Estimation of the effect of Income Inequality on Human Development: A Cross Sectional Study.

Ishan Parikh, Vivek Srinivasan, Jay Patel

Abstract

Throughout history, there has been overwhelming evidence that high levels of economic inequality correlate with low levels of human development, measured by the human development index. Our group's overall objective was to observe the effect of income inequality on human development levels by compiling GINI coefficient and Human Development Index statistics on 78 randomly selected countries from across the world. We ultimately found that the GINI index predominantly has a negative relationship with the human development index, but more so in developed countries than in developing nations. This is mainly due to the fact that globalization effects have not fully reached all of these countries' populations; as a result, these populations are still transitioning out of abject poverty and therefore have low levels of economic inequality. In addition, after running additional variables, including urban population, pollution levels, GDP Growth, external debt to GDP ratio, and index of economic freedom, we find that there are differences in significance with these variables with respect to their impacts on human development levels.

1.Introduction

Throughout the world, economic inequality has consistently been a hot topic of debate for governments and politicians. In addition, every country has always wanted to be able to boast high levels of human development. The US has traditionally had significantly high standards of living, with a relatively small level of inequality compared to developing nations across the globe, as many so-called "poor" US citizens have the luxury to purchase a wide array of consumer goods and accessibility to education and health insurance. So, what is the relationship between economic inequality and human development, and how much of an impact does such inequality have on overall human quality in different parts of the world? Governments in the past such as in the Soviet Union have had complete control over their countries' economies and politics, leaving the middle and lower classes to suffer significantly in the wake of the rich's luxury. The United States, on the other hand, has significantly better levels of equality and an actual middle class, despite a having a larger GINI coefficient. However, when we look at other countries around the world, we see that countries, such as Rwanda, with high income inequality, suffer from low standards of human development and severe poverty. We know that humans, and thereby countries, are unequal in more ways than just income, such as culture and genetic talent, but it is difficult to quantify the extent to which those factors affect income production, life expectancy, and education levels. This leaves us with just income inequality as the easiest quantifiable factor that can be included in the analysis. So the question becomes: Is income inequality a factor in producing low human development? And if it is, is income inequality the main cause of low human development or are there other, more significant factors?

To start our analysis we have to decide what form of human development to analyze. Because of the lack of available data on factors such as nutrition or education, we choose the Human Development Index as our main measure of human development. The reasoning for this is the Human Development Index is an all-encompassing measure that represents not only the developments earned through economic growth but also any development produced as a result of increases in education and public health, such as years of schooling and life expectancy, respectively. We include this to accurately represent the total amount of development a population has achieved at any point in time while controlling for exogenous effects such as economic shocks. It is our rationale that as income inequality increases, the economic disparity of a population increases; therefore, resulting in a decrease in the incentive to pursue higher education and a decrease in the living conditions of the population, which lowers life expectancy. The combination of these effects lowers the overall development of a population. For the multiple regression analysis, we wanted to generally introduce a host of other important variables that we believed could potentially affect income inequality and the human development index. First off, we decided to use urban population as an additional variable because of its relevance to urban economic theory, specifically studies on the urban wage premium and Central Place Theory (CTM) which will be discussed later. Put simply, large urban population sizes lead to positive effects on employment, consumption, overall GDP, which leads to higher levels of overall income across these populations- this leads to higher human development index levels. We then decided to include another concrete variable- pollution. Once again, simply put high pollutant levels lead to higher deaths and a lower standard of living, meaning a decrease in the human development index. The next variable we included was GDP growth, and higher GDP growth rates naturally lead to increased average levels of income in populations, and this leads to higher levels of human development measured by the index. Another additional quantifiable variable we included was the external debt to GDP ratio which measures exactly what it says. This variable is interesting because it takes into account a nation's balance of accounts surplus or deficit and then can be used to analyze a government's respective economic policies afterwards. The last variable we decided to include may have been the most reliable in determining its effects on income and human development indices. We believed economic freedom is a crucial impetus for higher levels of human development, and even though this is not always the case (for example, China), it is definitely a good predictor of human development levels. Typically, more intervening governments create societies where economic freedom for everyday citizens is strictly limited, and therefore human development levels for these developing nations' economies are very low.

2.Literature Review

Barro (2013) used factors such as male upper level schooling, government consumption as a percentage of GDP, rule-of-law index, openness ratio, inflation rate, fertility rate, and investment as a percentage of GDP as determinants of economic growth and investment. These determinants were analyzed in around 100 countries from 1960 to 1995. The data showed a pattern of conditional convergence with the growth rate of GDPPP is inversely related to the starting level of GDPPP, holding fixed other variables such as measures of government policies and institutions, initial stocks of human capital, and the character of the national population. There is little to no relation between years of schooling at the primary level to growth. However, growth has a positive correlation with starting level of the mean schooling years at secondary and further education for males while there is no meaningful correlation with females. Individuals with this level of education are more open when it comes to new technologies and help develop these technologies. The lack of female education significance shows that they are not being used will within the economy. Females have primary education does lead to increased economic growth through lower fertility rates. Data from scores on internationally comparable tests in science, mathematics, and reading were used to measure the quality of schooling. Scores on science tests have a particularly strong positive correlation with economic growth. If the quality of schooling is equivalent, then increasing the quantity of schooling will increase growth. But, quality of education is more important than quantity. Differences between rich nations and poor ones that emerge for the determination of economic growth are a higher convergence rate in rich countries, larger effects from international openness and terms-of-trade changes in poor countries, and more negative effects from government consumption in poor countries.

Brueckner and Lederman (2018) attempt to verify the prevailing economic theory developed by Galor and Zeira (1993). The prevailing theory is that poorer countries will have a positive correlation between inequality and aggregate output, while richer countries will have a negative correlation between inequality and aggregate output. The instrumental variables indicate that long term GDPPP growth and overall country growth are negatively impacted by inequality for countries with a GDPPP of 10,000 USD. Their model estimates illustrated that poor countries (nations will low initial GDP per capita) have a small to no relationship between inequality and GDPPP. They also show poor countries have a small to no relationship between inequality and human capital. For poor nations, income inequality has a positive correlation with growth. The main takeaway from this paper would be that income inequality helps growth for poor nations, but hurts growth for richer nations.

Ebenstein et al. (2015) examine the connection between income, pollution and mortality in China from 1991 to 2012. They found a strong positive correlation between a city's GDP and life expectancy. However, this is counteracted by the negative correlation between particulate air pollution and life expectancy. Most of the death causes the lowering of life expectancy come from cardiorespiratory illnesses. The gains in life expectancy have mostly been from infants and children while the decreases have been from those in adulthood and seniors who have had long term exposure.

Gould (2007) states that workers have higher wages in cities than in rural areas. This could come from cities making workers more productive or by the selection of workers with certain skills and abilities. The model in the paper works by choosing between urban or rural areas by career choices over time. The researcher controlled for all sources of selection and endogeneity, the results show that for white-collar jobs, the city pay better. Cities don't pay better for blue-collar workers. Because of this, the paper suggests that people move to cities not only because they like the location, but also to get a higher wage.

Lee (2010) test whether diversity in consumer products is a reason for lower wage premiums for high-skill workers. Product diversity is a luxury that is more valuable for high-skill workers. Therefore, these high-skill workers choose to live and work in large cities that can provide this diversity. A testable implication of the product diversity theory, different from productivity spillover theories, is that urban wage premiums (the difference between the wages of high-skill and low-skill workers) are decreasing and there may even be an urban wage discounts for high-skill workers. Testing data from the healthcare sector supports this theory finding that ability sorting accounts for 72% of the urban wage premium for the whole healthcare sector.

The contribution that our paper will have to the literature on the topic will mostly be on the comparative side of how different factors affect human development. Within this paper, we are including a vast array of variables which effect the three aspects of HDI (health, education, and income). These variables include the GINI index, urban population, pollution, GDP growth, external debt to GDP ratio, and the index of economic freedom. We wanted to use any different variables to reduce the likelihood of an omitted variable bias. Urban population was used as a factor on income, since wages in cities tend to be higher than those in rural areas. However, urban population can also affect health due to factors such as pollution. GDP growth and debt are used to see how much a nation can grow and how this growth is transferred to the average person's life in terms of health, education, and income. The index of economic freedom also provides a political angle which is generally lacking in other pieces of literature. This wide net of variables allows for a broad view of human development that maybe lacking in papers that focus in on just one aspect of HDI. The paper also asks whether some of these variables should be included in human development research. Should index of economic freedom be included in future research as vital factors or is it ignorable? We also hope to shed some light on the differences between developed and developing nations. Comparing the such variables with more common variables such as GINI, urban population, and pollution will also let future researchers to know what variables to include in their data sets and what to exclude. Due to the inherent differences in the economics of these two types of nations, the variables will have a different effect on them. While other papers have done this before, we are once again doing so with more variables that just GDP per capita.

3.Data

The data used in this analysis was gathered from a variety of sources including, The World Bank, the CIA, the United Nations, and the Heritage Foundation. This data was collected for 78 countries around the world for the 2010 year. List of countries found in **Appendix I.**

Table 1. Variables and Sources of Data used in Regression Models

Variables	Definitions	Abbreviation	Source
HDI	Human Development Index for year 2010	*hdi*	United Nations
GINI	Income inequality by country for year 2010	*gini*	World Bank
log(GDP)	log(GDP)	*lgdp*	
GDP	Real Gross Domestic Product by country for year 2010	*GDP*	World Bank
External Debt:GDP	External Debt to GDP ratio by country for year 2010	*debtratio*	CIA World FactBook
Urban Population	% of total population in cities for year 2010	*urbanpop*	World Bank
Index of Economic Freedom	Index of Economic Freedom score by country for year 2010	*econfreedom*	Heritage Foundation
Pollution	Amount of particulate matter in PPM for year 2010	*pollution*	World Bank

The Human Development Index is a composite measure of development in a country. Its creators emphasized that economic development is necessary but not sufficient to create a well-functioning free society. Freedom is central to the underlying motivation behind creating the index. The factors that are included in HDI are as follows: GNI per capita, the average of mean and expected years of schooling, and life expectancy at birth. The geometric mean of the three categories is taken to create the index score. A high HDI score indicates that a country's population is maximizing its potential human development. The GINI coefficient, which is our primary independent variable, is the measure of income inequality. The value comes from measuring the ratio of the areas under the perfect equality line, a line at 45 centered about the origin, and the Lorenz curve which plots an exponential curve, corresponding to a Pareto distribution. A Gini coefficient of 1 means that there is perfect inequality where one person has control of all income. A value of 0 is when a country has perfect income equality.

In order to control for other factors that may affect the human development index, additional environmental, economic, and social factors were considered. GDP and the External Debt to GDP ratio were added as economic variables. Urban population and Index of Economic Freedom were added as social variables. The last variable, pollution, was added as the environmental variable. GDP is the total measure of goods and services produced in a country annually. This variable was chosen because it is a widely used method of measuring national income since, in a general sense, total money spent on a good is equal to total income, hence, the total value of goods produced is equal to the national income. External Debt to GDP is the measure of how much a country owes to other country in proportion to its gross output. A low ratio indicates that a country's economy produces goods and services at a level where it can sufficiently pay back its debts without having to borrow money. It was included in this analysis because a high External Debt to GDP ratio indicates that a country will have to take monetary policy measures to solve the deficit through an increase in taxes, an increase in open market operations, or an increase in money supply. Examples such as the hyperinflated Germany economy in the post-World War I era shows that too much debt, as Germany was with war debts, can have extreme negative effects on the health and development of the citizens of such a country. Urban population measures the proportion of the total population that resides in urban areas i.e., the city. The reasoning behind this variable is that low skill workers benefit from higher wages due to positive wage premiums and high skill workers benefit from high consumption variety (Lee, 2010). In addition, both groups benefit from productivity and knowledge spillovers that result from the presence of universities and hospitals in cities (Christaller,1966). Higher wages and increased knowledge spillover effects mean that individuals that live in the city have greater incomes and are more likely to become better educated. The Index of Economic Freedom is an index created by the Heritage Foundation to measure the extent to which individuals are free to pursue goals and projects without the interference by a government entity of policy. In effect, the index is measuring the reduction in total welfare of society by the creation of deadweight loss through government intervention. A higher index score means less government intervention, and thereby higher the total welfare which contributes to an increase in the income portion of the HDI.

3.1 Descriptive Statistics

Table 2. Descriptive Statistics on all variables

Independent V.	Min	Max	Mean	Std Dev	Obs.
gini	24.8	63.4	36.8	8.5	78
urbanpop	15.5	97.6	63.9	20.6	78
pollution	5.4	90.4	22.1	15.7	78
lgdp	19.1	30.3	25.1	2.2	78
econfreedom	0	82.6	63.4	11.6	78
debtratio	.1	115.9	2.9	13.6	78
hdi	41.2	94.2	75.0	14.2	78
incomegroup	0	3	2.1	1	78

Table 2 includes descriptive statistics on all of the variables used in the regression analyses. Only 78 countries contained information on all of our variables. The explained variable, *hdi*, shows significant variation, ranging from 41.2 to 94.2. Such a wide variation in *hdi* points to the fact that the Human Development Index is worth looking at because given that more than half the countries in our dataset are developed countries, similar levels of industrialization do not fully explain differences in development between these countries. *pollution*, as well as the other explanatory variables, display large variations. We find this to be useful since it indicates that the variables we are studying vary country from country and thereby may be successful in explaining the notable variation we find in *hdi.*

Table 3. Descriptive Statistics on the Variables Including Dummy Variables

Independent V.	Min	Max	Mean	Std Dev	Obs.
gini	24.9	63.4	35.8	8.4	54
urbanpop	43.9	97.6	73.5	12.5	54
Pollution	5.4	36.0	16.6	7.2	54
lgdp	22.9	30.3	25.9	1.7	54
econfreedom	48.7	82.6	67.7	7.6	54
debtratio	.2	35.6	1.9	4.9	54
hdi	64.9	94.2	82.6	7.7	54
incomegroup	2	3	2.7	.5	54

To get a more detailed understanding of human development, the data was separated along the lines of stages of economic development, i.e., whether a country was developed or developing. Using the information set by the World Bank, countries that fit in the Upper Middle-Income bracket or higher were considered to be developed, the others were considered developing. The subset of data of containing the descriptive statistics on developed countries was generated and is shown above, in **Table 3**. The dummy variable *incomegroup* is a random variable which is given the value 0 for Low Income countries, the value 1 for Lower Middle-Income countries, the value 2 for Upper Middle-Income countries, and the value 3 for High Income countries. As displayed in **Table 3**, when considering the data on developed countries, the variation is less significant, but still notable, when compared to the values seen in **Table 2**.

3.2 Gauss-Markov Assumptions

Linear in Parameters Assumption

The model is linear in parameters such that it follows the model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Random Sampling Assumption

Data on the variables used in the regression were sourced from government data banks All countries that had available data were used in the regression analyses, therefore satisfying the random sampling assumption.

No Perfect Collinearity Assumption

In order to test for collinearity among the explanatory variables, correlation coefficients were computed using STATA. These coefficients, displayed in **Appendix II,** show that there is no perfect collinearity among the variables. The correlation coefficients remain at levels away from 1; therefore, there is no perfect collinearity among the variables and the collinearity condition is met.

Zero Conditional Mean Assumption

When we plot the residuals against the explanatory variable *gini* in the simple regression model, shown in **Appendix III**, we find that the residuals tend to sit below the x axis and therefore the expected error, $E(u|gini) \neq 0$. To address this fact, we add in several other variables in the multiple regression model to explain more of the data that is hidden in the error term in the simple regression model.

Homoskedasticity Assumption

To find if the simple regression model kept constant variance of the error term, given the x-value, *gini,* a residuals vs predicted value scatter plot was generated to visually confirm whether constant variance was maintained. It was not. As seen in **Appendix IV**, the residuals vs predicted value scatter plot shows that the data points tend to gather up and narrow towards to x-axis towards the right end. This shows that the variance or the error term does not remain constant.

3.3 Unfavorability of using *gdp and functional form*

A scatter plot of *hdi v gdp,* as shown in **Appendix V**, shows that the values of *gdp* are too large and data scaling on that variable will be necessary. However, even after scaling down *gdp* by 10^9 and generating a new variable denoted *GDP*, the scatter shows that the data still tends to stack up on the right side of the plot, making it visually ineffective to determine a relationship between the variables. Therefore, a functional form has been applied to the *gdp* variable, converting it from *gdp* to log(*gdp),* which is denoted *lgdp.* Using this form, a clearer relationship appears in the scatter plot. In addition, this functional form straightens out the interpretation. Using *lgdp* allows us to ask what happens to *hdi* when Gross Domestic Product increases by 1%, whereas in the cases of *gdp* or *GDP*, we could only ask what happens to *hdi* when Gross Domestic Product increases by $1 or $1 billion. The use of a percentage instead of a unit amount allows for a better understanding of how economic growth affects human development.

4.Results

The regression analysis was run on all the countries in the dataset. **Model 1** is the simple lienar regression of the Human Development Index on the Gini coefficient. STATA output located in **Appendix VI.**

Model 1: $hdi = 104.91 - .81(gini) + u$

As seen in **Table 4** below, there is a negative relationship between *gini* and *hdi* in **Model 1** as a 1 unit increase in the Gini coefficient results in approximately a .81 unit decrease in HDI. This fact remains consistent with our prediction. The effect of *gini* on *hdi* is significant all the way down to the 1% significance level with a t-value of -4.87. *gini* ,by itself, accounts for roughly 24% of the variation in *hdi* amongst all the countries in the dataset. Such a fact implies that *gini* plays a significant role alone in explaining *hdi* but also that other variables are needed to explain the remaining 75% of the variation in *hdi.* The constant is also positive, which fits our prediction, because if we hold *gini* at 0, meaning perfect income equality, the human development is high. However, the constant is 104.9, whereas, the Human Development Index is bounded in the interval [0,100]. This could possibly mean that the either the constant is purely a result of mathematical relation and not economic relation or that there is a factor notable factor that is not yet considered in the simple regression model that could provide a greater explanation for *hdi.*

Model 2 is the multiple regression model that only uses economic factors. In this regression, *hdi* analyzed against *gini*, *lgdp*, and *debtratio*. STATA output located in **Appendix VI.**

Model 2: $hdi = -3.61 - .61(gini) + 4.02(lgdp) + .08(debtratio) + u$

Shown in the second column of **Table 4**, this multiple regression model considers the effects of economic growth and government debt on the Human Development Index. Compared to **Model 1,** this regression model has a statistically insignificant constant at even 10% significance, implying that holding *gini, ldgp,* and debtratio at 0 leaves a *hdi* value of 0. In addition, a one unit increase in *gini* decreases *hdi* by .61 units, which implies that when accounting for more factors, income inequality may, in reality, have a smaller impact on human development than previously thought. A one percent increase in *lgdp* leads to a 4.02 unit increase in the *hdi.* **Model 2** shows that *lgdp* is the strongest predictor of *hdi* out of all the economic factors, while *debtratio* is insignificant at even 10% significance implying that it plays no role in predicting *hdi*. **Model 2** accounts for 58.92% of the variation in *hdi*.

Table 4: Table with models for all countries

$hdi=\beta_0+\beta_1(gini)+x\delta+u$	Model 1	Model 2	Model 3
gini	-.81*** (t=-4.87)	-.61*** (t=-4.79)	-.46*** (t=-6.75)
urbanpop			.26*** (t=7.04)
pollution			-.23*** (t=-5.27)
lgdp		4.02*** (t=7.90)	2.02*** (t=6.44)
econfreedom			.22*** (t=3.95)
debtratio		.08 (t=.96)	.03 (t=.60)
constant	104.91*** (t=16.64)	-3.61 (t=-.25)	15.76* (t=1.94)
R^2	.2376	.5892	.8915

Significance levels: 10%*, 5%**, 1%***

In the third regression analysis, economic, social, and environmental factors are considered. The variables *hdi* is regressed against are *gini, lgdp, debtratio, urbanpop, econfreedom,* and *pollution*. STATA output located in **Appendix VI.**

Model 3: $hdi = 15.76 - .46(gini) + .26(urbanpop) - .23(pollution) + 2.02(lgdp) + .22(econfreedom) + .03(debtratio) + u$

In this multiple regression analysis, *lgdp* continues to be the strongest predictor for *hdi*. Both *debtratio* and the constant are insignificant at 1% and 5% significance levels, but only the constant out of these two is significant at 10% significance. All the other variables are significant at 1%. *gini*'s impact on *hdi* decreases yet again, with a one unit increase in *gini* decreasing *hdi* by only .45 units, holding all other factors constant. A one percent increase in the urban population leads to a .22 unit increase in the Human Development Index. A similar result occurs for *econfreedom,* while a one unit increase in *pollution* leads to a predictable decrease of *hdi* by .23 units. This multiple regression model which has economic, social, and environmental factors combined accounts for 89.15% of the variation in *hdi.*

5. Extensions

5.1 F-tests

We conduct an F-test on the social variables to see if they are jointly significant on **Model 3.** STATA output on the regression without the social variables is in **Appendix VII.** The calculated F value is

$$\frac{.8915 - .7764}{1 - .8915} * \frac{71}{2} = 37.65$$

The F value for $F_{.05,2,71}$ is 3.15. Since 37.65> 3.15, the social variables, *urbanpop* and *econfreedom* are jointly significant. The F value for $F_{.01,2,71}$ is 4.97. 37.65 > 4.97, so the social variables are jointly significant at 1% as well. This implies that the social variables are not limited in their effect on *hdi.*

5.2 Dummy Variables

Introduced earlier in Section 3, a random variable, *incomegroup*, was generated to discern between developed and developing nations. A country is considered developed if it *incomegroup* variable takes on the values of 2 or 3. A value lower than 2, more specifically, either 1 or 0, indicates that the country is developing. STATA outputs located in **Appendix VIII.**

The mean *hdi* for developed countries is 82.6. When considering our dummy variable, our full regression model, **Model 3**, takes on the values seen in **Table 5** below. *gini* becomes statistically insignificant as does *econfreedom,debtratio,*and the constant. In developing nations, the distribution of income is likely to have no effect on *hdi. The* most impactful factors in developing nations are *lgdp, urbanpop, and pollution*, with a one percent increase in *lgdp* increasing *hdi* by 2.97 units, a one percent increase in urban population increasing *hdi* by .34 units and a one unit increase in particulate pollution decreasing *hdi* by .15 units, holding all other factors constant. This implies that increased agglomeration effects and economic growth, as well as, decreased pollution will make the greatest impact on Human Development in developing nations. On the other hand, Human Development in developed nations is additionally impacted by government intervention and income distribution schemes, due to *gini* and *econfreedom* being statistically significant in these nations.

Table 5. Multiple regression with all variables using dummy variable, *incomegroup*

$HDI=\beta_0+\beta_1(gini)+x\delta+u$	**Model 3 (developed only)**	**Model 3 (developing only)**
gini	**-.48*** (t=-9.61)**	**-.20 (t=-1.23)**
urbanpop	**.08** (t=2.30)**	**.34*** (t=4.53)**
pollution	**-.35*** (t=-5.84)**	**-.15*** (t=-2.17)**
lgdp	**1.17*** (t=4.87)**	**2.97*** (t=3.50)**
econfreedom	**.2*** (t=3.33)**	**.13 (t=1.27)**
debtratio	**.1 (t=-1.23)**	**.09 (t=1.59)**
constant	**56.1*** (t=7.43)**	**-19.26 (t=-.84)**
R^2	**.8824**	**.7441**

Significance levels: 10%*, 5%**, 1%***

6.Conclusions

In conclusion, GINI does a have negative correlation with HDI regardless of the development of the country though it is stronger in developed nations. When adding in additional variables such as the percentage of urban population, pollution, the log of GDP, economic freedom, and debt ratio, there is a decrease in not only the correlation of GINI, but also its statistical significance for developing countries. This may indicate that GINI is not an important factor of HDI for developing nations. The external debt to GDP ratio does not have a significant statistical presence within the model for either developing or developed nations. Other factors that actually did affect the HDI were urban population, pollution, log GDP, and economic freedom. Urban population and log GDP both had positive correlations with HDI which we expected. Both had more effect on developing nations than developed nations. This makes sense as countries that are developing have an influx of population to cities as well as higher growth rates. In developed nations, there is sustained economic growth and the fruits of industrialization have already been distributed across the populace regardless of whether they are in urban or rural areas. Economic freedom also had an overall statistically significant and positive correlation with HDI. It lost this significance with developing nations though. This may indicate that for developing nations, business do not require as much freedom to contribute to HDI. This could be due to already low wages and prioritizing growth over prosperity. In the developed world, however, having economic freedom is key to driving innovation and improvement. Pollution had a statistically significant negative correlation to HDI in all cases. This is to be expected since increase in pollution decreases health. An interesting result is that there is a higher negative correlation in the developed countries than in the developing one. Perhaps people in developed countries view pollution even worse than people in developing nations or it could be that developed nations have worse pollution than developing ones. This is something that future research can get into.

7.Bibliography

Barro, R. (2013). Education and Economic Growth. *Annals of Economics and Finance* **14(2)** 301- 328.

Brueckner, M. and D. Lederman (2018). Inequality and Economic Growth: The Role of Initial Income. *Journal of Economic Growth* **25(3)** 341-366

Ebenstein A. et al. (2015). Growth, Pollution, and Life Expectancy: China from 1991 -2012. *American Economic Review* **105(5)** 226-231

Gould, E. (2007). Cities, Workers, and Wages: A Structural Analysis of Urban Wage Premium. *Oxford University Press: The Review of Economic Studies* **74(2)** 477-506

Lee, S. (2010). Ability Sorting and Consumer City. *Journal of Urban Economics* **68(1)** 20-33

Appendix

Appendix I. List of countries in dataset

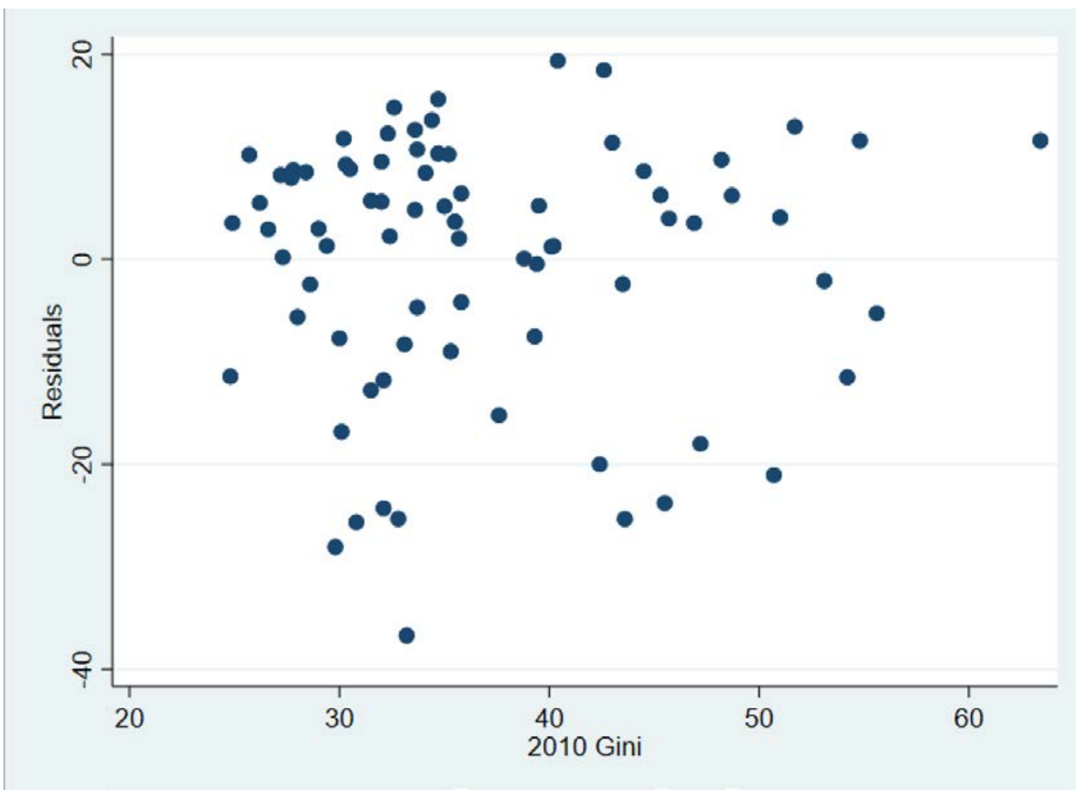
Argentina	Finland	Luxembourg	Romania
Armenia	France	Latvia	Russia
Australia	Great Britain	Moldova*	Rwanda*
Austria	Georgia*	Madagascar*	El Salvador
Belgium	The Gambia*	Mexico	Sao Tome and
Bangladesh*	Guinea-Bissau*	Macedonia	Principe*
Bulgaria	Greece	Malta	Slovak Republic
Canada	Honduras*	Montenegro	Slovenia*
Switzerland	Croatia	Mongolia*	Sweden
Colombia	Hungary	Malawi*	Thailand
Costa Rica	Ireland	Netherlands	Tunisia*
Cyprus	Iceland	Norway	Turkey
Czech Republic	Israel	Nepal*	Ukraine*
Germany	Italy	Pakistan*	Uruguay
Denmark	Jordan	Panama	United States
Dominican	Kazakhstan	Peru	of America
Republic	Kyrgyz*	Portugal	Vietnam*
Ecuador	Republic	Paraguay	Vanuatu*
Egypt*	South Korea	West Bank and	South Africa
Spain	Lesotho*	Gaza*	Zambia*
Estonia	Lithuania		

Developing nations*

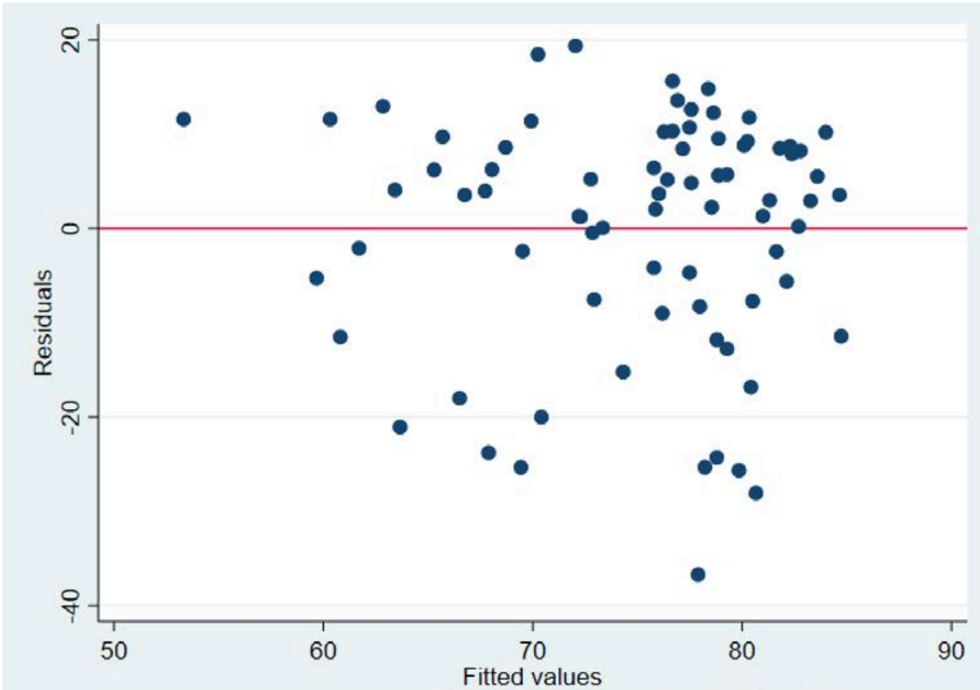
Appendix II. Correlation coefficients for all the explanatory variables.

	gini	gdp	urbanpop	pollut~n	debtra~o	econfr~m
gini	1.0000					
gdp	-0.0050	1.0000				
urbanpop	-0.2328	0.2071	1.0000			
pollution	0.1095	-0.1754	-0.5074	1.0000		
debtratio	-0.0480	-0.0376	-0.1434	-0.1437	1.0000	
econfreedom	-0.2082	0.2315	0.4146	-0.3278	0.0075	1.0000

Appendix III. Residuals vs *gini* plot to determine whether the zero conditional mean assumption was met.

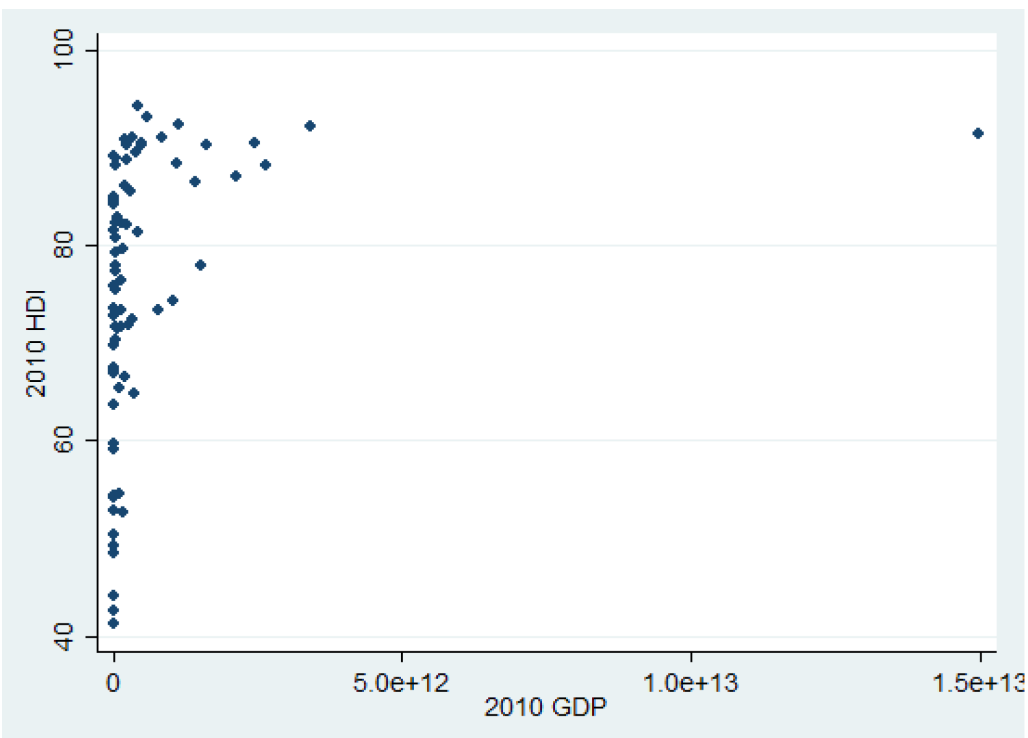


Appendix IV. A scatter plot showing residuals vs fitted values used to determine whether the homoskedasticity assumption was met.

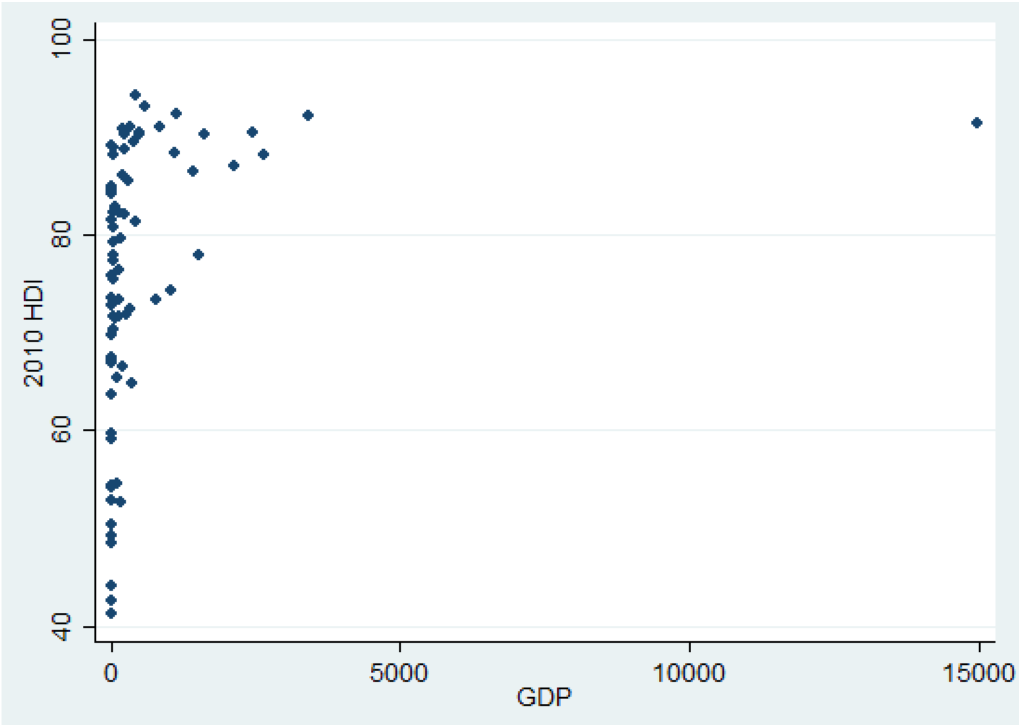


Appendix V. Scatter Plots

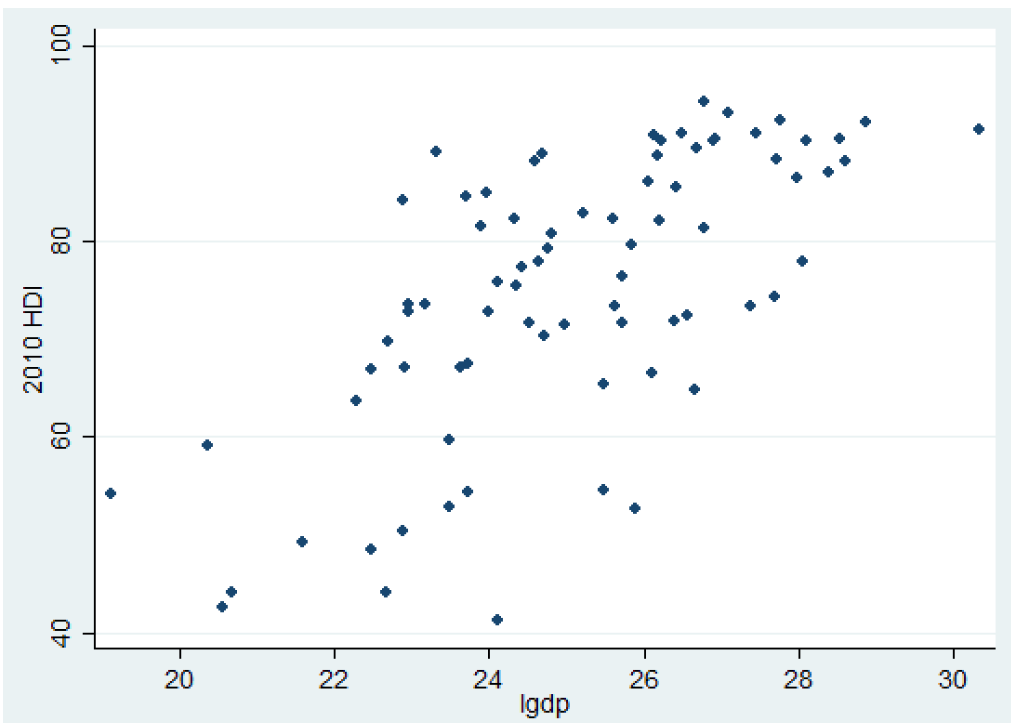
<u>Scatter plot between *hdi* and *gdp*</u>



Scatter plot between *hdi* and scaled *gdp*



Scatter plot between *hdi* and log(*gdp)*



Appendix VI. STATA outputs

Simple Linear Regression STATA output **Model 1**

```
regress hdi gini
```

Source	SS	df	MS
Model	3708.40164	1	3708.40164
Residual	11898.4821	76	156.558976
Total	15606.8838	77	202.686802

Number of obs	=	78
F(1, 76)	=	23.69
Prob > F	=	0.0000
R-squared	=	0.2376
Adj R-squared	=	0.2276
Root MSE	=	12.512

hdi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gini	-.8137094	.1671918	-4.87	0.000	-1.146701	-.4807181
_cons	104.9107	6.305101	16.64	0.000	92.35301	117.4684

Multiple Linear Regression STATA output **Model 2**

```
regress hdi gini lgdp debtratio
```

Source	SS	df	MS
Model	9195.8088	3	3065.2696
Residual	6411.07498	74	86.6361484
Total	15606.8838	77	202.686802

Number of obs	=	78
F(3, 74)	=	35.38
Prob > F	=	0.0000
R-squared	=	0.5892
Adj R-squared	=	0.5726
Root MSE	=	9.3079

hdi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gini	-.6098829	.1273044	-4.79	0.000	-.8635425	-.3562232
lgdp	4.0183	.508427	7.90	0.000	3.005237	5.031362
debtratio	.0771929	.0803701	0.96	0.340	-.0829481	.2373339
_cons	-3.610863	14.5677	-0.25	0.805	-32.63764	25.41591

Multiple Linear Regression STATA output **Model 3**

```
regress hdi gini urbanpop pollution lgdp econfreedom debtratio
```

Source	SS	df	MS			
				Number of obs	=	78
				F(6, 71)	=	97.22
Model	13913.3682	6	2318.89471	Prob > F	=	0.0000
Residual	1693.51555	71	23.8523317	R-squared	=	0.8915
				Adj R-squared	=	0.8823
Total	15606.8838	77	202.686802	Root MSE	=	4.8839

hdi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gini	-.4596817	.0681145	-6.75	0.000	-.5954981	-.3238653
urbanpop	.2553396	.0362912	7.04	0.000	.1829769	.3277023
pollution	-.2258849	.0428811	-5.27	0.000	-.3113874	-.1403824
lgdp	2.016415	.3129055	6.44	0.000	1.3925	2.640331
econfreedom	.2232982	.056555	3.95	0.000	.1105306	.3360657
debtratio	.0264855	.0438459	0.60	0.548	-.0609409	.1139118
_cons	15.76096	8.10374	1.94	0.056	-.3974373	31.91936

Appendix VII. STATA outputs used for Robustness Tests

```
regress hdi gini lgdp debtratio pollution
```

Source	SS	df	MS			
				Number of obs	=	78
				F(4, 73)	=	63.36
Model	12116.9193	4	3029.22982	Prob > F	=	0.0000
Residual	3489.9645	73	47.8077329	R-squared	=	0.7764
				Adj R-squared	=	0.7641
Total	15606.8838	77	202.686802	Root MSE	=	6.9143

hdi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gini	-.5710732	.094698	-6.03	0.000	-.7598061	-.3823404
lgdp	3.268812	.3896644	8.39	0.000	2.492212	4.045412
debtratio	-.0167157	.0608995	-0.27	0.784	-.1380882	.1046568
pollution	-.4100458	.0524575	-7.82	0.000	-.5145934	-.3054982
_cons	23.09181	11.34796	2.03	0.045	.4753568	45.70825

Appendix VIII. STATA outputs with dummy variables

<u>Multiple Linear Regression **Model 3** on developed countries only</u>

```
regress hdi gini urbanpop pollution lgdp econfreedom debtratio if incomegroup>=2
```

Source	SS	df	MS			
				Number of obs	=	54
				F(6, 47)	=	58.78
Model	2793.86022	6	465.64337	Prob > F	=	0.0000
Residual	372.307587	47	7.92143803	R-squared	=	0.8824
				Adj R-squared	=	0.8674
Total	3166.16781	53	59.7390153	Root MSE	=	2.8145

hdi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gini	-.4833516	.0502779	-9.61	0.000	-.5844977	-.3822055
urbanpop	.0766742	.0333828	2.30	0.026	.0095167	.1438318
pollution	-.3529689	.0604848	-5.84	0.000	-.4746485	-.2312893
lgdp	1.169713	.2402137	4.87	0.000	.6864653	1.652961
econfreedom	.1982523	.0595404	3.33	0.002	.0784725	.3180321
debtratio	.1021611	.0832684	1.23	0.226	-.0653534	.2696756
_cons	56.08475	7.550524	7.43	0.000	40.89506	71.27445

<u>Multiple Linear Regression **Model 3** on developing countries</u>

```
regress hdi gini urbanpop pollution lgdp econfreedom debtratio if incomegroup>=2
```

Source	SS	df	MS			
				Number of obs	=	54
				F(6, 47)	=	58.78
Model	2793.86022	6	465.64337	Prob > F	=	0.0000
Residual	372.307587	47	7.92143803	R-squared	=	0.8824
				Adj R-squared	=	0.8674
Total	3166.16781	53	59.7390153	Root MSE	=	2.8145

hdi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gini	-.4833516	.0502779	-9.61	0.000	-.5844977	-.3822055
urbanpop	.0766742	.0333828	2.30	0.026	.0095167	.1438318
pollution	-.3529689	.0604848	-5.84	0.000	-.4746485	-.2312893
lgdp	1.169713	.2402137	4.87	0.000	.6864653	1.652961
econfreedom	.1982523	.0595404	3.33	0.002	.0784725	.3180321
debtratio	.1021611	.0832684	1.23	0.226	-.0653534	.2696756
_cons	56.08475	7.550524	7.43	0.000	40.89506	71.27445