Econometrics Lab (HS49002)

Lab Project: Group 12

Date: 06 April 2022

Submitted to Prof. Pulak Mishra

Submitted by:-

Anjali Meena	19HS20008
Ankit Pal	19AR10007
Macharla Sai Manohar	19HS20027
Maitri Das	19HS20028
Rohit Jain	19AR10028
Turre Sai Girish	19HS20045

OBJECTIVE

In this report, we aim to provide an analysis of the determinants of the Human Development Index (HDI) and how various factors influence it across different countries (for the year 2019).
The Human Development Index (HDI) is a statistic developed and compiled by the United Nations to measure various countries' levels of social and economic development. It is composed of four principal areas of interest: Mean years of schooling, expected years of schooling, life expectancy at birth, and gross national income (GNI) per capita. This index is a tool used to follow changes in development levels over time and compare the development levels of different countries. We expect that after observing the results, we might have a better understanding of the contribution of each factor and correlation between them.

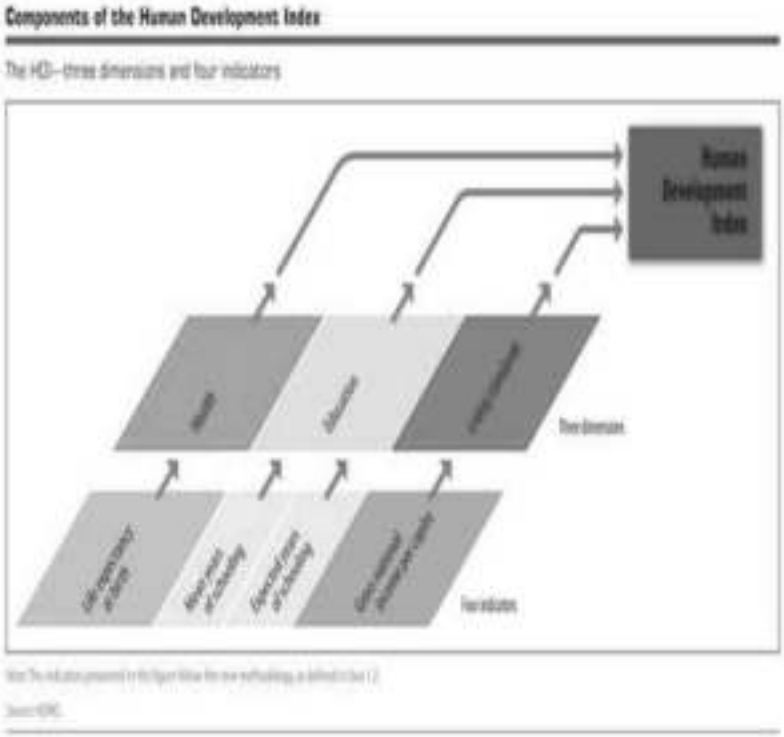
SPECIFICATION

Initially, we chose the following independent variables that affect the HDI

- Life expectancy at birth
- Expected years of schooling
- Mean years of schooling
- Gross national income (GNI) per capita

Our Dependent Variable (Human Development Index) was of the year 2019.

Our base model is
Human Development Index$_{2019}$ = f(Life expectancy at birth$_{2019}$, Expected years of schooling$_{2019}$, Mean years of schooling$_{2019}$, Gross national income (GNI) per capita$_{2019}$)



Source: UNDP.org

MEASUREMENT OF VARIABLES

I. Dependent Variable

Human Development Index - The Human Development Index (HDI) is an index that measures key dimensions of human development. The data is taken for the year 2019.

II. Independent Variables

Life expectancy at birth -

Life expectancy at birth reflects the overall mortality level of a population. It summarizes the mortality pattern that prevails across all age groups - children and adolescents, adults and the elderly. Average number of years that a newborn is expected to live if current mortality rates continue to apply.

Expected years of schooling

Expected years of schooling is the number of years a child of school entrance age is expected to spend at school, or university, including years spent on repetition. It is the sum

of the age-specific enrolment ratios for primary, secondary, post-secondary non-tertiary and tertiary education.

Mean years of schooling

Mean years of schooling (MYS), the average number of completed years of education of a population, is a widely used measure of a country's stock of human capital.

Gross national income (GNI) per capita

The GNI per capita is the dollar value of a country's final income in a year, divided by its population. It Reflects the average before tax income of a country's citizens.

SOURCE OF DATA

United Nations Development Programme - Human development reports
https://hdr.undp.org/en/content/download-data

IMPACT OF INDEPENDENT VARIABLES

We expect a positive impact of all of the following independent variables on the HDI

Life expectancy at birth
Expected years of schooling
Mean years of schooling
Gross national income (GNI) per capita

Model 1:

ESTIMATION TECHNIQUE & PROCESS FOLLOWED

We estimate our model using the method of OLS. We used the following independent variables in our initial baseline model: Life expectancy at birth, Expected years of schooling Mean years of schooling & Gross national income (GNI) per capita.

```
. regress HumanDevelopmentIndexHDI Lifeexpectancyatbirth Expectedyearsofschooling Meanyearsofschooling Gross
> nationalincomeGNIperc
```

Source	SS	df	MS
Model	13.0201194	4	3.25502986
Residual	.287825935	184	.001564271
Total	13.3079454	188	.070786943

Number of obs = 189
F(4, 184) = 2080.86
Prob > F = 0.0000
R-squared = 0.9784
Adj R-squared = 0.9779
Root MSE = .03955

HumanDevelopmentIndexHDI	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Lifeexpectancyatbirth	.3725542	.0226817	16.43	0.000	.3278045	.4173038
Expectedyearsofschooling	.3898064	.0319071	12.22	0.000	.3268556	.4527571
Meanyearsofschooling	.3935353	.0218474	18.01	0.000	.3504317	.4366389
GrossnationalincomeGNIperc	.1730628	.0182428	9.49	0.000	.1370709	.2090548
_cons	-.0958464	.0100154	-9.57	0.000	-.1156062	-.0760866

We'll now perform various tests such as tests for multicollinearity, heteroscedasticity, and testing for normality of random disturbance term. Also since the data is available only over a year, we need not test for autocorrelation or structural breaks.

Multicollinearity:

Multicollinearity is caused due to reasons such as inclusion of variables computed from other variables in the equation, inclusion of similar variables more than once etc. It is important to reduce the extent of multicollinearity as it has various implications like R squares can be very high and greater multicollinearity raises standard errors.
In our model after computing the regression results and individual significance statistical tests, we move towards checking multicollinearity.

VIF:

Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables. Mathematically, the VIF for a regression model variable is equal to the ratio of the overall model variance to the variance of a model that includes only that single independent variable. We checked the VIF of our current model which came

to be 3.16. The value of VIF indicated that multicollinearity is severe.

```
8 . vif

    Variable |       VIF       1/VIF
-------------+----------------------
Expectedye~g |      3.66    0.273441
Meanyearso~g |      3.44    0.290555
Lifeexpect~h |      3.38    0.296019
Grossnatio~c |      2.15    0.466129
-------------+----------------------
    Mean VIF |      3.16
```

Farrar Glauber Test:

To confirm this, we performed further analysis using the Farrar Glauber test. We fail to reject the hypothesis that the variables are orthogonal. Hence, the results from the Farrar Glauber test showed that multicollinearity existed.

```
. fgtest HumanDevelopmentIndexHDI Lifeexpectancyatbirth averageedu GrossnationalincomeGNIperc

-----------------------------------------------------------------------------
* Farrar-Glauber Multicollinearity Tests
-----------------------------------------------------------------------------
  Ho: No Multicollinearity - Ha: Multicollinearity

* (1) Farrar-Glauber Multicollinearity Chi2-Test:
    Chi2 Test =  346.3453    P-Value > Chi2(3) 0.0000

* (2) Farrar-Glauber Multicollinearity F-Test:
```

Variable	F_Test	DF1	DF2	P_Value
Lifeexpe~h	215.068	186.000	2.000	0.005
averageedu	214.852	186.000	2.000	0.005
Grossnat~c	106.466	186.000	2.000	0.009

```
* (3) Farrar-Glauber Multicollinearity t-Test:
```

Variable	Life~h	aver~u	Gros~c
Lifeex~h	.		
averag~u	19.269	.	
Grossn~c	13.332	13.223	.

Heteroskedasticity Test:

We also tried the heteroskedasticity test. Heteroskedasticity arises when the variance of the random disturbance term is not constant across observations. It usually arises in

cross-sectional data, but may also be present in time-series data. Since our data is cross sectional data, there is a possibility that the problem of heteroscedasticity may be present

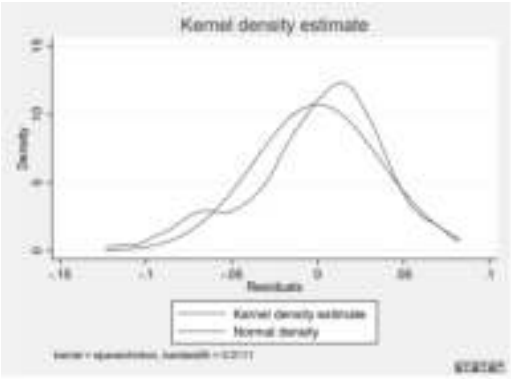
We perform the Breusch-Pagan test to check for heteroskedasticity which has constant variance across observations as its null hypothesis. On performing the test we get a very low p-value and hence accept the null hypothesis. Hence our estimated model does suffer from the problem of heteroscedasticity. To reaffirm our results, we also perform White's test, which uses the null hypothesis that our estimated model has homoskedasticity. We again get a p-value lower than 0.1 and hence accept the null-hypothesis. Hence we can conclude that the estimated model suffers from the problem of heteroscedasticity.

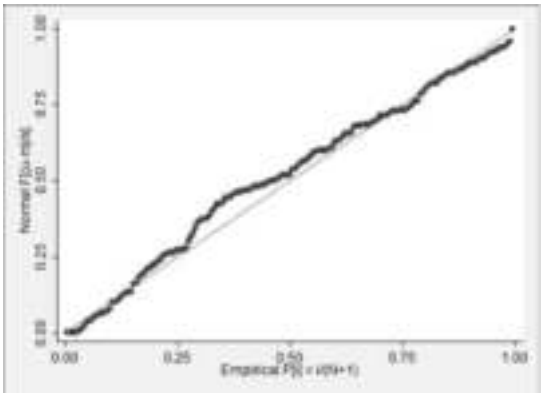
```
11 . estat hettest

   Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
            Ho: Constant variance
            Variables: fitted values of HumanDevelopmentIndexHDI

            chi2(1)      =     8.69
            Prob > chi2  =   0.0032

12 . estat hettest Lifeexpectancyatbirth Expectedyearsofschooling Meanyearsofschoo
   > ling GrossnationalincomeGNIperc

   Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
            Ho: Constant variance
            Variables: Lifeexpectancyatbirth Expectedyearsofschooling Meanyearsofs
   > chooling
                      GrossnationalincomeGNIperc

            chi2(4)      =    28.86
            Prob > chi2  =   0.0000

13 . estat imtest, white

   White's test for Ho: homoskedasticity
            against Ha: unrestricted heteroskedasticity

            chi2(14)     =     43.16
            Prob > chi2  =    0.0001

   Cameron & Trivedi's decomposition of IM-test

   ---------------------------------------------------
                Source |       chi2     df      p
   ---------------------+-----------------------------
     Heteroskedasticity |      43.16     14    0.0001
               Skewness |      10.90      4    0.0277
               Kurtosis |       3.25      1    0.0716
   ---------------------+-----------------------------
                  Total |      57.31     19    0.0000
   ---------------------------------------------------
```

Skewness/Kurtosis tests for Normality:

We conducted this test to check if the random disturbance term follows normal Distribution.

```
4 . sktest u

                     Skewness/Kurtosis tests for Normality
                                                       ------- joint ------
    Variable |   Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)    Prob>chi2
-------------+-------------------------------------------------------------
           u |   189       0.0008         0.0534        12.71          0.0017
```





So to solve the problem with the previous model we replace the variables.

SPECIFICATION II

We chose the following independent variables that affect the HDI

- Current Health expenditure
- Government expenditure on education
- Gross national income (GNI) per capita

Our Dependent Variable (Human Development Index) was of the year 2019.

Our new model is:
Human Development Index2019 = f(Current Health expenditure2019, Government expenditure on education2019, Gross national income (GNI) per capita2019).

MEASUREMENT OF VARIABLES

I. Dependent Variable

Human Development Index - The Human Development Index (HDI) is an index that measures key dimensions of human development. The data is taken for the year 2019.

II. Independent Variable

Current Health expenditure

Current health expenditure as a share of GDP provides an indication on the level of resources channeled to health relative to other uses. It shows the importance of the health sector in the whole economy and indicates the societal priority which health is given measured in monetary terms.

Government expenditure on education

General government expenditure on education (current, capital, and transfers) is expressed as a percentage of total general government expenditure on all sectors (including education, social services, etc.). It includes expenditure funded by transfers from international sources to the government.

Gross national income (GNI) per capita

The GNI per capita is the dollar value of a country's final income in a year, divided by its population. It Reflects the average before tax income of a country's citizens.

IMPACT OF INDEPENDENT VARIABLES

We expect a positive impact of all of the following independent variables on the HDI

Current Health expenditure
Government expenditure on education
Gross national income (GNI) per capita

Model 2:

ESTIMATION TECHNIQUE & PROCESS FOLLOWED

We estimate our model using the method of OLS. We used the following independent variables in our initial baseline model: Current Health expenditure, Government expenditure on education & Gross national income (GNI) per capita.

```
. regress HumanDevelopmentIndexHDI CurrentHealthexpenditure Governmentexpenditureoneducati NGrossnationalincomeGNIpe
```

Source	SS	df	MS
Model	3.29305934	3	1.09435311
Residual	.198292662	140	.001416376
Total	3.491352	143	.024345119

Number of obs	= 144
F(3, 140)	= 772.64
Prob > F	= 0.0000
R-squared	= 0.9432
Adj R-squared	= 0.9418
Root MSE	= .03763

HumanDevelopmentIndexHDI	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
CurrentHealthexpenditure	.0040846	.0013563	3.01	0.003	.0014031	.0067661
Governmentexpenditureoneducati	.0038469	.002194	1.75	0.082	-.0004909	.0081846
NGrossnationalincomeGNIpe	.1229325	.0027582	44.57	0.000	.1174795	.1283856
_cons	-.46706	.0280441	-16.65	0.000	-.5225736	-.4115465

The adjusted R-squared value is 0.9418. Hence, the regression has a very high explanatory power. All the three variables have a positive coefficient that means all of the variables impact positively on the HDI which is as expected.

We'll now perform various tests such as tests for multicollinearity, heteroscedasticity, and testing for normality of random disturbance term. Also since the data is available only over a year, we need not test for autocorrelation or structural breaks.

Following the regression we perform the VIF test to check for multicollinearity.

VIF

```
. vif
```

Variable	VIF	1/VIF
CurrentHea~e	1.26	0.793884
Government~i	1.21	0.824440
NGrossnati~e	1.09	0.913851
Mean VIF	1.19	

As the mean VIF is below 1.5, the revised variables for our dataset do not show multicollinearity. The probable reason for our previous dataset to show multicollinearity was the variables we used to determine the HDI were the variables which we were using in our model and due to this there might be some overlapping.

As VIF is between 1 and 1.5 we do not perform the Farrar-Glauber Multicollinearity test and we test for heteroskedasticity.

Breusch-Pagan / Cook-Weisberg test

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of HumanDevelopmentIndexHDI

         chi2(1)      =     6.82
         Prob > chi2  =   0.204
```

As the hypothesis of constant variables is rejected there is no heteroskedasticity.

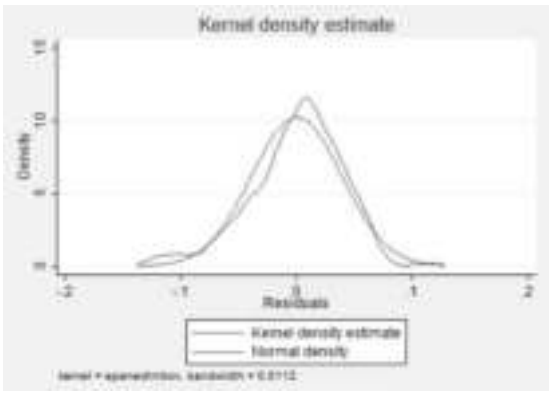
Skewness/Kurtosis tests for Normality:

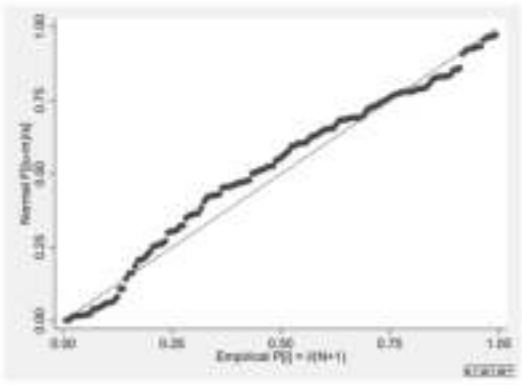
We conducted this test to check if the random disturbance term follows normal Distribution.

```
. kdensity u,normal

. pnorm u

. sktest u
```

Skewness/Kurtosis tests for Normality

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	joint adj chi2(2)	joint Prob>chi2
u	144	0.0016	0.3616	9.49	0.0087





Here an interesting fact to notice is that the graph follows bell curve (less skewed) in the revised data than the previous data and also the normality graph for both the datasets is similar, Hence we can say that the selection of the revised variables are a good substitution for the principle areas of interest we had in our objective.

Test for Linearity

We performed a test for linearity to find if our function is linear or multinomial. We predicted Y then generated Y^2, Y^3 and Y^4. Then we regress the data using our variables and Y^2, Y^3 and Y^4

```
. regress HumanDevelopmentIndexHDI CurrentHealthexpenditure Governmentexpenditureoneeducati NGrossnationalincomeGNIpe Y4 Y3 Y2

      Source |       SS       df       MS              Number of obs =     144
-------------+------------------------------           F(  6,   137) =  426.31
       Model |  3.30436943     6  .550728238           Prob > F      =  0.0000
    Residual |  .17698257    137  .001291844           R-squared     =  0.9492
-------------+------------------------------           Adj R-squared =  0.9469
       Total |  3.481352     143  .024345119           Root MSE      =  .03594

------------------------------------------------------------------------------
HumanDevelopmentIndexHDI |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------------+----------------------------------------------------
CurrentHealthexpenditure |  -.0265827   .0306483    -0.87   0.387    -.0871875    .0340221
Governmentexpenditureone~i|  -.0255179   .0289624    -0.88   0.380     -.082789    .0317534
NGrossnationalincomeGNIpe|  -.8226084   .9196936    -0.89   0.373    -2.641239    .9960224
                      Y4 |   2.82735   5.865517     0.48   0.630    -8.770855    14.42555
                      Y3 |  -10.38297   16.39459    -0.63   0.528    -42.80214    22.03621
                      Y2 |   14.02519   16.76714     0.84   0.404    -19.17023    47.22061
                   _cons |   4.591305   4.716492     0.97   0.332    -4.738232    13.91084
------------------------------------------------------------------------------

. test Y2 Y3 Y4

 ( 1)  Y2 = 0
 ( 2)  Y3 = 0
 ( 3)  Y4 = 0

       F(  3,   137) =    5.50
            Prob > F =    0.0013
```

As the result is significant at 1% it is not a linear function.

RESULT & IMPLICATION

Our final model comes out to be -
Human Development Index2019 = f(Current Health expenditure2019, Government expenditure on education2019, Gross national income (GNI) per capita2019).

```
. regress HumanDevelopmentIndexHDI CurrentHealthexpenditure Governmentexpenditureoneducati NGrossnationalincomeGNIpe
```

Source	SS	df	MS	Number of obs =	144
				F(3, 140) =	772.64
Model	3.29305934	3	1.09435311	Prob > F =	0.0000
Residual	.198292662	140	.001416376	R-squared =	0.9430
				Adj R-squared =	0.9418
Total	3.491352	143	.024345119	Root MSE =	.03763

HumanDevelopmentIndexHDI	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
CurrentHealthexpenditure	.0040846	.0013563	3.01	0.003	.0014031	.006766
Governmentexpenditureoneducati	.0038469	.002194	1.75	0.082	-.0004909	.0081846
NGrossnationalincomeGNIpe	.1229325	.0027562	44.87	0.000	.1174796	.1283856
_cons	-.46706	.0250441	-18.65	0.000	-.5165736	-.4175465

The model estimated is statistically significant with an adjusted R-squared value of 0.9430. Hence it has a very high explanatory power.

All the variables have a positive impact on HDI.

CONCLUSION & FURTHER TESTING/EXAMINATION

We conclude that determinant factors of human development index (HDI) like Current Health expenditure, Government expenditure on education, Gross national income (GNI) per capita are significant in empirical regression analysis and are positively related. Hence, The improvement in HDI can be achieved through three dimensions – Education index, Income index and Health index.

Robustness tests can be carried out to test whether our results would still hold true if a different estimation technique than OLS is used or if a different measure of HDI is used.

Tests must also be carried out to mitigate endogeneity concerns because it is possible that an omitted variable drives both the dependent and independent variables or it's possible

that the dependent variable also affects the distribution of independent variables. So we must apply tests to prove the exogenous variation of the independent variables.

We must also check for selection bias. Selection bias occurs when the procedures used to select subjects and other factors that influence participation in the study produce a result that is different from what would have been obtained if all members of the target population were included in the study. Since data contains only the variables from the year 2019, it is possible that if a different year was chosen we might have got different results.

We must check whether there is any information bias since systematic errors in the measurement of independent variables is possible.