value very high. Moreover, its coefficient is in accordance with its theoretical prediction (which is negative). In fact, as we have analyzed before for the Gini Index, it’s more important to understand the distribution of wealth in a country, than how much wealth is produced per person if we want to predict crime rates. Therefore, knowing the Gini Index is statistically more significant than GDP per Capita.

Unemployment Rate. Looking at the regression, the unemployment rate is statistically significant at the 5% level. Furthermore, its coefficient sign assimilates to its theoretical assumption, as this explanatory variable is an indicator of the general performance of the economy. It is reasonable to assume that a higher unemployment rate leads to a higher crime rate, as some studies mentioned in the literature review suggested. A one percentage point increase in the unemployment rate is expected to increase the crime rate by 0.55 points, on average, ceteris paribus, since unemployment was measured from 0 to 1, we divided the obtained estimate of the coefficient by 100.

Share of Drug Use. The share of drug usage in percentage of population is seen not to be statistically significant. Drug usage is an important indicator of the mental health of the population. In other words, an increase in the share of drug usage should increase the crime index. Looking at the results, the estimated coefficient sign and impact is also compatible with our assumptions.

Median Age and Population Density. According to some researches done between age and crime rate, most of the crimes are committed by people in adult age (15-64). Moreover, it is also important to note that high population concentration in urban areas/rural areas may have an extended impact on the number of crimes. These two explanatory variables may have a high correlation in our regression model and we expect them to move in the same direction given that countries with high population concentration also have a high share of the adult population. Looking at the results, Median Age is statistically significant at 1% level and Population Density is shown to be statistically significant at only 10% level. However, in our model, the respective coefficients are negative, contradicting the initial theoretical predictions (an increase in one year of age and population density increases the number of crimes). On one hand, an increase in one year of median age is expected to decreases the Crime Index by approx. 1.365 points, on average, ceteris paribus. On the other hand, an increase in one unit of population per sqr meter is expected to decrease the Crime Index by, on average, approx. 0.00211 points, all else remaining the same. It is worth mentioning that the 95% confidence interval for the estimator of population density contains positive and negative values, so our estimate being negative instead of the theoretical prediction may be due to sample variability.

Judging by the selection of countries (due to unavailability of data) in our model, we can find reasons that may explain the sign of the coefficients. First of all, in our 60 observations, most of the countries are either developed or in an advanced phase of development, impeding the analysis of certain countries (mostly African Countries), which will demonstrate the theoretical value of these variables. Furthermore, the association between the developed countries population density and share of population (high) and its respective Crime Index (low) turns the result around (i.e. Singapore and Hong Kong, locations with high GDP per Capita). Finally, there are other explanations that are omitted in this study (i.e. Venezuela´s Crime Rate is extremely high (due to political problems), even though the values of both explanatory variables are small).

Overall Fit and Significance and R-Squared

Using the F-statistics (figure 7), one can conclude that the explanatory variables are jointly statistically significant in the model, meaning that at least one of the variables included in the model is relevant. We can also confirm that the model is significant because the p-value (table) is extremely small. Regarding the R-squared, approximately 55.04% of the variation in the Crime Index in the 60 observations in 2013 is explained by the variables chosen. Also, the adjusted R-squared reveals that our variables have explanatory power, given that it is around 49%. Even though our model is statistically significant, we have to take into consideration that our analysis is simplified and that it only constitutes a starting point for a more complex study.

CONCLUSION

The main purpose of this study was to analyze the impact of levels of education in a country on its crime index. Following the work of Rathie, Sipos, and Ahuja (2016), we compiled data on crime and education for 60 countries. However, unlike the previously mentioned authors, we decided to include more variables in our model in order to test further hypothesis and also control for more factors that could influence both education and crime rate leading to biases in our estimators.

We concluded that, after controlling for median age, natural logarithm of GDP per capita, unemployment rate, Gini index, share of drug use and population density, education was not statistically significant at explaining differences in crime rate across countries. Out of all the variables we included, 4 of them were significant at the 10% level: median age, unemployment rate, Gini index and population density with the first two also being significant at the 5% level. The variables were jointly significant with an R-squared of 0.5504, so we managed to explain a lot of variation in crime index using our model.

One limitation of this work consists of its low sample size. With only 60 observations, the variances of our estimators were considerably high, leading to unlikely values of our estimates. Indeed, in several variables the 95% confidence intervals included large positive and negative values, which implies that the estimate we obtained is not very reliable, since it would vary substantially from sample to sample. A further study should focus on including more countries, although data can be quite limited on the variables we chose, and, more importantly, more years of analysis since focusing only on 2013 limited our sample size greatly.

Additionally, it would be interesting to analyze crime rates in more detail by focusing on different types of crimes. Presumably, the impact of education on theft will be quite different from its impact on homicide or financial crime, for example and it would be interesting as well as beneficial for policy makers to understand those differences. Also, following the work of Gonzales (2015), a study that includes different measures of education could yield different insights (using for example, % of population with a college degree or average years of education, instead of choosing a wider variable such as the education index). We could also have included dummy variables on gender and developed vs developing countries, for instance, to study the impact of education on different scenarios.

Even though we found some variables to be statistically significant and our model has strong explanatory power, there is still a lot of room for further studies to improve on our work and come up with more complete and statistically interesting results in the future.

BIBLIOGRAPHY

Numbeo:

https://www.numbeo.com/crime/rankings_by_country.jsp?title=2013-Q1&fbclid=IwAR3GWIS-GeWdUb_e2XE_oztidDxRDg5exGM2vjeSaJOlWFSd3OvX2STNZEI

United Nations:

https://www.un.org/en/development/desa/population/about/index.asp

World Bank:

https://data.worldbank.org/

Delotto and Ortranto (2010):

https://www.researchgate.net/publication/227353658_Does_Crime_Affect_Economic_Growth

Whitley and Prince (2005):

https://www.ncbi.nlm.nih.gov/pubmed/15908089

Rathie, Sipos and Ahuja (2016):

https://smartech.gatech.edu/handle/1853/54739

Gonzales (2015):

https://as.nyu.edu/content/dam/nyu-as/politics/documents/Gonzalez.pdf

APPENDIX

Figure 1 – Variables Summary

Variable	Obs	Mean	Std. Dev.	Min	Max
CrimeIdx	60	40.8265	13.96928	13.11	85.7
EducIdx	60	.7578506	.1046131	.4726667	.9265355
MedianAge	60	36.87	6.446423	24.7	46.3
lnGDPcap	60	9.805631	.9865675	7.280835	11.54306
UnempRate	60	.0890167	.0630137	.003	.275
GiniIdx	60	37.01833	8.040417	24.6	63.1
ShareofDrug	60	.050915	.0198638	0	.1113
PopDensity	60	353.9588	1296.555	1.846795	7636.722

Table 1 – Correlation matrix of independent variables used in regression analysis

	CrimeIdx	**EducIdx**	**MedianAge**	**lnGDPcap**	**UnempRate**	**GiniIdx**	**ShareofDrug**	**PopDensity**
CrimeIdx	1							
EducIdx	-0.3639	1						
MedianAge	-0.6246	0.6828	1					
lnGDPcap	-0.3706	0.765	0.522	1				
UnempRate	0.1135	0.0813	0.2693	-0.161	1			
GiniIdx	0.4657	-0.4792	-0.5893	-0.2645	-0.091	1		
ShareofDrug	0.0256	0.4046	0.2138	0.0482	0.0953	-0.2923	1	
PopDensity	-0.3156	0.0144	0.1424	0.1863	-0.1694	0.1735	-0.36	1

Figure 2 – Breusch-Pagan Test for Homoskedasticity

Source	SS	df	MS
Model	111156.334	7	15879.4763
Residual	1439543.15	52	27683.5221
Total	1550699.48	59	26283.0421

Number of obs =	60
F(7, 52) =	0.57
Prob > F =	0.7740
R-squared =	0.0717
Adj R-squared =	-0.0533
Root MSE =	166.38

uhatsq	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
EducIdx	361.2743	459.9672	0.79	0.436	-561.7174	1284.266
MedianAge	-.3397563	5.726823	-0.06	0.953	-11.83147	11.15196
lnGDPcap	-27.32809	41.45992	-0.66	0.513	-110.5235	55.86732
UnempRate	-153.88	404.0447	-0.38	0.705	-964.655	656.8949
GiniIdx	3.809479	3.645021	1.05	0.301	-3.504791	11.12375
ShareofDrug	-1734.968	1420.701	-1.22	0.228	-4585.813	1115.877
PopDensity	-.0234843	.0198078	-1.19	0.241	-.0632315	.0162629
_cons	62.30388	321.887	0.19	0.847	-583.6097	708.2174

Figure 3 – White-Test Test for Homoskedasticity

Source	SS	df	MS
Model	140245.33	2	70122.6651
Residual	1410454.15	57	24744.8097
Total	1550699.48	59	26283.0421

Number of obs =	60
F(2, 57) =	2.83
Prob > F =	0.0671
R-squared =	0.0904
Adj R-squared =	0.0585
Root MSE =	157.3

uhatsq	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
yhat	.0253499	1.015229	0.02	0.980	-2.007611	2.058311
yhatsq	.0065102	.0060045	1.08	0.283	-.0055135	.0185339
_cons	23.56914	50.4174	0.47	0.642	-77.38996	124.5282

Figure 4 – Crime Index Histogram

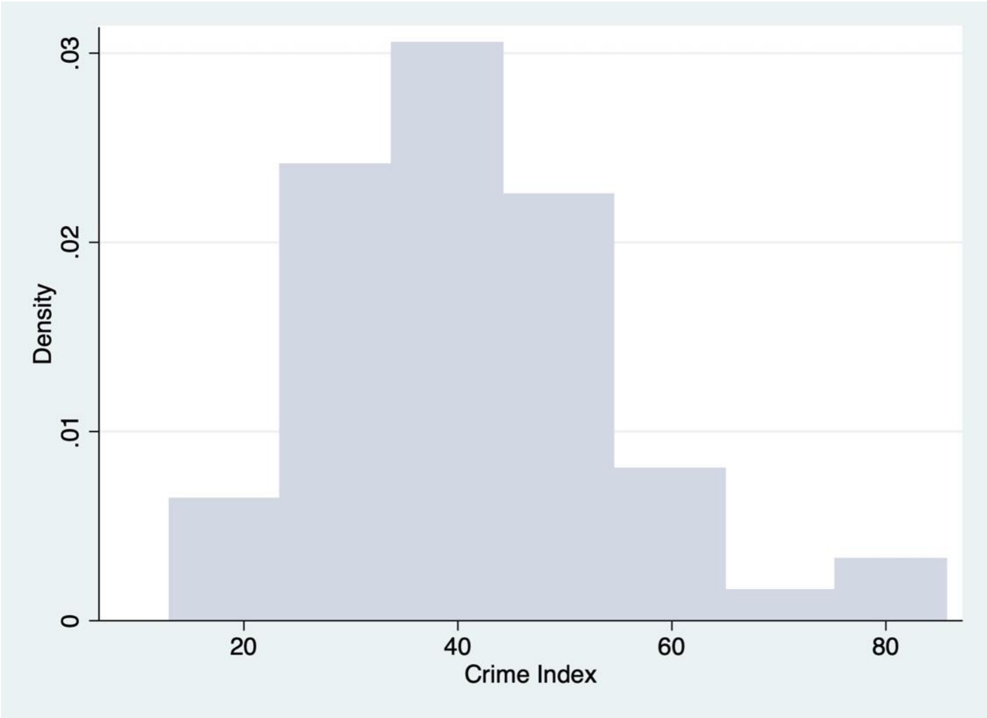

Figure 5 – Simple Linear Regression: Crime Index on Education Index

Source	SS	df	MS
Model	**1524.02736**	**1**	**1524.02736**
Residual	**9989.27461**	**58**	**172.228873**
Total	**11513.302**	**59**	**195.140711**

Number of obs	=	**60**
F(1, 58)	=	**8.85**
Prob > F	=	**0.0043**
R-squared	=	**0.1324**
Adj R-squared	=	**0.1174**
Root MSE	=	**13.124**

CrimeIdx	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
EdudIdx	**-48.58301**	**16.33206**	**-2.97**	**0.004**	**-81.27519**	**-15.89084**
_cons	**77.64516**	**12.49268**	**6.22**	**0.000**	**52.63834**	**102.652**

Figure 6 – Scatter Plot: Crime Index and Education Index

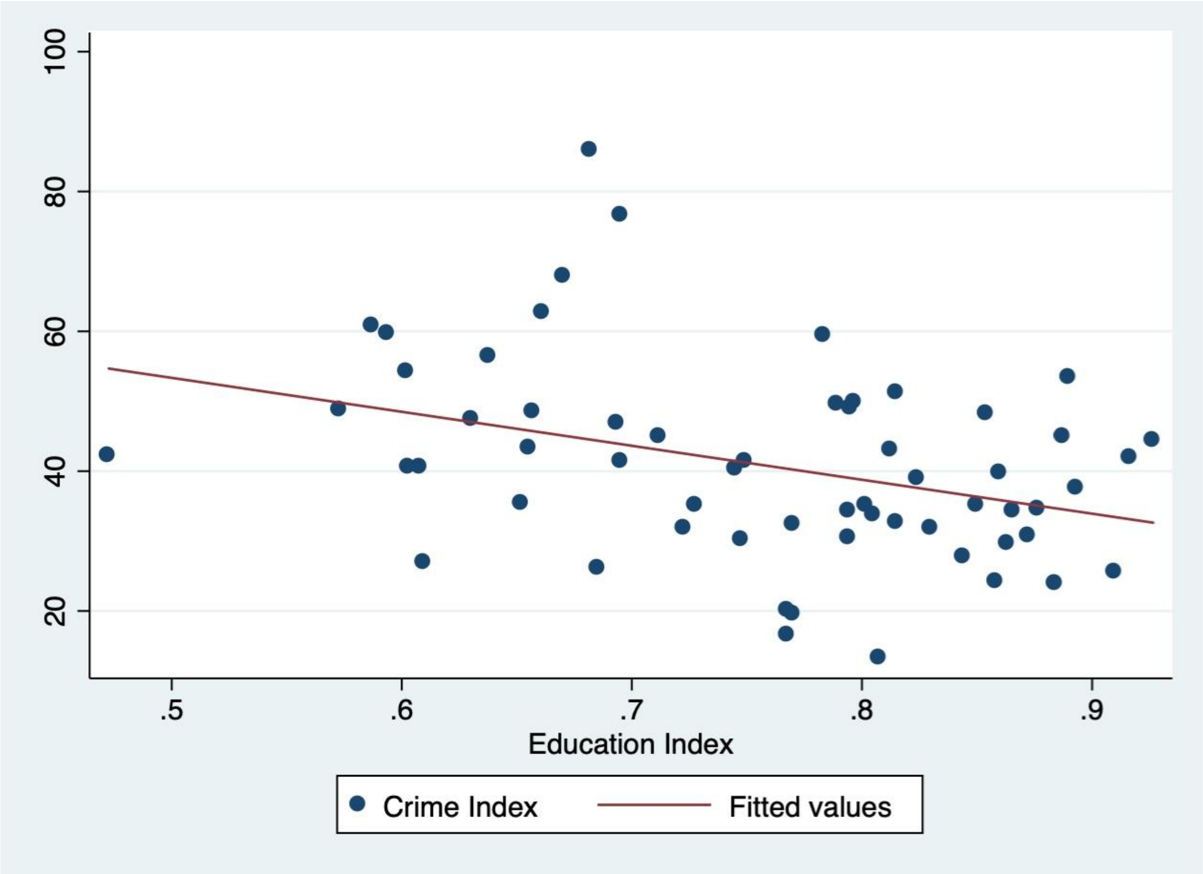

Figure 7 – Multiple Linear Regression

Source	SS	df	MS
Model	6336.85501	7	905.265002
Residual	5176.44695	52	99.5470568
Total	11513.302	59	195.140711

Number of obs	=	60
F(7, 52)	=	9.09
Prob > F	=	0.0000
R-squared	=	0.5504
Adj R-squared	=	0.4899
Root MSE	=	9.9773

CrimeIdx	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
EducIdx	19.02321	27.58229	0.69	0.493	-36.32471	74.37113
MedianAge	-1.365064	.3434135	-3.97	0.000	-2.054174	-.6759547
lnGDPcap	-.2070306	2.486177	-0.08	0.934	-5.195909	4.781848
UnempRate	55.47724	24.22886	2.29	0.026	6.858485	104.096
GiniIdx	.4145288	.2185765	1.90	0.063	-.0240771	.8531347
ShareofDrug	55.27026	85.19345	0.65	0.519	-115.6829	226.2234
PopDensity	-.00211	.0011878	-1.78	0.082	-.0044935	.0002735
_cons	56.41894	19.30221	2.92	0.005	17.68622	95.15165