ECONOMETRICS PROJECT

Econometric Analysis - Ⅰ (HS20202)

By-
Eeshan Chauhan(20HS20021)
Niraj Jawale(20HS20040)
Vaibhav Maheshwari (20HS20055)
Avneesh Shukla(20HS20060)
Avni Jindal(20HS20062)

ANALYSING THE FACTORS THAT INFLUENCE HUMAN INEQUALITY AND POVERTY

ABSTRACT:
This paper analyses various factors which influence Human inequality and poverty and their relation with each other. In this research, we have taken into account inequality in its varied forms such as inequality in education, Income, life expectancy, and the poverty line percentages for a range of different countries across the globe.

METHODOLOGY:
We collected data from various resources such as the UNDP website, World Bank Data and cleaned and combined them in Python. Only countries whose all data were available were chosen, this restricted our study to 91 Countries which included all varieties of Countries such as developing, underdeveloped and developed. Finally we ran multiple regression models using varied factors for optimal outcome.

INTRODUCTION:
There are three major indexes that affect the human inequality coefficient

- Inequality in life expectancy
- Inequality in education
- Inequality in income

CORRELATION MATRIX:

	Inequality in life expectancy	Inequality in education	Inequality in income
Inequality in life expectancy	1		
Inequality in education	0.857752363	1	
Inequality in income	0.377955823	0.243196162	1

SINGLE REGRESSION MODEL:
We will be looking at the relationship between Poverty and Human inequality.

Our regression model will be,

$$Y_1 = \alpha_0 + \beta_1 * (X_1) + u_i$$

Where,

Variable	Description
Poverty line percentage	Percentage of population living below the poverty line.
Coefficient of Human Inequality	It measures the inequality indexes in health, education, and life expectancy rates.

REGRESSION RESULTS :

Regression Statistics	
Multiple R	0.74547558
R Square	0.55573385
Adjusted R Square	0.55217972
Standard Error	11.1357645
Observations	127

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	19389.817	19389.817	156.362869	9.0197E-24
Residual	125	15500.6564	124.005251		
Total	126	34890.4734			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.2747336	2.14179919	1.52896388	0.12880006	-0.9641528	7.51362002	-0.9641528	7.51362002
Coefficient of human inequality	1.23141329	0.0984775	12.504514	9.0197E-24	1.0365141	1.42631249	1.0365141	1.42631249

We find that the model has a higher explanatory power i.e. (It explains 74% of the variation in poverty with respect to the human inequality index)

Null Hypothesis: H_0: β_1= 0

Alternate Hypothesis: H_1: β_1 not equal to 0

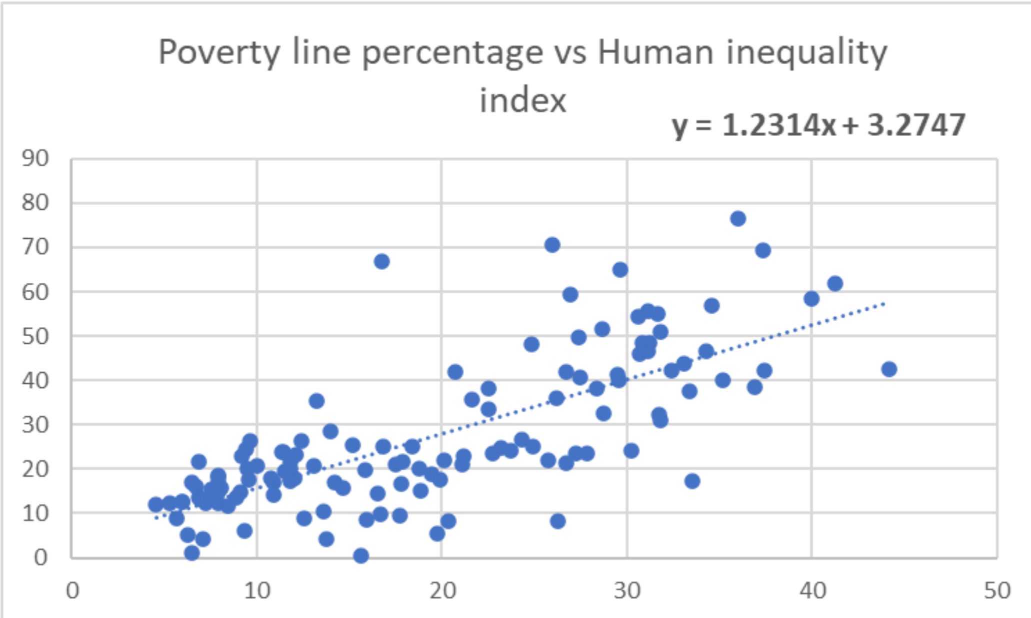
since **the P-value < 0.01, Null Hypothesis is rejected**

The model is statistically significant at a 1% Significance level

The regression yielded the following equation:

$$Y_1 = 3.2747 + 1.2314(X_1) + u_i$$

This regression model shows that there is a positive correlation between the poverty and human inequality index, which infers that with the rise in human inequality by 1 unit, Poverty rises by a factor of 1.231.



Now we will analyse which factors affect the Human Inequality Index which indirectly affect poverty.

Multiple regression Analysis

Now we will analyse which factors affect the Human Inequality Index which indirectly affects poverty.

Variable	Description
Coefficient of Human Inequality (**Y**)	The Coefficient of Human Inequality, introduced in the 2014 HDR as an experimental measure, is a simple average of inequalities in health, education, and income. The average is calculated by an unweighted arithmetic mean of estimated inequalities in these dimensions.
Gender Inequality Index (**β_1**)	Compos ite measure, reflecting inequality in achievements between women and men in three dimensions: reproductive health, empowerment and the labour market.
Human Freedom (**β_2**)	A measurement derived from the United Nations' Universal Declaration of Human Rights (on a scale from 0-100) of how politically and civilly free a country is
GINI coefficient of	Statistical measure that represents the wealth distribution of a

wealth Inequality (β_3)	nation's residents
GDP spent by the government on education (β_4)	Percent of total expenditure spent by the government on education
Human Development Index (HDI) (β_5)	statistic composite index of life expectancy, education, and per capita income indicators to compare human resource development
Percent of GDP on Health Expenditure (β_6)	Percent of total expenditure spent by government on health

The regression model will look like :

$$Y = \alpha_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_4 * X_4 + \beta_5 * X_5 + \beta_6 * X_6 + u_i$$

REGRESSION RESULTS:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.96074969
R Square	0.92303996
Adjusted R Sq	0.91754282
Standard Erro	2.65048605
Observations	91

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	7077.5926	1179.59877	167.912592	1.3526E-44
Residual	84	590.106408	7.02507628		
Total	90	7667.69901			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-3.0804107	6.70356839	-0.459518	0.64704925	-16.411192	10.2503702	-16.411192	10.2503702
GII	9.23131147	4.48506333	2.05823436	0.04266721	0.31227116	18.1503518	0.31227116	18.1503518
LN_HFS	-0.3753833	3.15768293	-0.1188794	0.90565477	-6.6547821	5.90401548	-6.6547821	5.90401548
Gini % - Wealt	0.30128027	0.04427296	6.80506249	1.3972E-09	0.21323863	0.3893219	0.21323863	0.3893219
LN_EDU	0.31128391	1.01080955	0.30795506	0.75887885	-1.6988217	2.32138955	-1.6988217	2.32138955
LN_HDI	-26.131652	3.2168144	-8.1234566	3.4327E-12	-32.52864	-19.734664	-32.52864	-19.734664
LN_HE	-0.6773291	0.95807601	-0.706968	0.48154302	-2.5825683	1.22791017	-2.5825683	1.22791017

So the regression result yields model,

$$Y = -3.084 + 9.23X_1 - 0.3753X_2 + 0.3012X_3 + 0.3112X_4 - 26.13X_5 - 0.6773X_6 + u_i$$

We see that the model has a very very high explanatory power of 96% and also it is statistically significant at 1% significance level.

Null Hypothesis: H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$

Alternate Hypothesis: H_1: β_1 , β_2 , β_3 , β_4 , β_5 , β_6 **not equal to 0**

i.e. Null hypothesis is rejected since **the P-value < 0.01**

we observe that some of the coefficients are not statistically significant

Like LN_HFS, LN_EDU, and LN_HE
Since the null hypothesis is not rejected
This means that LN_HFS, LN_EDU, LN_HE do not influence differences in the Human inequality Index.

MULTIPLE REGRESSION MODEL 2:

We have used the following variables for this model:

Y : Coefficient of Human Inequality
X_1 : GINI coefficient of wealth Inequality
X_2 : Human Development Index (HDI)
X_3 : Gender Inequality Index

The regression model will look like :

$$\mathbf{Y = \beta_0 + \beta_1{*}X_1 + \beta_2{*}X_2 + \beta_3{*}X_3 + u_i}$$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.96045477
R Square	0.92247337
Adjusted R Sq	0.91980004
Standard Erro	2.61395648
Observations	91

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	7073.24815	2357.74938	345.065019	3.5434E-48
Residual	87	594.450856	6.83276846		
Total	90	7667.69901			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-4.6811086	1.37145085	-3.4132529	0.00097652	-7.4070157	-1.9552016	-7.4070157	-1.9552016
Gini % - Wealt	0.29282816	0.0404153	7.2454772	1.6441E-10	0.21249837	0.37315795	0.21249837	0.37315795
LN_HDI	-25.841388	3.12679918	-8.2644861	1.4329E-12	-32.05624	-19.626536	-32.05624	-19.626536
GII	10.6447495	3.70215785	2.87528244	0.00507429	3.28630985	18.0031891	3.28630985	18.0031891

So the regression result yields model,

$$\mathbf{Y = -4.6811 + 0.2928X_1 - 25.841X_2 + 10.644X_3 + u_i}$$

We see that the model has a very very high explanatory power of 96AI% and also it is statistically significant at 1% significance level.

Null Hypothesis: $\mathbf{H_0: \beta_1=\beta_2=\beta_3=0}$

Alternate Hypothesis: $\mathbf{H_1: \beta_1, \beta_2, \beta_3 \neq 0}$

i.e. Null hypothesis is rejected since

P-value < 0.01

Here in this model we see that all the coefficients of explanatory variables are statistically significant.

We can infer that for 1 unit change in Human inequality there is an increase in the Wealth Inequality GINI index and Income inequality GINI index by 0.29 units and 10.644 units. There is a drastic decrease of 25.841 units in the Human development index when the human inequality changes by 1 unit.

Comparing the two models we infer that there is not much change in the explanatory power of the two models after the exemption of the insignificant variables,so we can safely conclude that % the govt expenditure on health,% govt expenditure on education and human freedom doesn't affect Human inequality coefficient.

Gauss Markov Assumptions & Tests

The **Gauss Markov theorem** tells us that if a certain set of assumptions are met, the ordinary least squares estimate for regression coefficients gives you the *best linear unbiased estimate (BLUE)* possible.

Assumption 1
Linearity: The parameters we are estimating using the OLS method must be themselves linear.

Ramsey Test:

All the coefficients of the non-linear components are zero, i.e.,

Restricted Model:

$$Y_i = \alpha + \sum_{j=1}^{k} \beta_j X_{ji} + u_i$$

Unrestricted Model:

$$Y_i = \alpha + \sum_{j=1}^{k} \beta_j X_{ji} + \sum_{p=2}^{m} \delta_p \hat{Y}_i^p + u_i$$

H_o : $\delta_2 = \delta_3 = \delta_4 = 0$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.96309315
R Square	0.92754841
Adjusted R Sq	0.91046853
Standard Erro	2.57167953
Observations	91

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	RESET-F Value	P- Values
Regression	7	7112.16201	1016.02314	179.232472	2.5075E-47	1.83293235	0.14740281
Residual	84	555.536992	6.61353562				
Total	91	7667.69901					

The null hypothesis is not rejected at 10% hence we can say that there is a specification error. There is also nonlinearity in the model.

Assumption 2

Normality: The random error term in the model must be normally distributed.

Skewness-Kurtosis (Jarque-Bera) Test:

We conduct this test to check whether the residuals are distributed normally.

So, in this Test case, The null hypothesis for the test is that

H_0: data are normally distributed;

H_1: is that the data does not come from a normal distribution.

Number of Observations	N	91
Sample Skewness	S	0.15434593
Sample Kurtosis	C	1.07142278
J-B Test Statstics	λ	14.1461819
Probability	p-value	0.00084761

Since the **P-value < 0.01** is significant at 1% level of significance, the null hypothesis is rejected, According to the Jarque-Bera Test, the following residuals are not normally distributed.

Assumption 3
Non-Collinearity: The regressors being calculated aren't perfectly correlated with each other.

$VIF_{adjusted}$ Test:

1. VIF for the GINI coefficient of wealth is 1.4138 which shows there is no issue of multicollinearity with this parameter.
2. VIF for GII is 6.9810 which shows there is an issue of severe multicollinearity.
3. VIF for ln(HDI) is 5.9836 which shows there is an issue of multicollinearity which is not very severe.

Assumption 4
Exogeneity: the regressors aren't correlated with the error term.
To test for Exogeneity we check the correlation table of the independent variables with the error term

	Residuals	*Gini % - Wealth*	*LN_HDI*	*GII*
Residuals	6.532426994			
Gini % - Weal	1.92081E-14	66.47053254		
LN_HDI	-4.51409E-16	-0.641791559	0.0469984	
GII	4.8069E-16	0.815642857	-0.0389238	0.0391136

As $\mathbf{cov(X_{1i}, U_i) \approx cov(X_{2i}, U_i) \approx cov(X_{3i}, U_i) \approx 0}$ We can safely say that there is no problem of Endogeneity and hence the estimators are unbiased and consistent.

Assumption 5
Homoscedasticity: no matter what the values of our regressors might be, the error of the variance is constant.
Goldfeld - Quandt Test :

Model:

$$Y_i = \alpha + \beta X_i + u_i$$

Assumption:

$$\sigma_{i^2} = \sigma^2 X_{i^2}$$

H_0 : $\sigma_1^2 = \sigma_2^2$

RSS1	928.016601
RSS2	135.4374942
λ	6.851991812
f-stat	3.51152E-06

The λ **= 6.851**, so the **f-stats come out to be 3.51E-06** which shows that the null hypothesis has been rejected at 1% and hence there is severe heteroscedasticity.

CONCLUSION:

We observe that Human Inequality is related to poverty with an increase in human inequality we observe a significant increase in the poverty rate from which we can conclude that both exhibit positive relation.
From the second regression result we see Human Inequality is mainly dependent on the Gender inequality index, Gini Wealth Coefficient Index, and the Human development index.
While checking the Gauss-Markov Assumptions for the model we find that the model is non-linear, normality is violated, the existence of severe heteroscedasticity but the estimators are unbiased and consistent.

REFERENCES:

1. Current health expenditure (% of GDP) | Data World Bank
2. Compulsory education - Wikipedia
3. Government expenditure on education, total (% of GDP) | Data World Bank Data
4. List of countries by spending on education (% of GDP) - Wikipedia

5.Excel File

Countries List:

Angola, Albania, Argentina, Armenia, Australia, Austria, Belgium, Benin, Burkina Faso, Bangladesh, Bulgaria, Belize, Brazil, Canada, Switzerland, Chile, China, Colombia, Costa Rica, Cyprus, Germany, Denmark, Dominican Republic, Algeria, Ecuador, Spain, Estonia, Ethiopia, Finland, France, Gabon, United Kingdom, Georgia, Ghana, Greece, Guatemala, Guyana, Honduras, Croatia, Haiti, Hungary, Indonesia, India, Ireland, Iraq, Iceland, Israel, Italy, Jamaica, Jordan, Japan, Kenya, Sri Lanka, Lithuania, Luxembourg, Latvia, Mexico, Mali, Mongolia, Mauritania, Mauritius, Malawi, Namibia, Nicaragua, Netherlands, Norway, Nepal, Pakistan, Panama, Peru, Philippines, Poland, Portugal, Senegal, Sierra Leone, El Salvador, Serbia, Suriname, Slovenia, Sweden, Chad, Togo, Thailand, Tunisia, Turkey, Uganda, Uruguay, United States, South Africa, Zambia, Zimbabwe