**TEAM MEMBERS: Devansh Singh (200319)**

**Jaiki Sagar (200464)**

**INTRODUCTION**

**BACKGROUND INFORMATION:**

**Income inequality is the measure of how unequally the income is distributed among the people. As the income distribution among the population gets more skewed, the value of income inequality increases.**The initial studies on income distribution were on how the distribution changes with economic growth.

However, many studies have shown that increased crime rates, health problems, lower value of goods and decreased level of satisfaction and happiness among people can be traced back to high levels of income inequality. These negative effects of high-income inequality shifted the focus of the latter studies to finding various ways to reduce income inequality. The above factors, combined with the availability of new datasets and the dynamic conditions across the globe resulted in topics like income distributions, income inequality, skewed distribution, and the ways to prevent the unequal distribution of income being researched with renewed interest.

**DATA**

Now in order to find the relationship between government expenditure on education, education and income inequality, we need to find suitable variables which can accurately represent the three abovementioned entities. The secondary data for the first two entities have been taken from the World Bank at this link: [Education | Data (worldbank.org)](https://data.worldbank.org/topic/4).

Data for education is taken care of for the time being. Now coming to data for income inequality, the variable I chose was Ginni Index or Ginni coefficient. Ginni index is a number ranging from 0 to 100 with 0 showing perfect income equality and 100 showing perfect income inequality in a country. The secondary data for Gini index was taken from the World Bank at the following link: [Gini index (World Bank estimate) | Data](https://data.worldbank.org/indicator/SI.POV.GINI).

Additionally, I also took the GDP per capita data from the World Bank from the following link: [GDP per capita (current US$) | Data (worldbank.org)](https://data.worldbank.org/indicator/NY.GDP.PCAP.CD) to see how the regression results change with and without this variable.

**METHODOLOGY**

In this study we are working with panel data which is time series cross-sectional i.e it contains many instances of cross-sectional data over a period of time. The data consists of more than 250 countries with each country having 9 columns to describe its condition over the time period 2000 to 2019.

One of the regression models on which we can fit panel data is pool regression model.

Model 1 Summary

Table

Description automatically generated with medium confidence

Model 2 plot and summary

Chart, scatter chart

Description automatically generated

Table

Description automatically generated with medium confidence