Factors affecting income inequality

Khrystyna Kubatska

Abstract

The paper explores the role of different factors in income inequality, as measured by the Gini coefficient. Data collected for over 140 countries. Four regression models were created to show the probable connection between various factors in economic, social and demographic areas and income inequality. Additional methods to improve the models and suggestions are also provided.

1 Introduction

Income inequality is defined as a measure that highlights the gap between the disposable income of different individuals or households in a given year. It has many adverse effects on various aspects of society. In countries with higher rates, there are higher rates of medical and social problems, such as obesity, mental disorders, prison, drug use. It also harms social cohesion, political stability and the well-being of the population. Inequality leads to a decline in the country's economic growth when human capital is neglected for high consumption.

In this research, I wanted to establish a specific explanation for the impact of particular economic and demographic factors on income inequality. It is essential to understand how specific criteria affect it in order to improve policies to target the specific allocation of public spending for specific areas that can thus diminish the gap. I hope this model will give a clear picture of the impact of individual factors on income inequality.

2 Literature review

Today, there is a large number of papers trying to trace the factors that influence income inequality.

Jose'de Gregorio and Jong-Wha Lee (2002), in their study, show the relationship between education factors and income distribution. Their research covers a wide range of countries between 190 and 1990 and indicates that higher education and its more even distribution play an important role in income inequality.

Their results confirm the Kuznets's inverted-U curve for the relationship between income level and income inequality, which means that the t the growth process inequality first rises and then declines. They also found that government social spending increases fluctuations in income distribution.

In their research, Martín González and Alicia Menendez (2000) developed a methodology that allows quantifying the effect of changes in unemployment on income inequality of labor. They assessed individual earning functions as determined by employment status. Next, they established wages that are established for the unemployed and calculating inequality measures, using actual and simulated aggregates, they estimated the impact of unemployment on income inequality. Their results indicate that unemployment accounts for much of the increase in inequality in earnings.

Duc Hong Vo, Thang Cong Nguyen, Ngoc Phu Tran and Anh The Vo (2019) conducted studies about causation and the dynamic link between income inequality and economic growth. Their study was made on different data between 1960 and 2014 for all available countries and middle-income countries. Using the generalized method of moments, they established the causality of economic growth and vice versa in the data sample.

Markus Brückner and Daniel Lederman (2015) showed that income inequality has a significant negative impact on aggregate production for the average sample country by using an econometric model, which includes the interaction between measures of income inequality and the initial level of GDroduct per capita. However, for developing countries, income inequality has a significant positive effect. They concluded that income inequality is advantageous to economic growth in developing countries, but that it is adverse to economic growth in advanced economies

Mark J. Perry (2013) researched that certain household demographics, including average household and marital status, age and household education, correlate very strongly with household income in the United States. In contrast, people in lower-income households are much more likely than their counterparts in higher-income households to be less educated, part-time, or very young (up to 35 years) or ancient (over 65) and living and incomplete households, and vise versa.

3 Data description

Many different factors can affect income inequality. To build the most accurate models, I decided to include different data for a set of over 140 countries.

Dependent variable

Gini coefficient: Gini coefficient is a measure of income inequality. It based on a comparison of aggregate proportions population with the aggregate share of income they receive and is the most popular measure of income inequality among countries. The index takes values between 0 and 1, where 0 denotes absolute equality and 1 denotes total inequality.

Independent variables

GDP per capita: It measures the income level of a country relative to the population. GDP per capita reflects the performance of a country, affecting the standard of living of its citizens, which will ultimately affect income inequality. I hypothesize that if GDP per capita goes down, then income inequality increases. To make GDP per capita more statically meaningful, I took its natural log of it.

Unemployment rate: It is the number of unemployed as a percentage of the workforce. In theory, it should have a negative effect on GDP per capita, which in turn would increase income inequality.

Growth rate: It is GDP growth for a specific period. Data are taken as a percentage relative to the initial year for five periods. I hypothesize that a decrease in growth rate contributes to income inequality.

Government expenditure on education: It shows how much a country spends on education. The data reflects how much the country values education. Government spending on education is correlated with the mean level of education. However, the mean level of education does not fully explain how much the country invests in education. Data are taken as a percentage of GDP and the logarithm of spending in foreign currency (in millions of US dollars). I suppose that small government expenditures on education stimulate income inequality.

Government expenditure on healthcare: It measures the percentage of GDP that a country spends on health care services or goods such as research and development of new medicines, excluding capital expenditures on health care, such as buildings and machinery. I expect if the percentage of GDP that a country spends on health care is large, then income inequality is relatively small.

Birth Rate: It reflects total live births per 1000 population. I assume that a high birth rate will lead to more inequality, as the costs for children will be much higher.

Life expectancy: I suppose a higher life expectancy has a negative effect on income inequality since longer life expectancy would affect more social programs in a country.

Taxes on income, profits and capital gains: It is a percentage of tax on income, profits and capital gains. Data were taken for 2018-2019 (latest available). I assume that if the tax rate is low, then income inequality is high.

Gender gap score: It measures the relative gaps between women and men in four key areas: health, education, economics and politics. The highest possible score is 1 (equality) and the lowest is 0 (inequality). I suppose if the Gender gap score in the country is high, then income inequality is relatively large too.

Individuals using the Internet: It shows the percentage of the population of a country that uses the Internet. I expect that the higher this indicator, the lower the income inequality in the country.

Corruption Perceptions Index: It reflects the level of corruption in the public sector, as determined by expert opinions and opinion polls. The index takes values between 0 and 100, where 0 means absolute corruption and 100 means no corruption. I hypothesis, the low Corruption Perceptions Index corresponds to high rates of income inequality.

Other data characteristics are listed in the table below.

Variable	Abbreviation	Units	Year	Source
Gini Index	GiniIndex	index	2020	Central Intelligence Agency
GDP per capita 2019	logGDP_pc_2019	M$	2019	World Bank Data
Unemployment Rate	unem	%	2020	Central Intelligence Agency
Growth Rate	growth13.18	%	2013-2018*	World Bank Data
Gov.Expend. on Education	eduĠDP	%	2018*	World Bank Data
Gov.Expend. on Education	edu_US.m	M$	2018*	World Bank Data
Gov.Expend. on Healthcare	health.GDP	%	2017*	World Health Organization
Birth Rate	Birth_rate	index	2020	World Bank Data
Life Expectancy	LE.both	years	2020	Worldometer
Taxes on income	IncomeTax	%	2018*	World Bank Data
Gender gap score	Gender_gap	index	2016	World Economic Forum
Individuals using the Internet	Net	%	2016	World Bank Data
Corruption Perceptions Index	CPI.2019	index	2019	Transparency International

* The data was taken the latest available period.

Summary statistics for the described variables are shown in the table below.

Table 1: Summary Statistics

Abbreviation	Observations	Mean	St.Dev.	Min	Max
GiniIndex	144	38.84	8.73	22.70	63.20
logGDP_pc_2019	134	8.659	1.47	5.617	11.637
unem	140	10.117	10.35	0.3	77.0
growth13.18	135	3.307	2.74	-11.76	9.4
edu.GDP	110	4.501	1.44	0.981	7.976
edu_US.m	109	13858.79	30089.4	16.38	167800.24
health.GDP	128	6.857	2.53	2.274	17.061
Birth_rate	132	19.92	10.13	7.30	46.08
LE.both	138	73.66	7.58	54.36	85.29
IncomeTax	100	26.875	12.22	2.027	65.188
Gender_gap	119	0.7012	0.06	0.5160	0.8740
Net	141	50.14	27.99	4.00	98.24
CPI.2019	139	28.95	16.39	2.00	64.00

Gauss Markov Assumptions

1.*Linear in Parameters*

The dependent variable can be written as a linear combination of the independent variables plus an error term. The scatter graphs between each of the independent variables and the dependent variable *(included in Appendix)*. All models in this paper satisfy this assumption.

2.*Random sampling*

The data has random sampling since it reflects variables for 144 counties. Moreover, none of the chosen countries are from any particular region or economic background.

3.*No perfect collinearity*

Correlations between some independent variables are high but not perfect *(Correlation Matrix included in Appendix)*. There is no perfect collinearity between the independent variables.

4.*Zero Conditional Mean*

This assumption means that the independent variables contain no information about the size of the error. However, in the next models, $R\hat{2}$ is quite small. So there will be unobserved factors not accounted for in the independent variables.

5.*Homoskedasticity*

The independent variables contain no information about the variance of the error. It is shown by scatter plots between each of the independent variables and the dependent variable.

4 Methodology explanation

The main purpose of this exploration is to determine what factors have a significant impact on income inequality using linear models and tests for statistical significance.

The Gini Index was chosen as the criterion for income inequality, as it is the most popular measurement in this area. Independent variables were chosen based on a review of literature about income inequality and intuitively. The data used in the analysis are cross-sectional and were taken from various sources. All variables were analyzed and checked on Gauss Markov Assumptions.

The ordinary least squares models were used for analysis because they are BLUE and produce unbiased estimates that have the smallest variance of all possible linear estimators.

For each variable in the models, the statistical significance of this variable was determined. An F test for the joint significance of some variables in one model was also performed. The conclusion was made using the results of the analysis.

The analysis was performed using R software.

All graphs, tables and code are attached.

5 Results

Model 1: First Multiple Regression:

Equation:

GiniIndex = $\beta_0 logGDP\_pc\_2019 + \beta_1 growth13.18 + \beta_2 unem + \beta_3 edu.GDP + \beta_4 log(edu\_US.m) + \beta_5 health.GDP + \beta_6 Birth\_rate + \beta_7 LE.both + \beta_8 IncomeTax + \beta_9 Gender\_gap + \beta_1 0 Net + \beta_1 1 CPI.2019$

After Regression:

GiniIndex = 52.14 + 1.51 logGDP_pc_2019 - 0.82 growth13.18 + 0.03 unem + 0.05 edu.GDP - 0.09 log(edu_US.m) + 0.14 health.GDP + 0.22 Birth$_rate - 0.3LE.both + 0.2IncomeTax - 8.05Gender_gap - 0.06Net - 0.12CPI.2019$

N = 77

$R^2 = 0.342$

See Appendix Output 1: First Multiple Regression

Table 2: Estimation Results - Model 1

Variable	Coefficient	Std. Error	T-value	P-value	H0:b=0 H1:b$\bar{0}$
logGDP_pc_2019	1.50	2.36	0.63	0.52	Fail to reject at 10%
growth13.18	-0.82	0.63	-1.28	0.20	Fail to reject at 10%
unem	0.03	0.10	0.32	0.75	Fail to reject at 10%
edu.GDP	0.04	0.91	0.05	0.95	Fail to reject at 10%
log(edu_US.m)	-0.08	0.68	-0.12	0.89	Fail to reject at 10%
health.GDP	0.13	0.59	0.23	0.81	Fail to reject at 10%
Birth_rate	0.22	0.21	1.01	0.31	Fail to reject at 10%
LE.both	-0.29	0.35	-0.82	0.41	Fail to reject at 10%
IncomeTax*	0.19	0.08	2.34	0.02	Reject at 10%
Gender_gap	-8.04	22.83	-0.35	0.72	Fail to reject at 10%
Net	-0.06	0.09	-0.65	0.51	Fail to reject at 10%
CPI.2019	-0.12	0.12	-0.99	0.32	Fail to reject at 10%

(*Statistically Significant at 10%, **Statistically Significant at 5%, ***Statistically Significant at

Interpretation Model 1

All independent variables were included in the first model in order to see which variables are most significant, according to their P-values and t-statistics. In the model, only taxes on income, profits and capital gains are statistically significant at 90% confidence interval. So, the Null hypothesis that b8 is equal to 0 at confidence interval 90% can be rejected. However, $R^2 = 0.342$.

It means that the model explains only 34.2% of the Gini Index.

So, the variables tested before were not necessarily significant by themselves, but they could be jointly significant. *(See Extensions)*

Model 2: Second Multiple Regression:

Equation:

GiniIndex $\beta_0 + \beta_1 logGDP\_pc\_2019 + \beta_2 log(edu\_US.m) + \beta_3 Birth\_rate + \beta_4 LE.both + \beta_5 Gender\_gap + \beta_6 CPI.2019$

After Regression:

GiniIndex = 44.50 + 0.4 logGDP_pc_2019 + 0.7 log(edu_US.m) + 0.17 Birth_rate -0.26 LE.both + 5.1 Gender_gap -0.12 CPI.2019

N = 95

$R^2 = 0.2049$

See Appendix Output 2: Second Multiple Regression

Table 3: Estimation Results - Model 2

Variable	Coefficient	Std. Error	T-value	P-value	H0:b=0 H1:b/=0
logGDP_pc_2019	0.40	1.56	0.25	0.79	Fail to reject at 10%
log(edu_US.m)	0.70	0.53	1.31	0.19	Fail to reject at 10%
Birth_rate	0.17	0.17	1.01	0.31	Fail to reject at 10%
LE.both	-0.25	0.27	-0.95	0.34	Fail to reject at 10%
Gender_gap	5.10	16.90	0.30	0.76	Fail to reject at 10%
CPI.2019	-0.11	0.08	-1.29	0.20	Fail to reject at 10%

(*Statistically Significant at 10%, **Statistically Significant at 5%, ***Statistically Significant at

Interpretation Model 2

In the second, model log GDP per capita, log government spending on education in millions, birth rate, life expectancy, gender gap and corruption index were included in the regression. In that case, no variable has a statistically significant effect on Gini index. Also, $R^2 = 0.2049$,

that is very small, since the model explains only 20.49% of the Gini Index.

Model 3: Third Multiple Regression:

Equation:

GiniIndex $= \beta_0 + \beta_1 logGDP\_pc\_2019 + \beta_2 growth13.18 + \beta_3 unem + \beta_4 edu.GDP + \beta_5 health.GDP + \beta_6 IncomeTax + \beta_7 CPI.2019 + \beta_8 Net$

After Regression:

GiniIndex = 48.69 -0.8 logGDP_pc_2019 -0.8 growth13.18 + 0.07 unem + 0.3 edu.GDP -0.12 health.GDP + 0.23 IncomeTax + -0.09 CPI.2019 -0.09 Net

N = 82
$R^2 = 0.3483$

See Appendix Output 3: Third Multiple Regression

Table 4: Estimation Results-Model 3

Variable	**Coefficient**	**Std. Error**	**T-value**	**P-value**	**H0:b=0 H1:b/=0**
logGDP_pc_2019	-0.79	1.79	-0.44	0.65	Fail to reject at 10%
growth13.18	-0.78	0.54	-1.44	0.15	Fail to reject at 10%
unem	0.07	0.09	0.78	0.43	Fail to reject at 10%
edu.GDP	0.30	0.78	0.39	0.69	Fail to reject at 10%
health.GDP	-0.12	0.54	-0.22	0.82	Fail to reject at 10%
IncomeTax**	0.22	0.07	3.03	0.01	Reject at 5%
CPI.2019	-0.08	0.11	-0.78	0.43	Fail to reject at 10%
Net	-0.09	0.08	-1.08	0.28	Fail to reject at 10%

(*Statistically Significant at 10%, **Statistically Significant at 5%, ***Statistically Significant at

Interpretation Model 3

In the third model were included such variables log GDP per capita, growth rate, unemployment, gov.spending on education (%), gov.spending on health, taxes on income, Internet share and CPI. In the model, only taxes on income are statistically significant at 5% confidence interval as well. So, the Null hypothesis that b6 = 0 can be rejected. R squared is equal to 0.3483 that is better than in previous models, but still small.

Model 4: Forth Multiple Regression:

Equation:

GiniIndex $\beta_0+\beta_1 IncomeTax+\beta_2 growth13.18+\beta_3 log(edu\_US.m)+\beta_4 Birth\_rate+\beta_5 LE.both+\beta_6 CPI.2019$

After Regression:

GiniIndex 55.9 + 0.21 IncomeTax -0.86 growth13.18 +0.07 log(edu_US.m)+ 0.27 Birth_rate -0.31 LE.both -0.09 CPI.2019

N = 83
$R^2 = 0.3825$

See Appendix Output 4: Fourth Multiple Regression

(*Statistically Significant at 10%, **Statistically Significant at 5%, ***Statistically Significant at

Interpretation Model 4

Table 5: Estimation Results - Model 4

Variable	**Coefficient**	**Std. Error**	**T-value**	**P-value**	**H0:B=0 H1:B/=0**
IncomeTax**	0.21	0.07	2.80	0.00	Reject at 5%
growth13.18 .	-0.86	0.48	-1.77	0.08	Fail to reject at 10%
log(edu_US.m)	0.07	0.59	0.12	0.90	Fail to reject at 10%
$\text{Birth}_{r}ate$	0.27	0.18	1.47	0.14	Fail to reject at 10%
LE.both	-0.30	0.27	-1.13	0.25	Fail to reject at 10%
CPI.2019	-0.09	0.07	-1.22	0.22	Fail to reject at 10%

Taxes on income, growth rate, log gov.spending on education (in millions), birth rate, life expectancy and corruption index were included in the model. Only taxes on income are statically significant on confidence interval 95%. In this model, The R squared is the highest in comparison with the previous ones, but still very small.

Extensions Robustness Test

Since all variables, except IncomeTax, were statistically insignificant in Model 1, the F-Test was made to determine if these variables are jointly significant. *See Appendix Output 5 for Restricted Model.*

The hypothesis is follows:
$\text{H}_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_9 = \beta_{10} = \beta_{11} = 0$
$\text{H}_1 : H_0 \neq 0$

Unrestricted Model:
$\text{GiniIndex} = \beta_0 logGDP\_pc\_2019 + \beta_1 growth13.18 + \beta_2 unem + \beta_3 edu.GDP + \beta_4 log(edu\_US.m) + \beta_5 health.GDP + \beta_6 Birth\_rate + \beta_7 LE.both + \beta_8 IncomeTax + \beta_9 Gender\_gap + \beta_{10} Net + \beta_{11} CPI.2019$

Restricted Model:
$\text{GiniIndex} = \beta_0 + \beta_8 IncomeTax$

Critical Value at 10% significance $\text{F}_{11,65} = 1.672$
Model F_{Stat} : 2.6094

F is higher than the critical value. So, the null hypothesis can be rejected. The variables are jointly statistically significant at the 10% level. Therefore they are essential for our model in explaining income inequality.

6 Conclusion

This project explores the factors contributing to income inequality, which is measure as the Gini index. In order to take into different dimensions, I used such variables as GDP per capita, GDP growth over several years, income tax rate, unemployment rate, the share of GDP spending on education (and US currency data) and health services, fertility rate, life expectancy, gender gap, percent of people using the Internet and corruption coefficient.

Four models that included different variables were built, although all of them do not explain income inequality to a greater extent, because of the small R squared.

Model 1 includes all of these factors above, but it does not show the best R squared among the four models. The income tax is the only variable that has a significant impact on the 90% confidence interval. Although the other variables used in this model are not significant in themselves, testing their using the f-test shown that they have joint significance on confidence interval 90%.

Model 4 has the highest R squared among the other models in this paper (but still quite low - 38.25%). It includes taxes on income, GDP growth rate, log government spending on education (in millions $US), birth rate, life expectancy and corruption index. In this model, the only variable that is statistically significant on confidence interval 95% is the taxes on income as well. Contrary to my assumption, increasing income taxes is correlated with an increase in the Gini index in these models.

In the models above, there is omitted many economic, political, demographic and psychological factors (for example, a person's personal qualities) that could have an impact on income inequality. Reducing the missing factors may make the model more relevant. Additionally, including more observations (in this case, countries and data for them for different periods) would strengthen the model.

References

Jlose'de Gregorioand and Jong-Wha Lee 2002, EDUCATION AND INCOME INEQUALITY: NEW EVIDENCE FROM CROSS-COUNTRY DATA

Martín González and Alicia Menendez, 2000, The Effect of Unemployment on Labor Earnings Inequality: Argentina in the Nineties*

Duc Hong Vo , Thang Cong Nguyen, Ngoc Phu Tran and Anh The Vo 2019, What Factors Affect Income Inequality and Economic Growth in Middle-Income Countries?

Markus Brückner and Daniel Lederman 2015, How does income inequality affect economic growth?

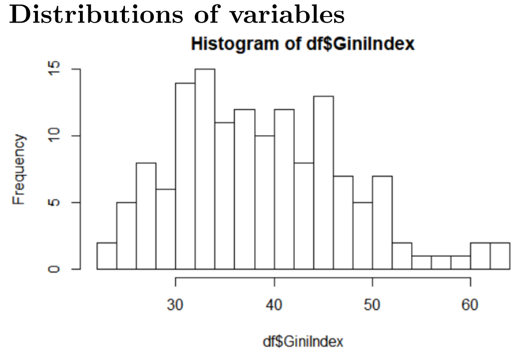
Mark J. Perry 2013, Income inequality can be explained by demographics, and because the demographics change, there's income mobility

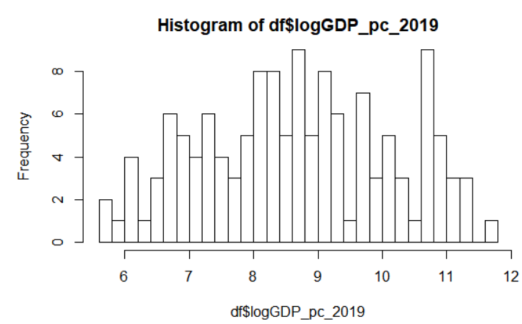
Appendix

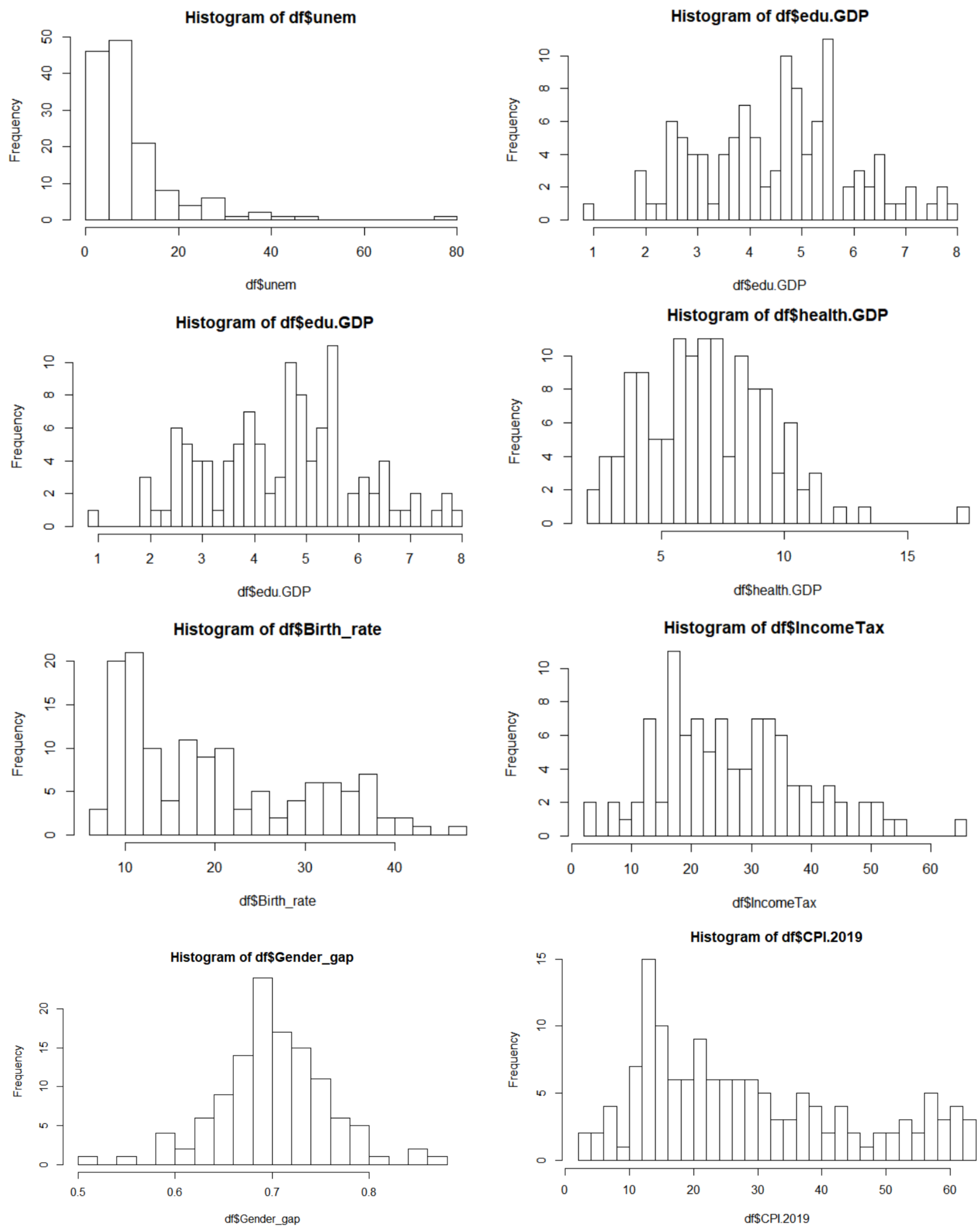
List with countries:

Albania, Algeria, Angola, Argentina, Armenia, Australia, Austria, Azerbaijan, Bangladesh, Belarus, Belgium, Benin, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Central African Republic, Chad, Chile, China, Colombia, Comoros, Costa Rica, Croatia, Cyprus, Czech Republic, Denmark, Djibouti, Dominican Republic, Ecuador, Egypt, El Salvador, Estonia, Ethiopia, Faroe Islands, Finland, France, Gabon, Georgia, Germany, Ghana, Greece, Greenland, Guatemala, Guinea, Guyana, Haiti, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kyrgyzstan, Laos, Latvia, Lesotho, Liberia, Lithuania, Luxembourg, Macau, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Mauritania, Mauritius, Mexico, Moldova, Mongolia, Montenegro, Morocco, Mozambique, Namibia, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russia, Rwanda, Sao Tome and Principe, Saudi Arabia, Senegal, Serbia, Sierra Leone, Singapore, Slovakia, Slovenia, South Africa, South Sudan, Spain, Sri Lanka, Swaziland, Sweden, Switzerland, Taiwan, Tajikistan, Tanzania, Thailand, Togo, Tunisia, Turkey, Turkmenistan, Uganda, Ukraine, United Kingdom, United States, Uruguay, Uzbekistan, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe

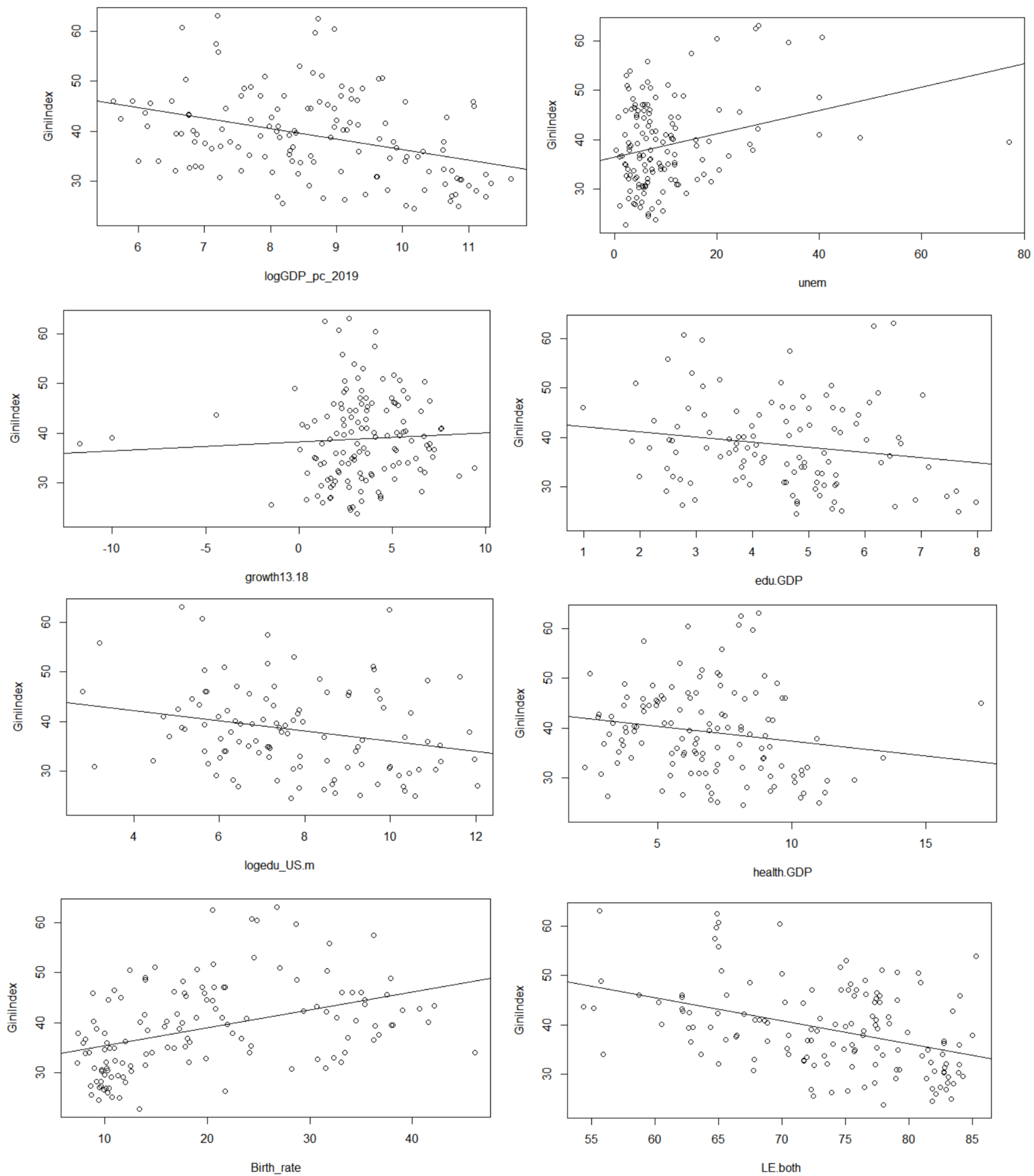
Distributions of variables

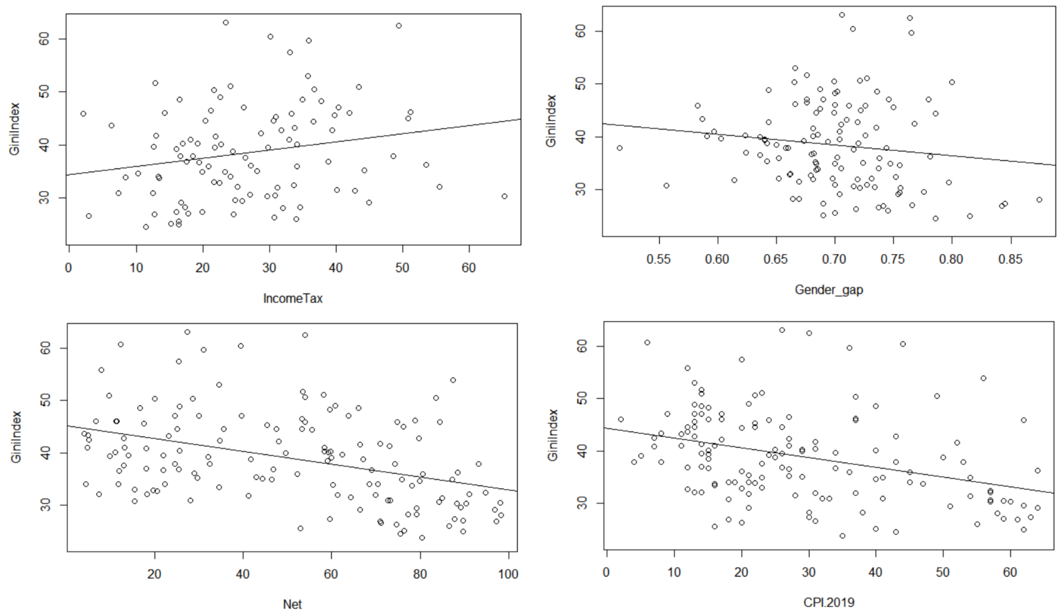






Scatter graphs between each of the independent variables and the dependent variable


GiniIndex
logGDP_pc_2019
unem
growth13.18
edu.GDP
logedu_US.m
health.GDP
Birth_rate
LE.both



Correlation matrix

	logGDP_pc_2019	unem	growth13.18	edu.GDP	edu_US.m	health.GDP
logGDP_pc_2019	1.00	-0.27	-0.14	0.34	0.42	0.42
unem	-0.27	1.00	-0.10	0.09	-0.13	-0.04
growth13.18	-0.14	-0.10	1.00	-0.26	-0.30	-0.45
edu.GDP	0.34	0.09	-0.26	1.00	0.15	0.43
edu_US.m	0.42	-0.13	-0.30	0.15	1.00	0.37
health.GDP	0.42	-0.04	-0.45	0.43	0.37	1.00
Birth_rate	-0.83	0.27	0.30	-0.27	-0.35	-0.40
LE.both	0.86	-0.32	-0.10	0.30	0.39	0.40
IncomeTax	0.15	0.03	0.14	0.02	0.24	0.16
Gender_gap	0.45	-0.15	0.04	0.46	0.18	0.45
Net	0.92	-0.26	-0.24	0.38	0.40	0.47
CPI.2019	0.81	-0.17	-0.08	0.48	0.40	0.46

	Birth_rate	LE.both	IncomeTax	Gender_gap	Net	CPI.2019
logGDP_pc_2019	-0.83	0.86	0.15	0.45	0.92	0.81
unem	0.27	-0.32	0.03	-0.15	-0.26	-0.17
growth13.18	0.30	-0.10	0.14	0.04	-0.24	-0.08
edu.GDP	-0.27	0.30	0.02	0.46	0.38	0.48
edu_US.m	-0.35	0.39	0.24	0.18	0.40	0.40
health.GDP	-0.40	0.40	0.16	0.45	0.47	0.46
Birth_rate	1.00	-0.88	0.06	-0.36	-0.85	-0.63
LE.both	-0.88	1.00	0.09	0.39	0.86	0.71
IncomeTax	0.06	0.09	1.00	0.12	0.06	0.18
Gender_gap	-0.36	0.39	0.12	1.00	0.45	0.51
Net	-0.85	0.86	0.06	0.45	1.00	0.78
CPI.2019	-0.63	0.71	0.18	0.51	0.78	1.00

Output 1: First Multiple Regression

```
Call:
lm(formula = GiniIndex ~ logGDP_pc_2019 + growth13.18 + unem +
    edu.GDP + log(edu_US.m) + health.GDP + Birth_rate + LE.both +
    IncomeTax + Gender_gap + Net + CPI.2019, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-14.228  -5.154  -1.684   3.969  16.397

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    52.13895   28.18851   1.850   0.0690 .
logGDP_pc_2019  1.50839    2.36286   0.638   0.5255
growth13.18    -0.82449    0.63974  -1.289   0.2021
unem            0.03486    0.10907   0.320   0.7503
edu.GDP         0.04970    0.91539   0.054   0.9569
log(edu_US.m)  -0.08729    0.68643  -0.127   0.8992
health.GDP      0.13971    0.59449   0.235   0.8150
Birth_rate      0.22340    0.21979   1.016   0.3132
LE.both        -0.29040    0.35287  -0.823   0.4136
IncomeTax       0.19953    0.08507   2.345   0.0221 *
Gender_gap     -8.04855   22.83926  -0.352   0.7257
Net            -0.06379    0.09710  -0.657   0.5136
CPI.2019       -0.12045    0.12123  -0.994   0.3242
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.725 on 64 degrees of freedom
  (67 observations deleted due to missingness)
Multiple R-squared:  0.342,     Adjusted R-squared:  0.2186
F-statistic: 2.772 on 12 and 64 DF,  p-value: 0.004261
```

Output 2: Second Multiple Regression

```
Call:
lm(formula = GiniIndex ~ logGDP_pc_2019 + log(edu_US.m) + Birth_rate +
    LE.both + Gender_gap + CPI.2019, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-13.818  -4.625  -1.600   4.253  21.351

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    44.57695   22.62877   1.970    0.052 .
logGDP_pc_2019  0.40228    1.56641   0.257    0.798
log(edu_US.m)   0.70168    0.53441   1.313    0.193
Birth_rate      0.17249    0.17046   1.012    0.314
LE.both        -0.25919    0.27054  -0.958    0.341
Gender_gap      5.10348   16.90319   0.302    0.763
CPI.2019       -0.11622    0.08999  -1.291    0.200
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.516 on 88 degrees of freedom
  (49 observations deleted due to missingness)
Multiple R-squared:  0.2049,    Adjusted R-squared:  0.1507
F-statistic: 3.779 on 6 and 88 DF,  p-value: 0.002144
```

Output 3: Third Multiple Regression

```
Call:
lm(formula = GiniIndex ~ logGDP_pc_2019 + growth13.18 + unem +
    edu.GDP + health.GDP + IncomeTax + CPI.2019 + Net, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-16.6105  -4.7131  -0.8574   4.1184  18.9548

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    48.69336   13.09790   3.718 0.000392 ***
logGDP_pc_2019 -0.79951    1.79137  -0.446 0.656692
growth13.18    -0.78838    0.54661  -1.442 0.153488
unem            0.07265    0.09304   0.781 0.437438
edu.GDP         0.30439    0.78082   0.390 0.697799
health.GDP     -0.12330    0.54305  -0.227 0.821014
IncomeTax       0.22825    0.07519   3.035 0.003327 **
CPI.2019       -0.08751    0.11218  -0.780 0.437837
Net            -0.09444    0.08732  -1.082 0.282981
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.597 on 73 degrees of freedom
  (62 observations deleted due to missingness)
Multiple R-squared:  0.3483,	Adjusted R-squared:  0.2769
F-statistic: 4.876 on 8 and 73 DF,  p-value: 7.644e-05
```

Output 4: Fourth Multiple Regression

```
Call:
lm(formula = GiniIndex ~ IncomeTax + growth13.18 + log(edu_US.m) +
    Birth_rate + LE.both + CPI.2019, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-15.121  -5.214  -1.937   4.077  16.546

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   55.90476   21.02496   2.659  0.00955 **
IncomeTax      0.21313    0.07600   2.804  0.00640 **
growth13.18   -0.86119    0.48558  -1.774  0.08015 .
log(edu_US.m)  0.07378    0.59947   0.123  0.90237
Birth_rate     0.27062    0.18367   1.473  0.14478
LE.both       -0.30925    0.27219  -1.136  0.25946
CPI.2019      -0.09033    0.07353  -1.228  0.22308
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.291 on 76 degrees of freedom
  (61 observations deleted due to missingness)
Multiple R-squared:  0.3825,	Adjusted R-squared:  0.3338
F-statistic: 7.847 on 6 and 76 DF,  p-value: 1.369e-06
```

Output 5 for Restricted Model

```
Call:
lm(formula = GiniIndex ~ IncomeTax, data = omitdf)

Residuals:
    Min      1Q  Median      3Q     Max 
-13.163  -7.482  -1.434   5.051  26.412 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 33.05497    2.45344  13.473   <2e-16 ***
IncomeTax    0.15966    0.08312   1.921   0.0585 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.589 on 75 degrees of freedom
Multiple R-squared:  0.04689,	Adjusted R-squared:  0.03418 
F-statistic:  3.69 on 1 and 75 DF,  p-value: 0.05854
```