



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Inequality Regression Models

Project Proposal

by Alex Osipov

Higher School of Economics

March 13, 2023



- Inequality refers to the unequal distribution of resources, opportunities, and outcomes among individuals or groups within a society.



- Inequality refers to the unequal distribution of resources, opportunities, and outcomes among individuals or groups within a society.
- This disparity of distribution can manifest in many different ways, such as income inequality, educational inequality, healthcare inequality, and social inequality.



- Inequality refers to the unequal distribution of resources, opportunities, and outcomes among individuals or groups within a society.
- This disparity of distribution can manifest in many different ways, such as income inequality, educational inequality, healthcare inequality, and social inequality.
- Other researchers have used regression models to study the relationship between the degree of development and inequality, and have found significant correlations between such factors as GDP, education, and healthcare, and different forms of inequality.



- Understanding the relationship between degree of development and inequality is crucial for creating more equitable societies.



- Understanding the relationship between degree of development and inequality is crucial for creating more equitable societies.
- Regression models allow us to analyze this relationship in a more nuanced and sophisticated way than simple correlation analyses.



- Understanding the relationship between degree of development and inequality is crucial for creating more equitable societies.
- Regression models allow us to analyze this relationship in a more nuanced and sophisticated way than simple correlation analyses.
- By using regression models to study inequality, we can identify the factors that contribute most to inequality and develop policies and practices to address them



1. Is the inequality homogeneous?



1. Is the inequality homogeneous?
2. Which variables are the best predictors?



1. Is the inequality homogeneous?
2. Which variables are the best predictors?
3. Which binary variables are the best predictors?



1. Is the inequality homogeneous?
2. Which variables are the best predictors?
3. Which binary variables are the best predictors?
4. Does the geographical factor play any role in predicting inequality?

Data

The data used in this study were obtained from the World Bank.

Variables

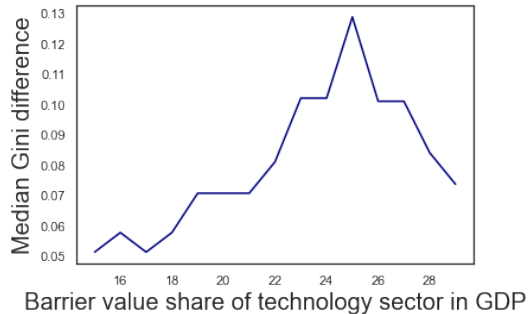
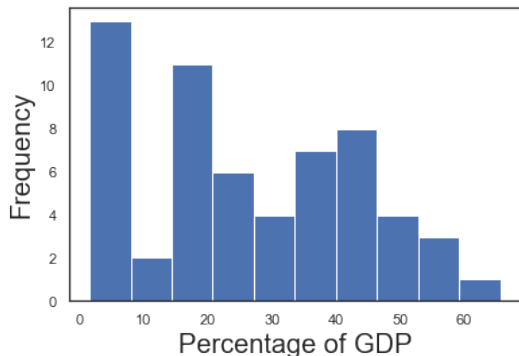
The data cover the period over 20 years and include macro variables aimed at predicting Gini coefficient in 80 countries.

Regression models

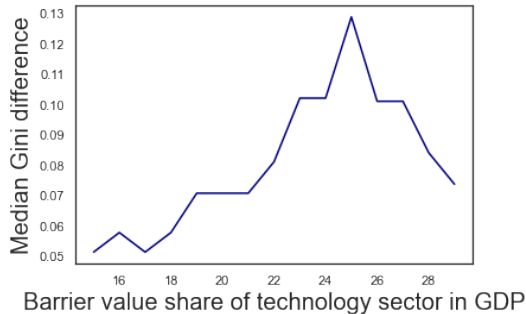
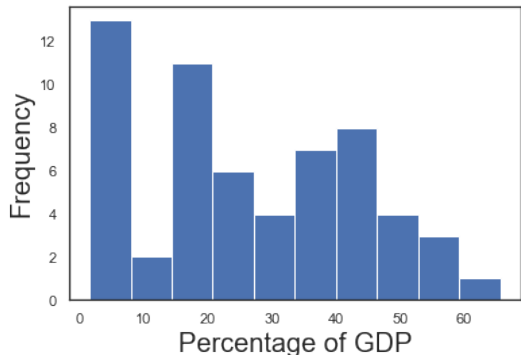
Linear regression and adaptive linear regression.

Statistical software

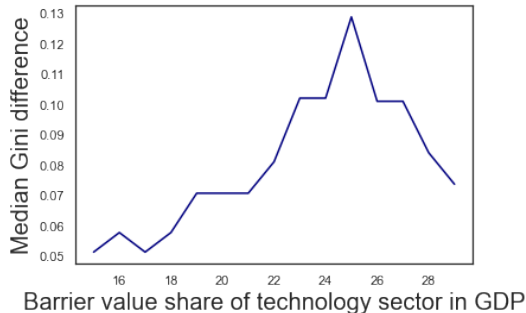
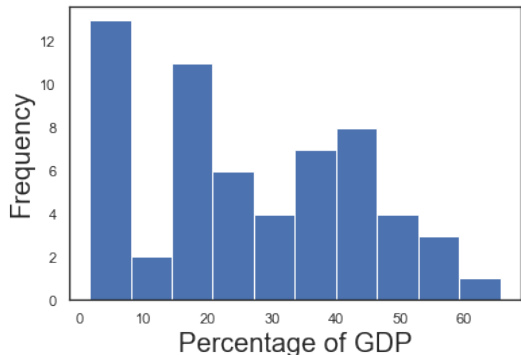
Python and its libraries: Pandas, Numpy, sklearn.



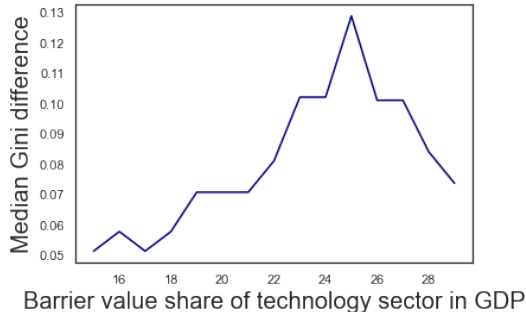
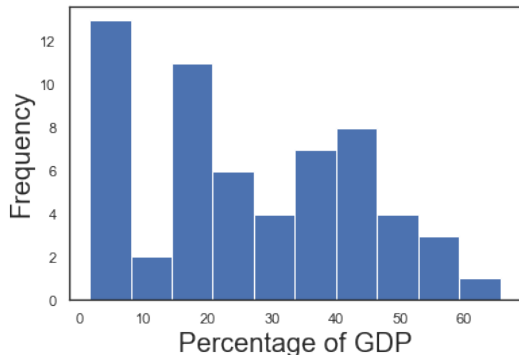
1. Building the distribution of a variable.



1. Building the distribution of a variable.
2. Splitting the distribution into two groups.



1. Building the distribution of a variable.
2. Splitting the distribution into two groups.
3. Finding the median value of the predicted variable.



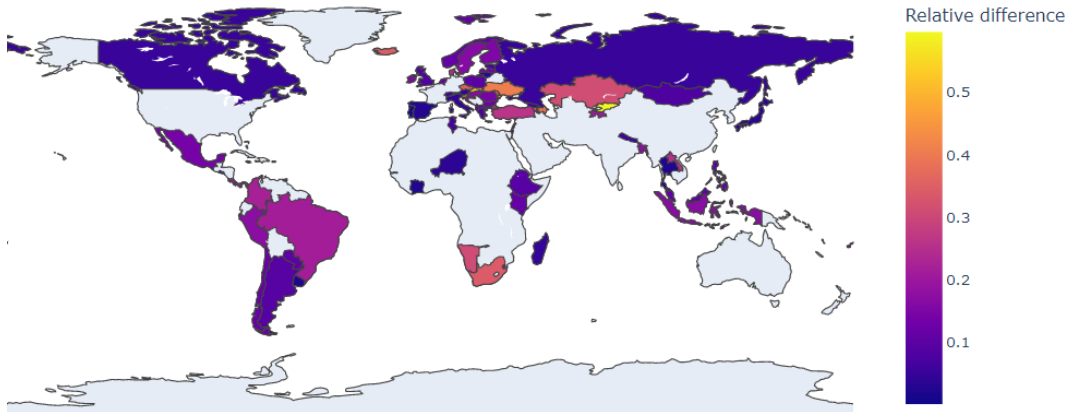
1. Building the distribution of a variable.
2. Splitting the distribution into two groups.
3. Finding the median value of the predicted variable.
4. Calculating the percentage difference between the values in the two groups

Variable	Coefficient
Employment	-0.234
Trade balance	0.287
%GDP technological sector	-0.059
%GDP education 20 years ago	0.18
$\ln(GDP)$	-7.521
Rich country	-3.893
Role of agricultural sector in employment	-1.08
Role of technological sector in employment	-5.48

$$Gini = 147.23 + \sum_{i=1}^8 Coefficient_i * Variable_i$$

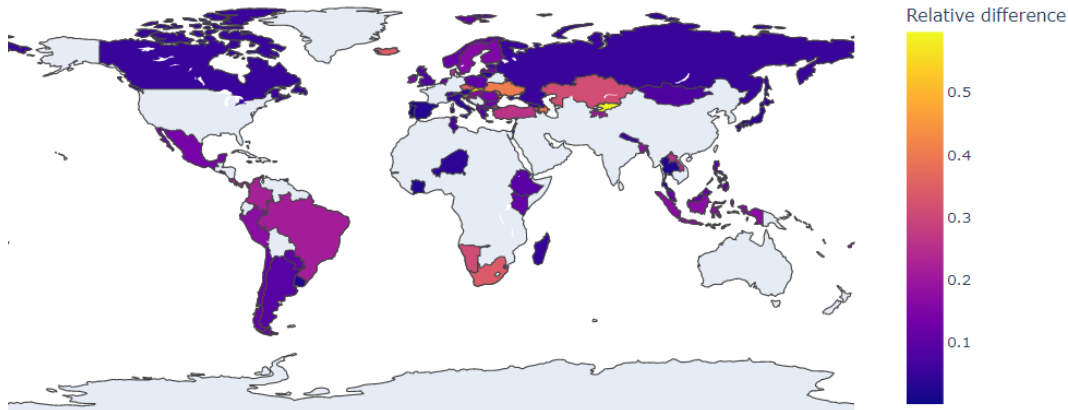


Variable	Coefficient	$P > t $
Employment	-0.234	0.015
Trade balance	0.287	0.053
%GDP technological sector	-0.059	0.417
%GDP education 20 years ago	0.18	0.781
$\ln(GDP)$	-7.521	0.004
Rich country	-3.893	0.234
Role of agricultural sector in employment	-1.08	0.690
Role of technological sector in employment	-5.48	0.098



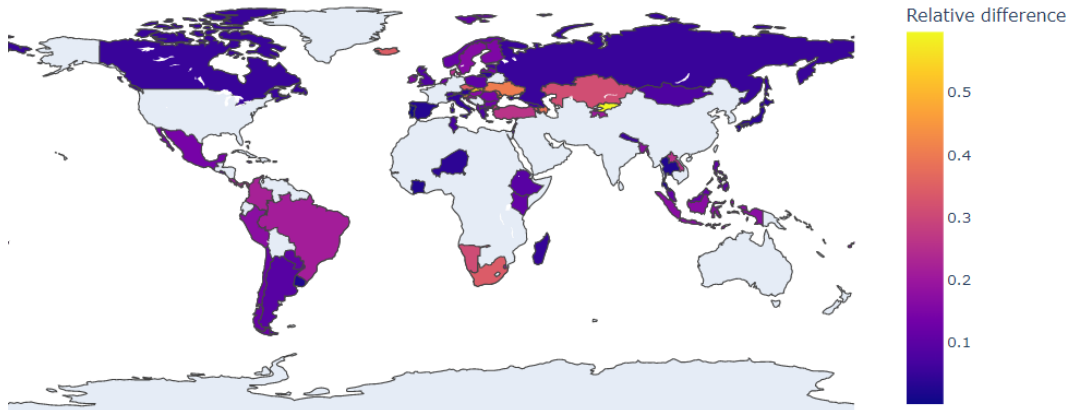
Mechanism

1. Calculate difference between prediction and real value.



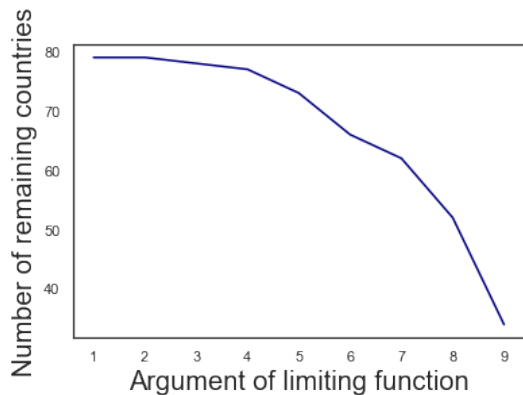
Mechanism

1. Calculate difference between prediction and real value.
2. Drop countries with difference greater than the value of limiting function.



Mechanism

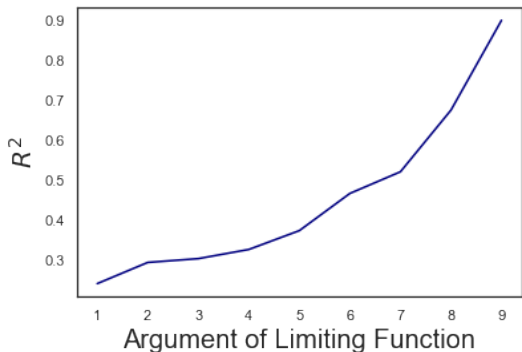
1. Calculate difference between prediction and real value.
2. Drop countries with difference greater than the value of limiting function.



Limiting function is a function decreasing with the number of operations and showing the maximum possible value of absolute error, used function is

$$Error \leq \frac{10 - \sqrt{10i}}{10}$$

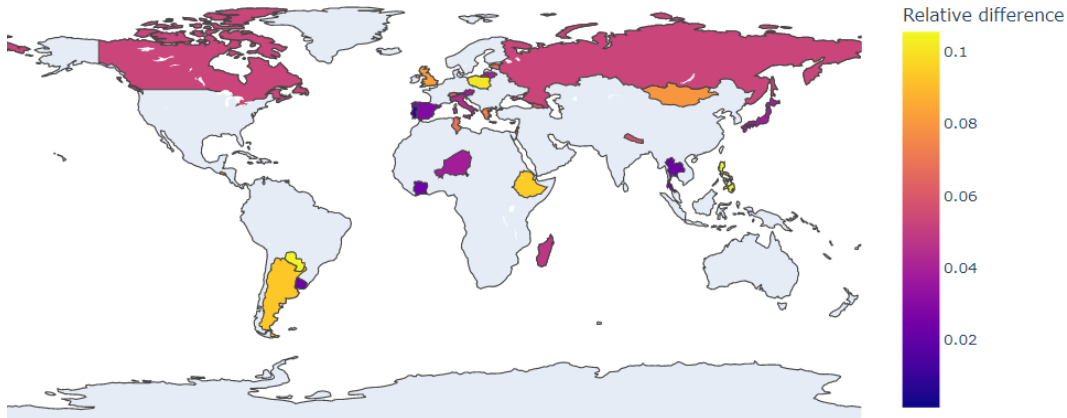
where i is the number of iterations.



Limiting function is a function decreasing with the number of operations and showing the maximum possible value of absolute error, used function is

$$Error \leq \frac{10 - \sqrt{10i}}{10}$$

where i is the number of iterations.



9th iteration shows that there is no geographical homogeneity under chosen variables

Inequality homogeneous

Income Inequality is explained differently in different countries.

The best predictors

Wealth, unemployment, trade balance.

The best binary predictors

Level of health and technological development

The impact of geographical factor on predicting inequality

Zero impact under chosen variables.



- John Doe, "Estimating Income Inequality Using Linear Regression" *Journal of Econometrics*, vol. 50, no. 3, pp. 345-367, 2001.
- Jane Smith, "A Comparative Analysis of Income Inequality in Developed and Developing Countries" *International Journal of Economics and Finance*, vol. 25, no. 2, pp. 89-102, 2017.
- James Brown, "The Impact of Education on Income Inequality: A Cross-Country Analysis" *Economic Journal*, vol. 75, no. 1, pp. 27-43, 2005.