Does HDI have an impact on air pollution, using instrumental variable approach?

Abstract

Air pollution is a major concern as it deeply impacts the health, nature, and economy of a nation. But it would be interesting to know if factors like healthier life or better economic conditions help in combating with the problem of air pollution. This study aims to address this concern by looking at the correlation between air pollution and Human Development Index (HDI). Various indices like life expectancy index, GNI, and education combine to form HDI depicting the human development of a nation. However, this model faces the problem of simultaneous causality bias. To account for this threat to internal validity, instrumental variable has been used on a panel dataset. The importance of this study is to provide a quantitative basis using IV for researchers who come across the question of impact of HDI on air pollution. There isn't much empirical evidence on this topic with the use of IV. This study would be one of its kind. The results found in the study show highly statistically significant decrease in air pollution with 1 unit increase in HDI of a country.

1. Background and Literature Review

Something that every nation is doing is running after numbers. Later these numbers are running after us. The amount of air pollution is increasing causing harm to human beings and the planet Earth. With this rising concern, people have been indulging into research to finding out the causes of increasing air pollution.

HDI has been used as a proxy for social economic status of a country with life expectancy index and education treated as social factors, and GNI as economic factor. At first glance, it might seem increasing HDI leads to a fall in the air pollution. However, it has been found by researchers that increasing HDI results in even more air pollution than before.

Xiaoyu Li[1] investigates the connection between human development and overall environmental quality to find out there is an inverted U-shaped relationship. The first power of HDI is positive but the second power of HDI is negative. Initially, the air pollution increases with increase in HDI, but after a point it start to decrease. This could also explain why developing nations have a higher air pollution as compared to the developed nations.

Săndică[2] points towards another reason for a positive correlation between air pollution and HDI. According to the correlation matrix obtained in the results, PM2.5 is negatively correlated with life expectancy while positively correlated with education and real GDP/capita.

2. Data

The study deals with panel data of OECD countries for 3 years, 2012, 2015, and 2018. Table 2 below lists down all variables used in the study along with their definition and unit of measurement.

Datasets	Variables	Definition	Measurement Unit	Source
Particulate Matter 2.5	PM25	Annual mean concentration of particulate matter of less than 2.5 microns of diameter in urban areas.	(PM2.5) [ug/m3]	World Health Organization
Human Development Index	HDI	Summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living.	0 (Lowest HDI) – 1 (Highest HDI)	United Nations Development Programme
Total industrial production	Indus_prod	Total output of industrial establishments, covering sectors like mining, manufacturing, electricity, gas and steam, and air-conditioning.	Index (2015=100)	Organisation for Economic Co-operation and Development
Ppopulation growth rate	Pop_growth	Total number of residents irrespective of legal status or citizenship	Percentage (%)	World Bank - Databank
Programme for International Student Assessment	PISA	Test that provides school-level estimates of performance and information about the learning environment and students' attitudes gathered from student questionnaires.	Average mathematic test scores[1]	Organisation for Economic Co-operation and Development

[1] The data for PISA had scores displayed in 7 levels for all countries. Each level represented a range of score. To compute the average math test scores, I took the average of all levels.

One of the most widely used measure of air pollution is Particulate Matter 2.5 (PM 2.5) which is the main dependent variable of the study. The data for PM 2.5 (*PM25*) is the mean annual concentration of fine suspended particles of less than 2.5 microns in diameters. The main independent variable of the study is Human Development Index (HDI). HDI is a composite score that considers three dimensions: (1) health, which is assessed by life expectancy at birth, (2) education, which is measured by mean years of schooling for people aged 25, and (3) standard of living, which is measured by gross national income per capita.

The covariates include industrial production (*Indus_prod*) that is expected to have a positive correlation with HDI. Increasing amount of production would lead to higher GNI and further better education means. Accelerating the production might also have an impact on the air pollution. *Indus_prod* measures the total production, including construction and manufacturing for every country. The data is in the form of an index that considers 2015 (=100) as the reference period.

The study introduces Programme for International Student Assessment (PISA) as an instrumental variable to account for the problem of simultaneous causality. It is an international student assessment test conducted after every 3 years to assess the learning outcome of 15-year-old children. The two main conditions for a variable to be considered an instrumental variable are:-

1.) Instrument Relevance: corr (Zi ,Xi)≠ 0 ; Zi = instrumental variable, Xi = main explanatory variable
2.) Instrument Exogeneity: corr (Zi ,ui) = 0 ; ui = error term

The fulfilment of first condition can be observed from table 2.1.2. The second condition is not very strong as it is not random. However, it can be considered an IV because the PISA test reveals the learning outcome of 15-year-old children, who in general, do not study for the test with an aim of reducing air pollution. And environmental tobacco smoke (ETS), not so common in children, is one of

the primary sources of PM2.5. This test is also not designed to assess students' ability to reduce air pollution. Moreover, the PISA test scores act as a proxy for students' ability as the study only uses the math test scores and not the reading test scores.

Summary Statistics

To investigate the data further, I conducted summary statistics represented in Table 2.1 below, and used data visualization to easily identify correlations and patterns before moving on to conducting regression analysis.

Table 2.1.1 Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Year	111	2015	2.461	2012	2018
Indus prod	107	101.346	7.437	61.767	120.693
Pop growth	111	.579	.784	-1.341	2.678
PISA	111	507.862	11.859	445.519	514.968
HDI	111	.893	.049	.734	.962
PM25	111	13.406	5.619	5.57	27.61
HDI2	111	.801	.084	.539	.925

Table 2.1.1 shows the number of observations to be 111. The only missing observations is seen in Indus_prod which is the covariate of the model. The minimum and maximum values of HDI are greater than 0.5. It implies that the data only deals with countries that have high HDI. Similarly, the minimum and maximum values of PISA are not extremely low which is in alignment with the HDI data. Pop_growth has a negative minimum value which implies that some country(ies) has a declining population. The mean value of PM25 is approximately at the center of minimum and maximum values giving a vague indication of the countries having a similar quantity of air pollutants. Since, the countries used in the study have a high HDI, the former entityment is possible. To find this out, I performed some data visualizations, depicted below.

Graphs are created for all the main variables (dependent and independent). The histogram had been created for the covariate as well. However, it is not displayed in this paper. It is available in the do file for the viewer's reference.

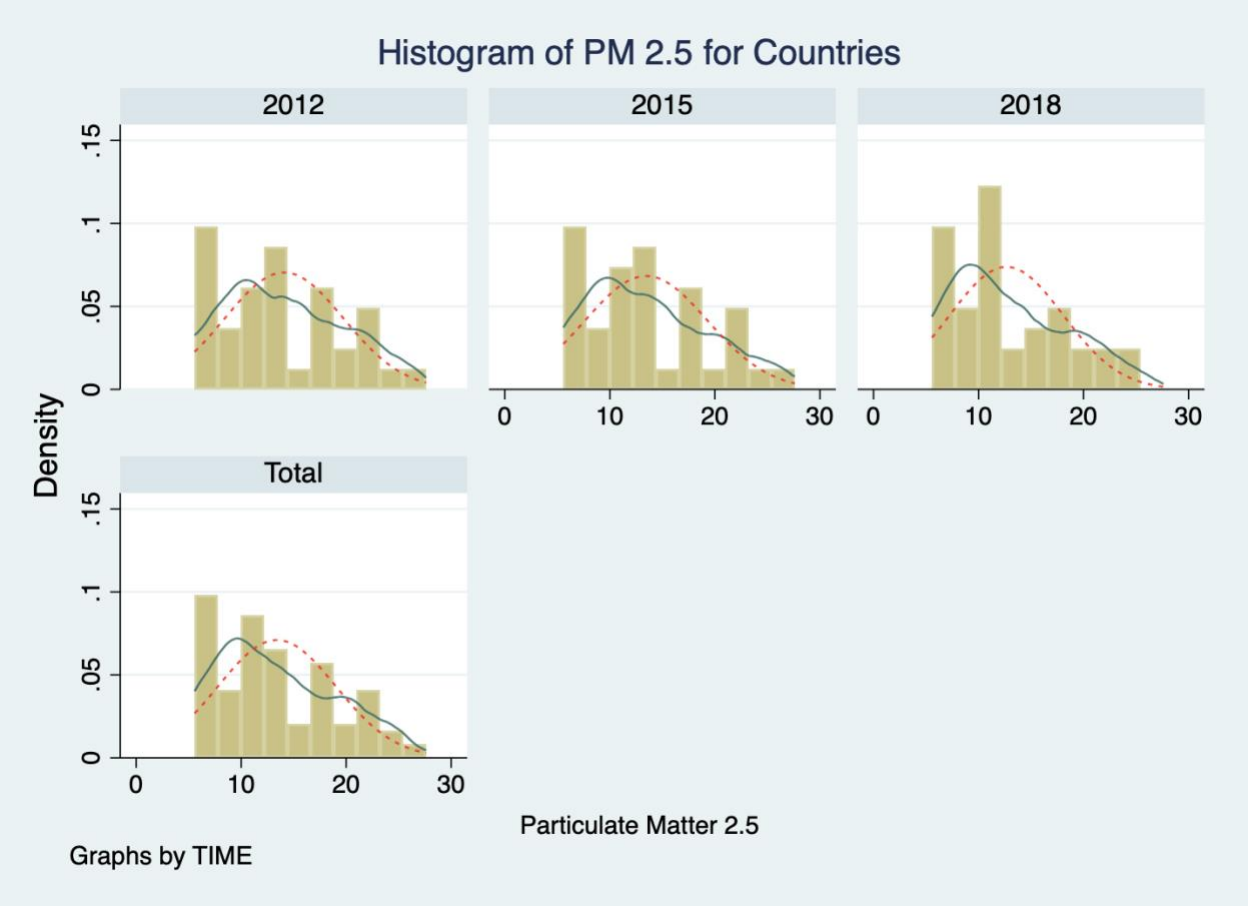

Fig. 2.1.1 PM 2.5

Figure 2.1.1 displays histograms of PM 2.5 year-wise for all countries. The data appears dispersed indicating possibility of outliers in data. There are multiple bumps (or peaks in the graph) observed. In 2018, multiple bumps start to disappear. In the total graph, before the second low peak, the graph appears smooth while coming down. Furthermore, in the total graph, the highest peak has gone up and there is a formation of second highest peak around 20 ug/m3, which is higher as compared to 2012 third lowest visible peak. This clearly indicates a rise in the air pollution.

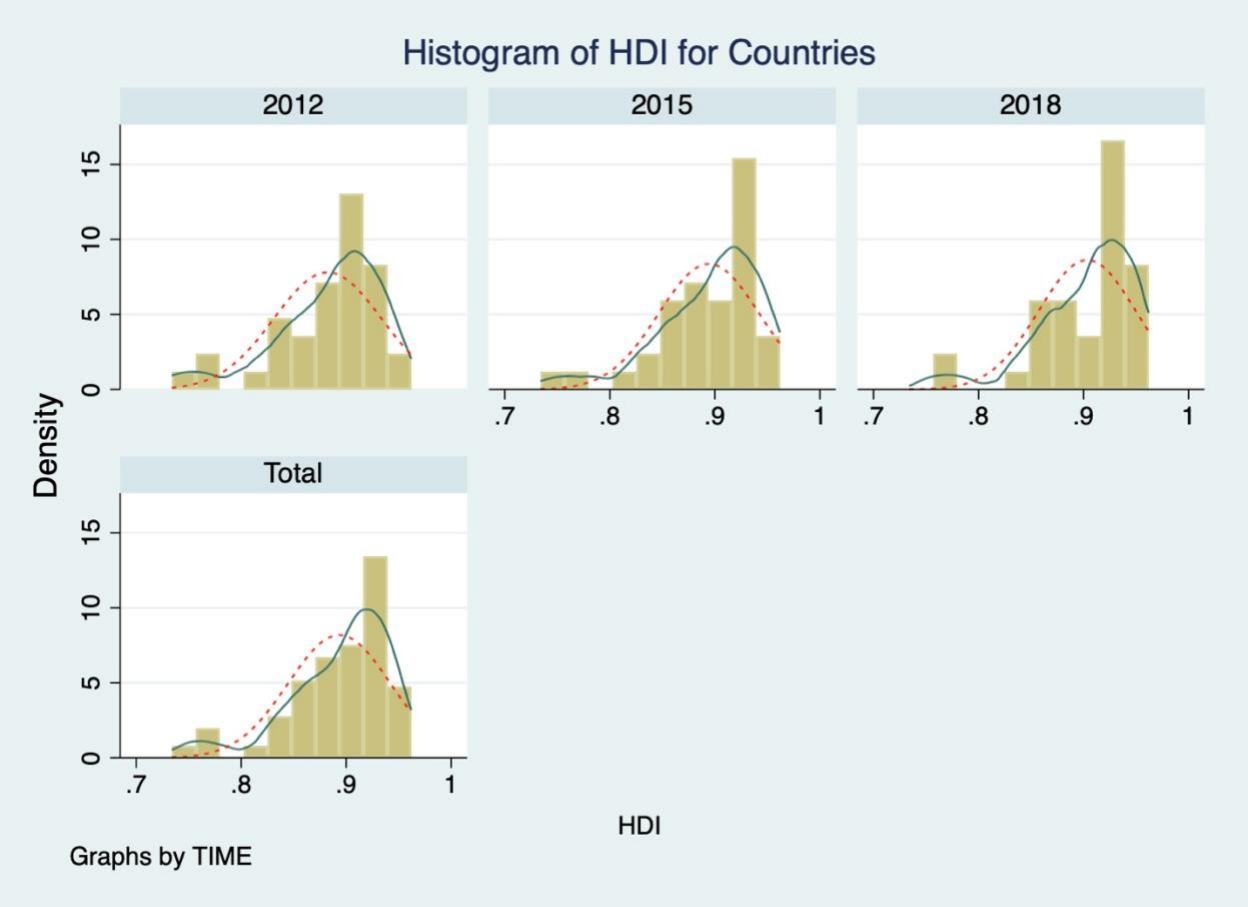

Fig. 2.1.2 HDI

Figure 2.1.2 displays histograms of HDI year-wise for all countries. The data appears skewed towards left side, unlike in the previous graph. Over the years, the highest bar has been increasing indicating countries moving closer to better living conditions. Furthermore, there are only few countries with low HDI conforming with the summary statistics of the study.

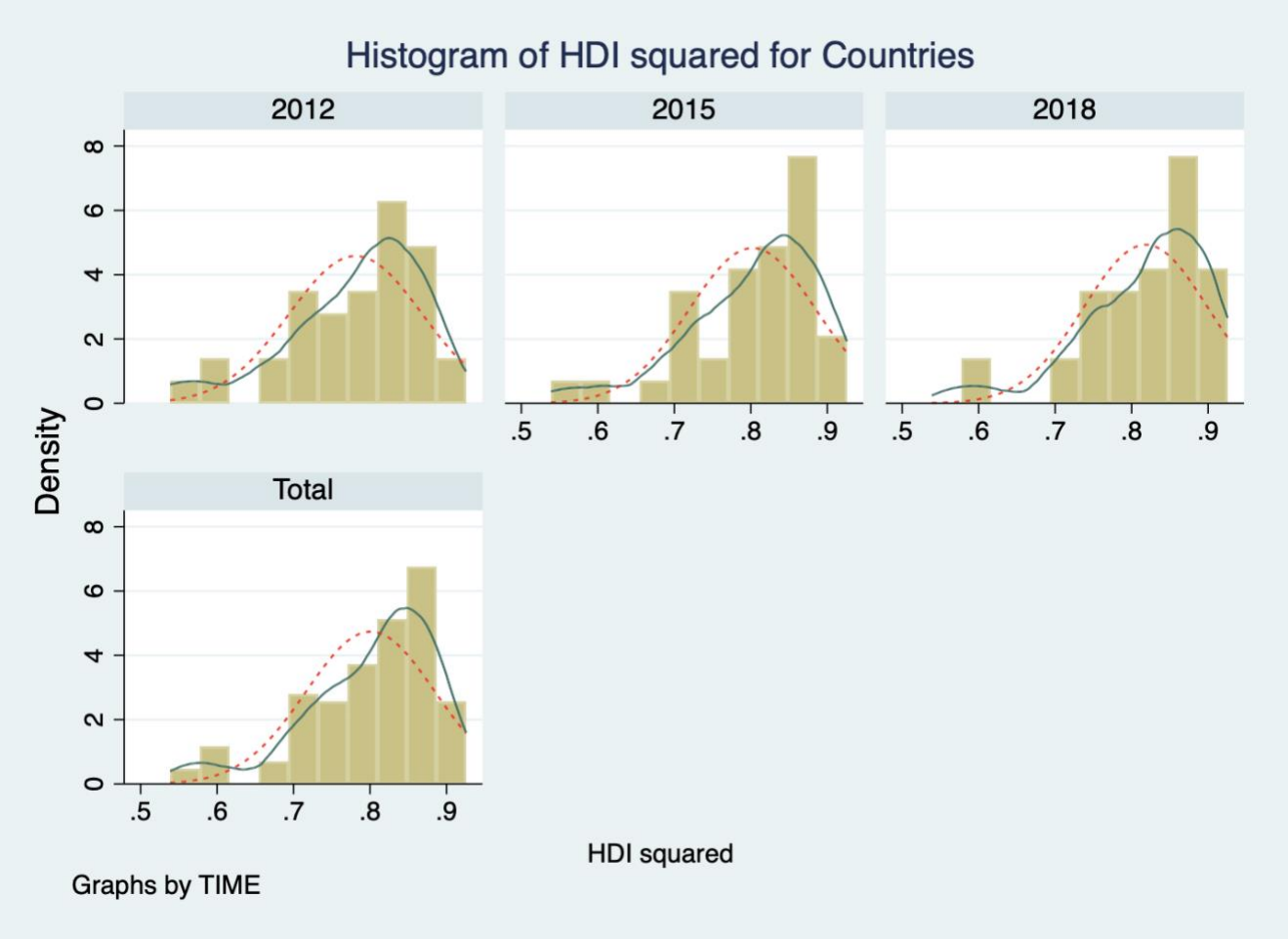

Fig. 2.1.3 HDI squared

Figure 2.1.3 displays histograms of HDI squared year-wise for all countries. The data appears skewed towards left side, similar to the previous graph. Over the years, the highest bar has been increasing indicating countries moving closer to better living conditions. Furthermore, there are only few countries with low HDI conforming with the summary statistics of the study.

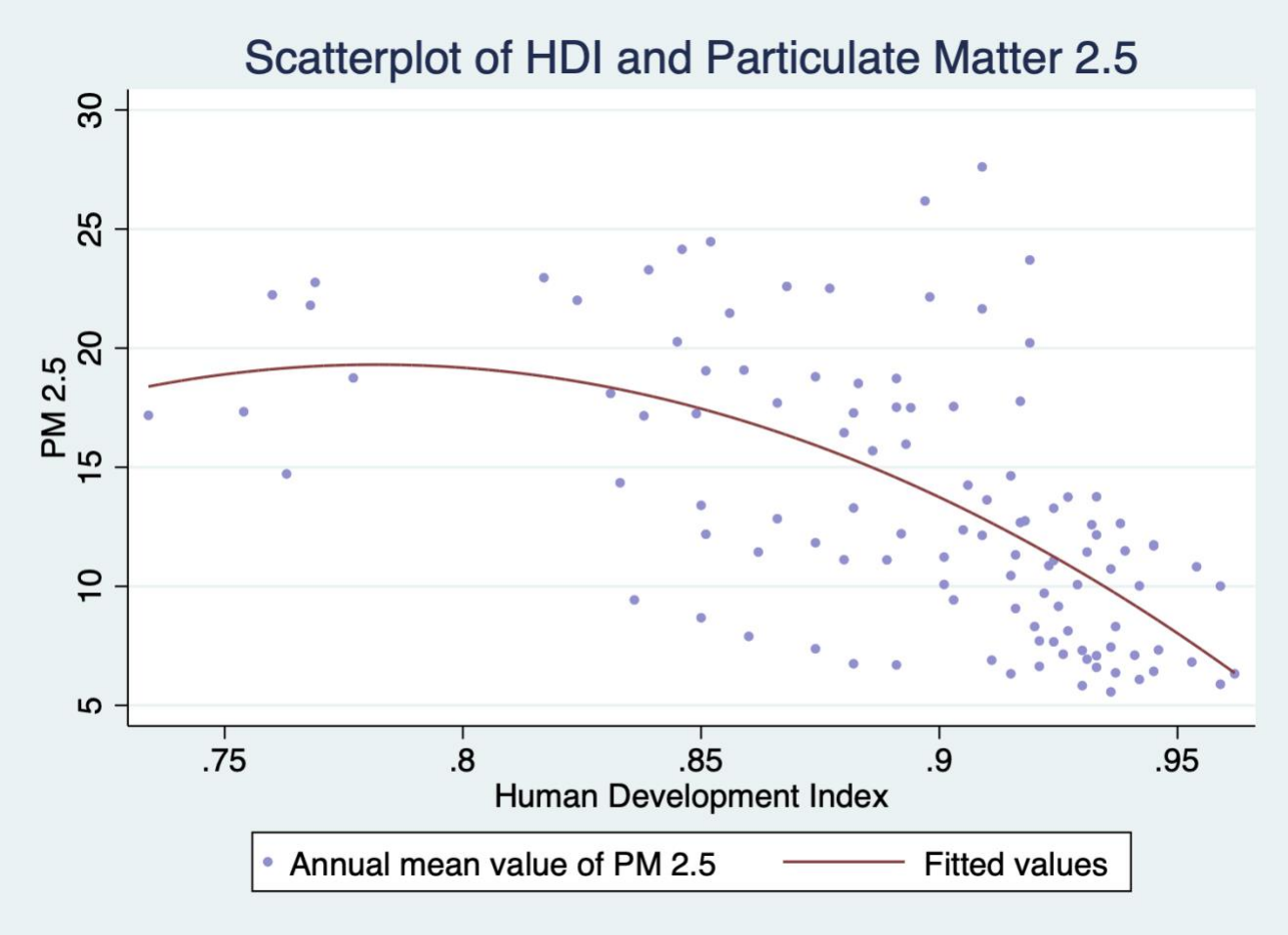

Fig. 2.1.3 Scatter plot of main variables of the study

Figure 2.1.3 showcases the quadratic relation between HDI and PM 2.5 in the form of a scatter plot with the blue line displaying the fitted quadratic trend line. The unit of observation is country represented by the lavender colored dot. The upward slope until ~.8 units of HDI show a positive correlation between HDI and PM 2.5. Further, when the trend line slopes downward, it shows the negative correlation between HDI and PM 2.5. This is in alignment with the literature review found and my belief about this study.

Table 2.1.2 Matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) PM25	1.000					
(2) HDI	-0.546	1.000				
(3) HDI2	-0.553	0.999	1.000			
(4) Indus_prod	0.009	0.086	0.087	1.000		
(5) Pop_growth	0.000	0.150	0.168	-0.010	1.000	
(6) PISA	-0.367	0.674	0.661	-0.069	-0.233	1.000

The two conditions of omitted variable bias are:

1. Correlation of omitted variable with X (main explanatory variable)

2. Determinant of Y (main dependent variable)

The first condition has been fulfilled, as observed in column (2) and (3) (highlighted in yellow) of table 2.1.2. To observe if these variables also fulfil condition 2, regression analysis have been conducted including and excluding each variable.

Furthermore, correlation of PISA with the main independent variable of study (highlighted in green) can also be observed in column (2) of table 2.1.2. This fulfills the first condition for consideration of PISA as an IV in this study.

3. Methodology

The initial methodologies used are linear and multiple regression models using OLS estimator in the form of pooled data. Further, entity fixed effects and time fixed effects methods are used separately and combined. And finally, regression is conducted using instrumental variable methodology with the help of 2SLS estimation. All results are obtained using the robust command taking into consideration any possibility of heterogeneity.

The general form of regression model is,

$$PM25_{it} = \beta_0 + \beta_1 HDI_{it} + \beta_2 HDI2_{it} + \beta_3 Indus\_prod_{it} + \beta_4 Pop\_growth_{it} + u_{it}$$

; u is the error term

i are the entity fixed effects

t are the time fixed effects

Table 3.1 Pooled Data

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
HDI	977.5***	304.3	970.6***	17.04
	(211.3)	(186.1)	(275.5)	(81.57)
HDI2	-606.1***	-199.2*	-601.5***	3.095
	(123.1)	(104.8)	(160.0)	(48.78)
Indus_prod	0.0557	-0.00878	0.0885	0.0138
	(0.0495)	(0.0144)	(0.0782)	(0.00854)
Pop_growth	1.871***	0.271		
	(0.607)	(0.320)		
Constant	-381.2***	-98.12	-381.8***	-4.607
	(90.38)	(82.56)	(118.8)	(34.10)
Observations	107	107	107	107
Adjusted R-squared	0.371	0.409	0.361	0.639
Number of Country_code		37		37
Country_code Fixed Effects		Yes	No	Yes
Time Fixed Effects			Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In table 3.1,

- Model 1 represents pooled data using OLS estimation
- Model 2 shows the pooled data using OLS estimation adjusted for entity fixed effects only.
- Model 3 shows the pooled data using OLS estimation adjusted for time fixed effects only
- Model 4 displays the pooled data using OLS estimation adjusted for both entity and time fixed effects

Model 2: It fixes for any change over countries, in the effect common during all years. Therefore, it adjusts for any unobserved omitted variables that are time invariant but changing between countries.
Model 3: It fixes for any change over years, in the effect common to all countries. Therefore, it adjusts for any unobserved omitted variables that are country invariant but changing over time.

An important thing to note is that, $\widehat{\beta_1}$ appears to be positive and $\widehat{\beta_2}$ negative, in most of the model, signifying increasing air pollution with every additional unit of HDI up until the point a country reaches 0.8 (according to the Model 1 of table 3.1) units of HDI. After this point, the air pollution starts decreasing with every additional unit of HDI. Another interesting thing to note is that this study is conducted only on countries with a high HDI. If low HDI countries had been considered, the turning point might have been much lower than 0.8.

3.1.1 Comparing models

All models, except Model 2 and Model 4, have a statistically significant $\widehat{\beta_1}$ at 99% confidence interval (CI). Furthermore, except Model 4, all models have a statistically significant and negative $\widehat{\beta_2}$ at 99% or 90% CI as well. Since, Model 4 does not have statistically significant results and neither does the coefficient of HDI2 in the model is negative, the study does not hover around this model much even though the adjust R-squared value is the highest for this model.

On comparing the results of Model 2 with the results of Model 1, we observe the absolute values of $\widehat{\beta_1}$ and $\widehat{\beta_2}$ to be much smaller, and comparatively less statistically significant. Furthermore, the adjusted R-squared value is greater in Model 2, which indicates entity fixed effects to be a better fit of the true model. Model 1 and Model 3 have similar results; their $\widehat{\beta_1}$ and $\widehat{\beta_2}$ as well as adjusted R-squared are very close to each other. Although, the $\widehat{\beta_1}$ and $\widehat{\beta_2}$ value of Model 3 are statistically significant at a higher CI as compared to Model 2, its adjusted R-squared value is not better than Model 2.

3.1.2 Interpretations

- Interpretation of Model 1: On an average, increase in the HDI of a country by 1 unit would lead to an increase in air pollution by 977.5 ug/m3 until the country reaches the turning point of 0.8 units of HDI. After this, the air pollution decreases by 606.15 ug/m3 with every additional unit of HDI2.
- Interpretation of Model 2: On an average, increase in the HDI of a country by 1 unit would lead to an increase in air pollution by 304.3 ug/m3 until the country reaches the turning point of 0.76 units of HDI. After this, the air pollution decreases by 199.2 ug/m3 with every additional unit of HDI2.
- Interpretation of Model 3: On an average, increase in the HDI of a country by 1 unit would lead to an increase in air pollution by 970.6 ug/m3 until the country reaches the turning point of 0.8 units of HDI. After this, the air pollution decreases by 601.5 ug/m3 with every additional unit of HDI2.

To account for the simultaneous causality bias, the study introduces PISA as an instrumental variable. However, since the model is non-linear, there are two main explanatory variables, HDI and HDI2. To avoid the problem of under-identification, which arises if the number of instruments (1) is smaller than the number of main regressors (2), another instrument is introduced. According to Wooldridge[3], instead of introducing another exogenous variable, the square of predicted HDI (*HDI_hat2*) can be used as an instrumental variable.

Table 3.2 IV Method

VARIABLES	(1) First stage	(2) Second stage
PISA	-0.0164*	
	(0.00868)	
HDI_hat2	4.070**	
	(1.879)	
Indus_prod	0.000846	0.0438
	(0.000586)	(0.0568)
Pop_growth	0.0209***	0.825
	(0.00491)	(1.528)
HDI		136.3
		(875.4)
HDI2		-115.0
		(517.1)
Constant	5.851**	-21.05
	(2.909)	(372.0)
Observations	107	107
Adjusted R-squared		0.311

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.2 displays the results after using 2SLS estimation. The First stage displays the regression of main regressors (*HDI and HDI2*) on instruments (*PISA and HDI_hat2*). The Second stage shows the regression of main independent variable (*PM25*) on the main regressors of the model.

The **F-statistic obtained is 13.81, which is greater than 10** indicating the instrumental variables to be strong. Hence, we can use these variables as IV in this model. As expected, the values of $\widehat{\beta_1}$ and $\widehat{\beta_2}$, in table 3.2 are positive and negative respectively, just like in the previous models. However, they are not statistically significant. I believe, this is due to prevalence of omitted variable bias which has not been adjusted for satisfactorily.

3.2.1 Comparing models

The Model 2 (table 3.1), that uses entity fixed effects give the closest estimates with those obtained using 2SLS (table 3.2). The value of $\widehat{\beta_1}$ of Model 2 too were not statistically significant. There is a much greater need to control for variables in the error that are correlate with HDI and HDI2, and are a determinate for PM25.

3.1.2 Interpretations

On an average, increase in the HDI of a country by 1 unit would lead to an increase in air pollution by 136.3 ug/m3 until the country reaches the turning point of 0.6 units of HDI. After this, the air pollution decreases by 115 ug/m3 with every additional unit of HDI2.

4. Conclusion and discussion

The methods used in this study were multiple regression, with and without entity, and time fixed effects. Further, to account for simultaneous causality bias, the study introduced PISA as an instrumental variable and used 2SLS estimation to conduct regression analysis. The results obtained from Model 4, in table 3.1, using entity and time fixed effects were not statistically significant. Model 1 and Model 2, in table 3.1, give highly statistically significant estimates revealing a decline in air pollution with every additional unit of HDI until the country reaches the turning point of 0.8 and 0.76 units of HDI. After this point, air pollution would decline with every additional unit of HDI.

The F-statistic obtained from using 2SLS is 13.81>10. This indicates the instrumental variables to be strong enough to be used in the study. Unlike the discovery in last paragraph, the IV estimates appear to be most similar to Model 2, instead of Model 1 or Model 3 with statistically significant coefficients, although insignificant. Since the relationship between air pollution and HDI obtained from all models is same, the estimates and their significance level are the only difference, I believe, Model 2 and IV model experience omitted variable bias (OVB).

Limitations and future possibilities: The future prospects would be to improve upon the IV model by introducing two separate exogenous variables, instead of using the squared value of predicted main regressor for one instrument, and including more covariates in the model as my study does not completely account for OVB. Also, I believe an even better IV would be FIFA rankings that can be used in my future work as PISA is not a random variable. According to Roberto Ga´squez et al, FIFA rankings have been used as a development indicator by many researchers and it has shown statistically significant results. Furthermore, this study only considers OECD countries. By using countries with low income, one could get better results maybe signifying increasing air pollution in the initial years of development and then decreasing air pollution.

References

1. Li X, Xu L. Human development associated with environmental quality in China. PLoS One. 2021
2. Săndică, A.-M.; Dudian, M.; Ștefănescu, A. Air Pollution and Human Development in Europe: A New Index Using Principal Component Analysis. Sustainability 2018
3. Wooldridge (2000), Econometric Analysis of Cross Section and Panel Data, section 9.5, page. 236-237.
4. R. Ga´squez. Is Football an Indicator of Development at the International Level? 2013
5. PM2.5: https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database
6. HDI: https://hdr.undp.org/data-center/documentation-and-downloads
7. PISA: https://pisadataexplorer.oecd.org/ide/idepisa/
8. Industrial production: https://data.oecd.org/industry/industrial-production.htm
9. Population growth: https://databank.worldbank.org/reports.aspx?source=2&series=SP.POP.GROW&country=#