Final Project - Income Inequality and Gross Capital Formation

Shahana Ayobi

2022-12-19

1. Introduction

There has been a lot of discussion recently about income disparity, particularly since the Great Recession and now following COVID-19. Many people believe that the wealthy are getting richer while the vast majority is falling behind. Determinants such as GDP per capita, education, and unemployment rate have been shown to have a correlation with a country's income disparity. In this project, I will examine the potential relationship between Gross Capital Formation, also known as investment, and income inequality as measured by the Gini Index.

2. Data

2.1. Source

My main source of data for this project was collected from World Development Indicators which accumulates its indicators based on data from nationally representative household surveys, which are generally conducted every three to five years. The data includes 217 observations for year 2018 where each observation is country, and the following variables have been considered for the analysis: 1. Dependent Variable: Gini Index which is a statistical measure of dispersion meant to depict the level of income or wealth inequality within a country or a social group. 2. Explanatory Variable: Gross capital formation, formerly known as gross domestic investment, is made up of expenditures for new fixed assets for the economy as well as net changes in inventory levels. 3. Control Variables: - GDP per capita (constant US$) which is a country's Gross Domestic Product divided by its population. - Unemployment rate that measures the percentage of workers in the labor force who do not currently have a job but are actively seeking one. Table 1 shows the descriptive statistics for dependent, explanatory, and control variables. It can be inferred that all of the variables have higher means than medians, meaning the distribution of the variables are skewed to right. Therefore, for Variables like `GDP per capita`, I took the logarithm of it to be able to interpret the results in percentages and transform the distribution closer to normal. The rest of my variables are already in percentages, Taking their logarithm would worsen the estimation of the models.

Table 1: Descriptive statistics

	Mean	Median	SD	Min	Max	P05	P95
Gini Index	38.71	38.60	7.60	24.60	63.00	27.35	51.24
Gross Capital Formation (% of GDP)	24.83	23.18	8.24	2.79	51.78	13.64	40.32
GDP per Capita	16 561.12	6381.60	23 704.46	281.93	171 223.98	632.25	59 718.36
Unemployment Rate (% of Labor Force)	7.48	5.59	5.77	0.11	26.91	1.17	19.75

2.2. Data Cleaning

The variables' original names from the World Bank Indicators were lengthier, so I modified them to make them easier to deal with when adding them into the model. Also, the variables type has been classified as characters initially; thus, I changed their type to numeric. Then, the values were rounded with two decimal places.

3. Non-Parametric Regression and Variable Transformations

Before moving on to the analysis, I plotted a Loess regression with a scatter plot of dependent versus independent variables. This is because, I wanted to display the joint distribution of two variables in the data and see how much the mean of the Gini Index varies among observations with different values of the explanatory and control variables. This aids in putting the regression's slope into perspective. Figure 1 shows the loess regression line for Gini Index and Investment, and it is apparent that the pattern of line is not linear. At lower than 22% investments, countries tend to have a decreasing trend in income inequality; however, inequality tend to increase at higher levels of investment. To better capture this relationship, I introduced two cutoffs points or splines for the explanatory variable in 22% and 32%.

Figure 1: Loess Regression: Gross Capital Formation and Gini Index

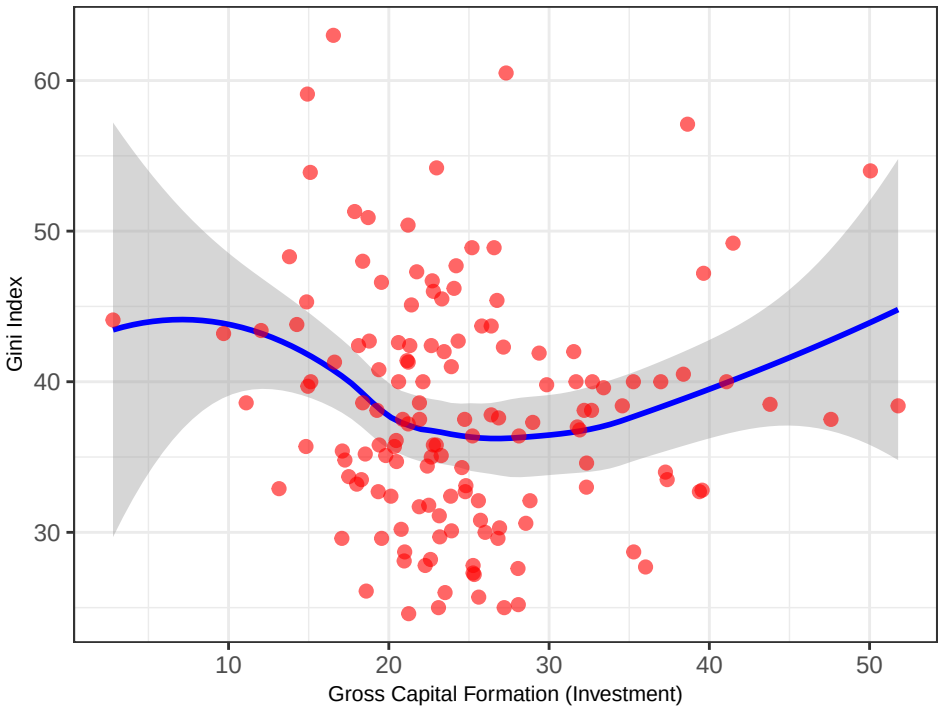

Figure 2 in the appendix shows the pattern of association between Logarithm of `GDP per Capita` and Gini index,and Figure 3 shows the relationship between Unemployment rate and Gini Index.The regression line has an increasing trend after 5% unemployment rate, therefore, I added a spline at that point.

4. Hypothesis

The main hypothesis of this research is that inequality tends to be higher at higher levels of Gross Capital Formation (% of GDP). This is because in terms of statistics, GCF calculates the value of purchases of new or

used fixed assets by businesses, governments, and "pure" households (apart from those with unincorporated businesses), minus disposals of fixed assets. It provides information regarding the proportion of newly added economic value that is invested as opposed to consumed. GCF is not a measure of total investment because all types of financial assets are excluded, along with stocks of inventories and other operating expenses, leaving just the value of net additions to fixed assets to be calculated. The sale and acquisition of land is the most significant exclusion from the GCF. Therefore, excluding those important factors, will make the investment (% GDP) relatively lower for developed countries as compared to some developing countries where some of the above mentioned expenditures are larger part of their expenses, so it would also weigh higher as their percentage of GDP. For instance, construction of roads and other infrastructures would not weigh as high in percentage of GDP of the United Stated compared to Timur-Leste, this is why the investment (% GDP) is higher for Timur-Leste (35.28%) compared to U.S. (21.14%). When comparing GDP per capita and income inequality, the Kuznet Curve developed by Simon Kuznet (1955) can be referred here. The main conclusion is that income inequality rises as a country is developing, particularly as its population shifts from rural to urban, and falls inversely as modern structures take hold and the GDP per capita increases. Figure 5 in the appendix depicts this relationship. Increased unemployment is positively correlated with income inequality, meaning that when a larger percentage of labor force in a country is unemployed, the income inequality increases in that country (Appendix, Figure 4).

5. Models

To confirm my hypothesis, I have first ran a simple linear regression with the GCF as explanatory and Gini index as dependent variables. I have then added the control variables one by one in the following models.

5.1. Model 1

$$GiniIndex = \alpha + \beta 1 * GCF(GCF < 22) + (\alpha 2 + \beta 2 * GCF)[22 <= GCF <= 32]$$
$$+(\alpha 3 + \beta 3 * GCF)[GCF > 32]$$

In the Table 2, the intercept suggests that, income inequality in a country is 50.83% when all other variables are zero. We can also conclude that that at a level of GCF lower than 22%, income inequality tends to decrease by 0.64% and it is statistically significant at 1 and 5 percent significant levels. This is because, most of the developed countries such as the ones in Europe have a GCF values around this specific cuttoff and they are the ones with lower values for gini index, that is why we have negative correlation among them.

5.2. Model 2

$$GiniIndex = \alpha + \beta 1 * GCF(< 22) + (\alpha 2 + \beta 2 * GCF)[22 <= GCF <= 32$$
$$+(\alpha 3 + \beta 3 * GCF)(> 32) + \beta 4 * log(GDP)$$

In this model, I added a control variable, GDP per capita. When all other variables are zero, income inequality tends to be 62.36%. At lower level of GCF lower than 22%, income inequality tends to decrease by 0.45% as GCF increases by 1% while at higher levels of GCF (more than 32%) inequality tends to increase by 0.38% and both are statistically significant 5 percent significant level. This in turn confirm our hypothesis that higher GCF at higher levels of GCF tends to increase income inequality.

Table 2: Simple Regressions Result

	Model 1	Model 2	Model 3
Constant	50.8286***	62.3659***	57.2234***
	(4.1767)	(5.1946)	(5.2033)
lspline(GCF, cutoff_GCF)1	−0.6382**	−0.4524*	−0.1723
	(0.2100)	(0.2048)	(0.2445)
Gross Capital Formation Spline 22-32%	−0.0426	−0.1493	−0.2684
	(0.1751)	(0.1713)	(0.1926)
Gross Capital Formation Spline 32%<	0.4252	0.3763*	0.5039*
	(0.2226)	(0.1876)	(0.1981)
Log(GDP)		−1.7069***	−1.6227***
		(0.3894)	(0.4195)
Unemployment Rate Spline <5%			−0.6901
			(0.4725)
Unemployment Rate Spline 5%<=			0.5743**
			(0.1849)
Num.Obs.	142	142	136
R2	0.077	0.162	0.266
R2 Adj.	0.057	0.138	0.232
RMSE	7.42	7.07	6.74

* p < 0.05, ** p < 0.01, *** p < 0.001

5.3. Model 3

$$GiniIndex = \alpha + \beta 1 * GCF(< 22) + (\alpha 2 + \beta 2 * GCF)[22 <= GCF <= 32]$$

$$+(\alpha 3 + \beta 3 * GCF)(> 32) + \beta 4 * log(GDP) + \beta 5 * UnemploymentRate(< 5) + \beta 6 * UnemploymentRate(>= 5)$$

While controlling for Logarithm of GDP per capita and Unemployment rate, GCF is still statistically significant at higher levels meaning that a 1% increase in Gross Capital Formation (% of GDP) where that value of GCF is more than 32%, inequality also tends to be higher by 0.5% and it is statistically significant at 5%. Confirmed by the Kuznet's theory, the model also shows that if log(GDP) increases by 1%, inequality decreases by 1.62%, ceteris paribus. Inequality also tends to be higher by 0.57% at higher than 5% level of unemployment rate when unemployment rate increases by 1%, ceteris paribus. In terms of R-squared, Model 3 definitely is a better fit since more than 26% of variation in gini index is explained by the variation in explanatory and control variables.Therefore, considering the fact that our alternative hypothesis is confirmed and we have higher R-squared in Model 3, it better represents our analysis.

6.Conclusion

To sum up, the findings of the models have met the expected results. At higher levels of Gross Capital Formation (Investment as % of GDP), the income inequality is indeed higher. However, this correlation does not mean causation, the purpose of this research was to find the patterns of association between the two variables which has been shown in the analysis. Also, control variables such as logarithm of GDP per capita and unemployment rate could indeed absorb some of the association with gini index, but more control variables are needed to be added further confirm a stronger association.

Appendix

Figure 2: Loess Regression: Log(GDP per Capita) and Gini Index

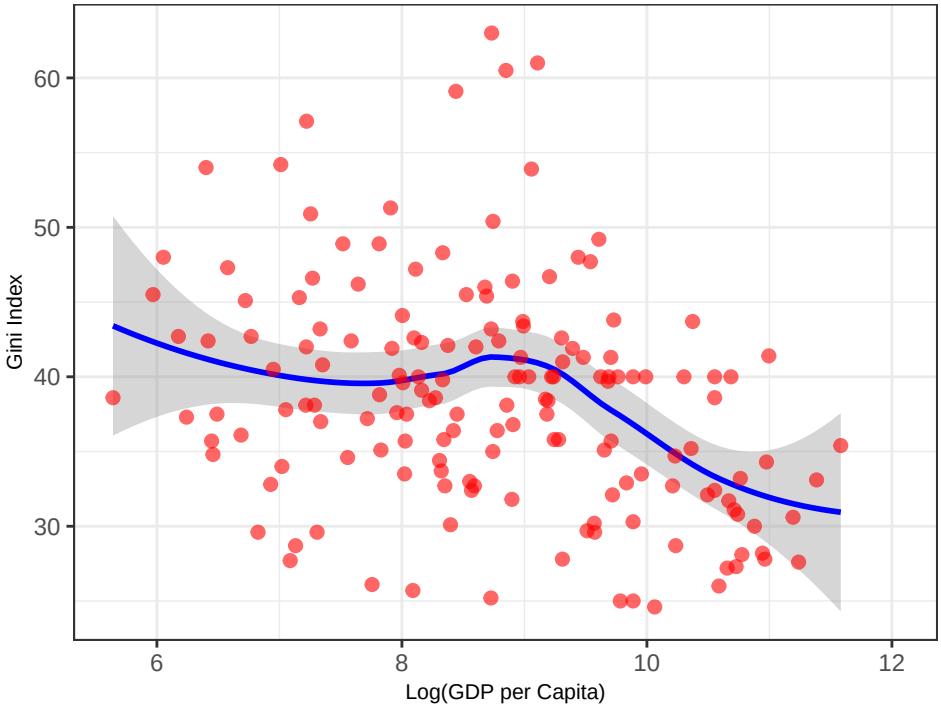

Figure 3: Loess Regression: Unemployement Rate and Gini Index

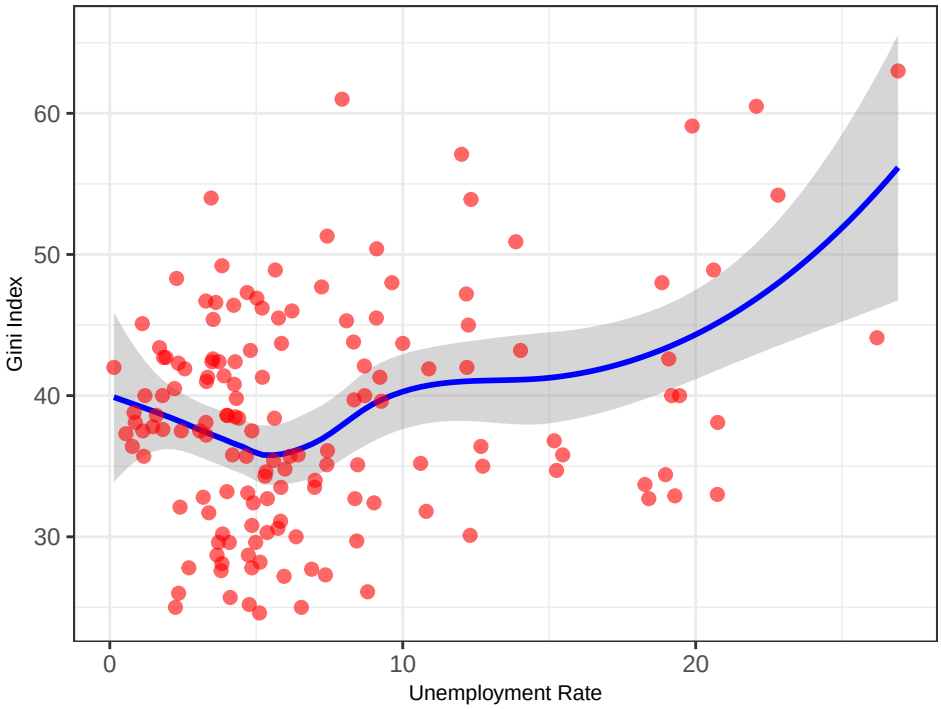

Figure 4: Gross Capital Formation percentage by country