Air pollution and the economic impact

Math 107 - Lie Detector: Introduction to the Practice of Statistics.

Huda Feroz Ahmed, Syed Bilal Hoda & Maaheen Ahmed Chandna

20^{th} June 2020

1. Introduction:

Air pollution is one of the most major threats and environmental risks in the 21st century. The consequences of air pollution on human health are so alarming that air pollution dominates all other major avoidable causes of deaths including tobacco smoking, liquor use, accidents and viruses (Dechezleprêtre, Rivers and Stadler, 2019). According to World Health Organization, 9 out of 10 people globally live in areas with inappropriate levels of air pollution, which causes 7 million deaths a year that is 1 out of 8 deaths (Dechezleprêtre, Rivers and Stadler, 2019). Additionally, air pollution has impacts on the environment, which has consequences on agricultural products, biodiversity, human activities, building structure and materials and even cultural heritage (OECD, 2016). The severe impacts of air pollution have globally led to the introduction of stricter air quality regulations and environmental policies and there is a very vocal and pushing demand for more stringent policies. However, there are arguments over suitable levels of strictness, as environmental regulations are seen as a 'trade-off' between protecting human health and *inflicting* its response of the economic development (Morgenstern, Pizer, and Shih, 2002) as resources are diverted away from increasing economic output towards producing healthier environments. The purpose of this paper is to investigate this debate that air pollution decreases economic development globally by reducing economic activities.

2. Background:

Literature on cost analysis of air pollutions recognizes that poor air quality may cause direct reductions in economic activity because it negatively impacts cognitive or physical ability. The journal, the impact of pollution on worker productivity, through evidence and examples reveal that exposure to air pollution not only reduces productivity of workers on a farm or factory, but also of high-skilled labor and students (Graff Zivin and Neidell, 2018). Air pollution can majorly affect economy through reduction in the size of working population which is caused by deaths and migration, through reduction in working hours of working population which is caused by health issues in workers or their family, by reduction in productivity of workers, and by decrease in quality of natural capital sources such as agriculture, fish and poultry.

Modern cohort studies confirm that mortality is the most important effect of ambient air pollution and air pollution increases death rates. Air pollution caused 8.8 million premature deaths worldwide in 2015 which corresponded to an average reduction in life expectancy per capita of 2.9 years (Graff Zivin and Neidell, 2018). Researchers have found that, despite significant expense of relocation and government restrictions on migration, large numbers of people in China have been moving away when air pollution increases (Close and Phaneuf, 2017). It was found that if a country experiences 10% increase in pollution in a year, it would also experience a 2.7% reduction in its population (Close and Phaneuf, 2017).

Other than its impact on overall population, air pollution also increases sickness and as a consequence absenteeism. Studies have revealed that prevalence of pollution not only increases school absenteeism, but also absenteeism from work (Dechezleprêtre, Rivers and Stadler, 2019). Dechezleprêtre also explains in his work that absenteeism from work, may not just be direct but sickness in household, such as in children, is also a key factor in work absenteeism. In addition to causing health issues, air pollution also reduces productivity. Various focus group studies have confirmed that air pollution impacts the amount of work which is handled by workers such as count of crop harvested by workers in a farm in California and number of clothes tailored by workers in Bangladesh (Graff Zivin and Neidell, 2018). Furthermore, air pollution also negatively impacts land and water resources such as direct damage to agricultural and forestry output, which reduces productivity of natural resources (Dechezleprêtre, Rivers and Stadler, 2019).

The discussed literature clearly recognizes that air pollution impacts productive outputs from populations worldwide. The aim of this paper is to study the impact of air pollution on economic performance in light of the discussed background.

3. Research question and hypothesis:

The paper aims to answer the following questions:

1) What is the relationship between air pollution and value added by agriculture, forestry and fishing in the economic development globally?
2) What is the relationship between air pollution and expenditure of health in view of economic development globally?
3) What is the relationship between air pollution and economic development globally?

In view of our established background, we can hypothesize that:

1) IAir pollution negatively impacts the value added by agriculture, forestry and fishing in the economic development.
2) Air pollution increases health expenditures of total economy of a country.
3) Air pollution impacts economic activities negatively and reduces economic development

4. Method:

4.1 Explaining Data:

Our data consists of 2 main components, air pollution variables and economic development variables for the year 2017.

1. Location: name of the countries, which acts as primary key for our analysis. We have 142 countries.

Air pollution:

2. PM2. 5: is a continuous variable which refers to outdoor fine particulate matter (PM) that has a diameter of less than 2.5 micrometers. The variable has the unit micrograms per cubic meter (μg/m3). For our analysis, we used median PM2.5 values for each country, as some countries might have outlier values which does not make mean a suitable measure.
3. Ozone: is a continuous variable which refers to a colorless and highly irritating gas that forms just above the earth's surface. The variables have the unit parts per billion (ppb).

For our analysis, we used median ozone values for each country, as some countries might have outlier values which does not make mean a suitable measure.

4. Population proportion exposed to HAP (house air pollution): is the proportion of people exposed to house air pollution with respect to the total population of the country. The variables has values between 0 and 1 inclusive. Here HAP is the contamination of indoor air caused due to solid fuels.

Economic Development:

5. GDP (Gross Domestic Product): is a continuous variable which refers to the monetary value of all finished goods and services made within a country during a specific period. The variable is represented in US dollar terms.
6. Agriculture, forestry, and fishing, value added: is a continuous variable which refers to the contribution of agriculture, forestry, and fishing in GDP, in terms of US dollars.
7. Current health expenditure per capita: is a continuous variable which refers to the amount that each country spends on health per person, in terms of US dollars.

Data was taken from publicly available datasets: Data for air pollution was taken from State of Global air, and data for economic development was taken from World Bank Open Data.

4.2 Data Distributions:

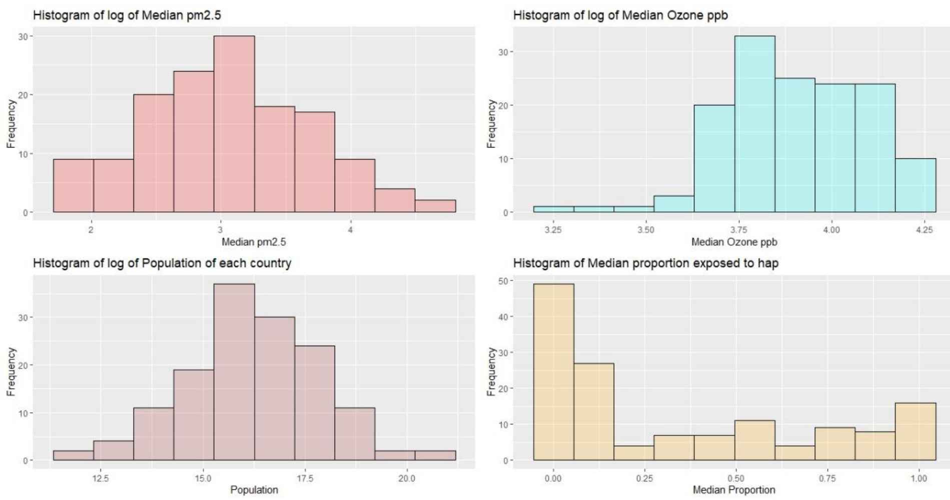


Figure 1 – histograms for air-pollutants: i) PM2.5 ii) ozone iii) population iv) population proportion exposed to HAP

In order to summarize data and to find out its measures of spread and center we made histograms for each variable. Figure 1 shows distributions of air-pollution variables. Figure 1(i) shows that PM2.5 follows normal distribution which is unimodal. The center lies around 3 and the spread shows broad distribution. Figure 1(ii) shows that the data is skewed left which means that the mean is higher than its median with its data spread towards the right. Figure 1(iii) shows the log of population of each country has a normal distribution, with the mean at the center and data spread broadly. Lastly, figure 1(iv) shows the distribution of proportion of population exposed to

HAP follows a distribution that is skewed to the right, indicating that the mean is lower than its median.

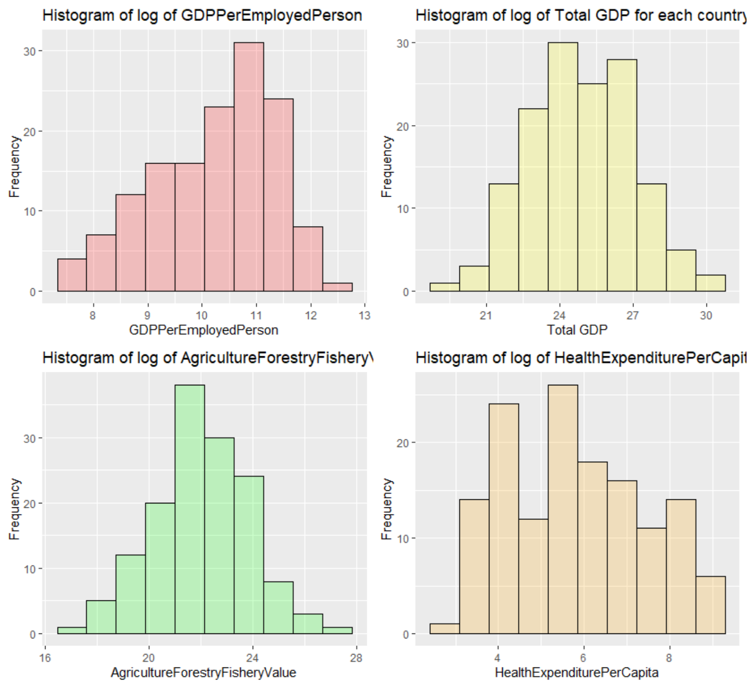


Figure 2 – histograms for economic development

Figure 2 shows us the distribution of variables focusing on economic development. The top right histogram that shows the distribution of total GDP for each country shows normally distributed data with the mean almost around the center. The third histogram shows the distribution of Agriculture, Forestry and Fishing value and shows us that the data again shows normal distribution with the mean at the center. The distribution however seems narrower. Finally, the bottom right histogram shows health expenditure per capita which is skewed to the right and shows us that the mean is lower than the median.

The boxplots in figures 3 summarize data too. Figure 3(i) shows pollution variables against GDP. The range for all pollution variables is same and there are no outliers. Figure 3(ii) shows us three boxplots demonstrating PM2.5, ozone and population proportion exposed to HAP against agriculture, forestry and fishing (AFF) values. All three of them show the same range with no outliers. Figure 3(iii) shows us three different boxplots that summarize PM2.5, ozone, and population proportion exposed to HAP against health expenditure. All three boxplots show the same range without any outliers and have the same shape, center and distribution.

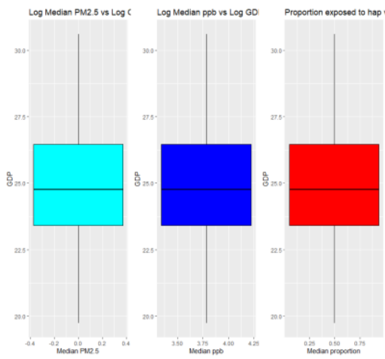


Figure 3(i) - air-pollution vs GDP

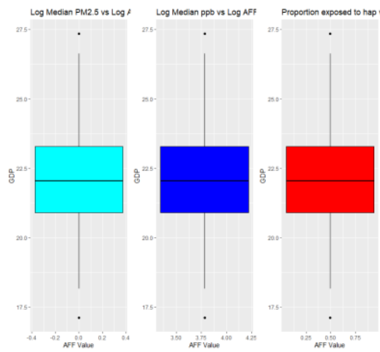


Figure 3(ii)- air-pollution variables vs agriculture, forestry and fishing values

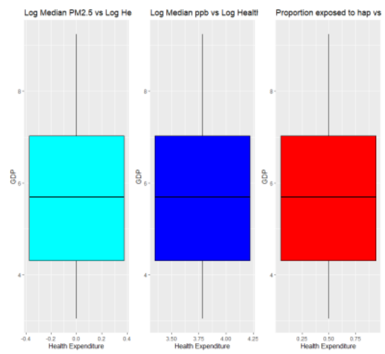


Figure 3(iii)- air-pollution vs health expenditure per capita

Figure 3 – boxplots of air pollutant against economic development

4.3 Analysis:

4.3.1 Regression Analysis:

As visible in our data distributions, the data is highly skewed. To counter this, we log transform our data, except for the variable proportion of people exposed to HAP because it was already between 0 and 1. However, all results were unlogged for interpreations. Our objective is to capture the effect of pollution on economic activity. A simple regression of economic outcomes on pollution can help us investigate the impact we seek. We study three different set of relationships. Firstly, the relationship between GDP and air- pollution variables. Secondly, the relationship between agriculture, forestry, and fishing value and air-pollution variables. Thirdly, the relationship between health expenditure and air-pollution variables.

The forward linear regression between GDP and air-pollutant variables after unlogging the log-log model yields 37.405 as alpha value. For every 1% increase in PM2.5, there is 0.58% decrease in GDP. For every 1% increase in ozone, there is a 3.5% increase in GDP. For every 1% increase in population proportion exposed to HAP, there is an increase of 0.1% in GDP.

The linear regression between agriculture, forestry, and fishing value and air-pollutant variables after unlogging the log-log model yields 40.07 as alpha value. For every 1% increase in PM2.5, there is 0.27 % decrease in agriculture, forestry, and fishing value. For every 1% increase in ozone, there is a 2.09% increase in agriculture, forestry, and fishing value. For every 1% increase in population proportion exposed to HAP, there is an increase of 0.91% in agriculture, forestry, and fishing value.

The linear regression between health expenditure and air-pollutant variables after unlogging the log-log model yields 9.01 as alpha value. For every 1% increase in PM2.5, there is a 1.02 % decrease in health expenditure. For every 1% increase in ozone, there is a 1.69% increase in health expenditure value. For every 1% increase in population proportion exposed to HAP, there is an increase of 0.06% in health expenditure.

We used a level of 0.05 to check for significance. We used the dredge function for model selection where the model with the lowest AIC value is the best model. We predicted our outcomes for the best models. To evaluate our model we used the following plots: residual vs fitted plots to validate for non-linearity, scale vs location plot to validate for homoscedasticity, residuals vs leverage to check for outliers and normal quantile-quantile plot to validate for normal distribution.

4.3.2 Principal Component Analysis:

Since our data has multiple variables for air pollution and economic activity, there arises a need for multivariate analysis. To do so, we checked that can we apply unimodal methods or linear methods by using detrended correlational analysis. Since our value of the longest gradient is not greater than or near 4, we use linear methods. The linear model that we use here is principal component analysis. As our variables are on different scales, i.e. all variables except population proportion exposed to HAP are log transformed; therefore, we use PCA based on a correlation matrix. PCA represents the major features of data along a reduced number of axes; therefore, to select components of importance we used Guttman-Kaiser and broken stick model. We compared a cluster analysis and an ordination to explain or confirm the differences between groups of sites. We used two ways for combining these results, the first differentiates clusters of sites by colors on the ordination plot and the second overlays a dendrogram on the plot. We clustered the objects using the air-pollution data, by using Euclidean distance after standardizing the variables, followed by Ward clustering.

5. Results:

5.1 Linear Regression:

5.1.1 GDP and Air-pollutants:

For the linear regression between GDP and air-pollutants ozone and population proportion exposed to HAP are significant predictors of the response variable (GDP), while PM2. 5 does not seem significant. Table.1 below summarized the p-values.

Model selection showed that ozone and population proportion exposed to HAP have the lowest AIC value, which means it was a relatively better model. The linear regression of ozone with GDP, as shown in figure 4, shows that as median ozone parts per billion increases, GDP also increases. The linear regression of population proportion exposed to GDP with GDP, as shown in figure 5, shows that as greater population proportion is exposed to HAP global GDP decreases.

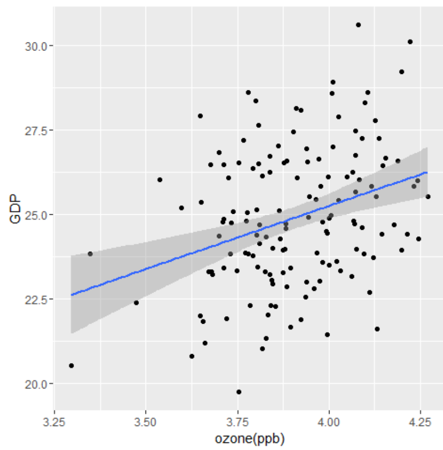


Figure 4 – Regression prediction of ozone vs GDP

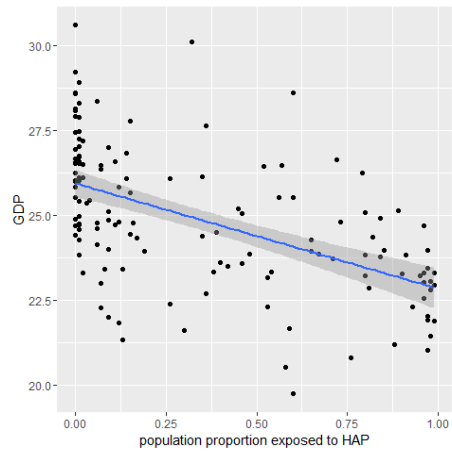


Figure 5 – Regression prediction of HAP vs GDP

We can see from figure 6(i) that residual vs fitted model seems to be random there is no increasing or decreasing pattern. Similarly, the scale location plot as seen in figure 6(ii) seems to be random hence homoscedasticity holds. Figure 6(iii) i.e. residuals vs leverage plot, shows there are no outliers. Finally, figure 6(iv), the normal q-q plot, shows a very slight deviation from normality but overall, it fits the line.

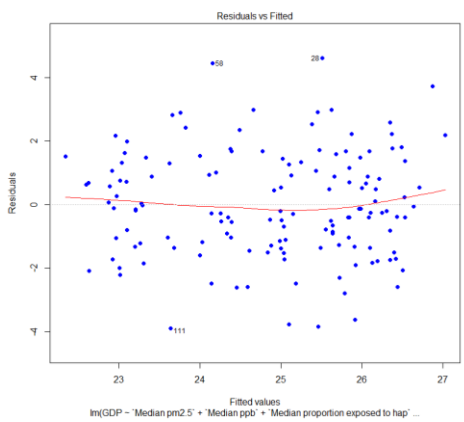


Figure 6(i) - residual vs fitted plot

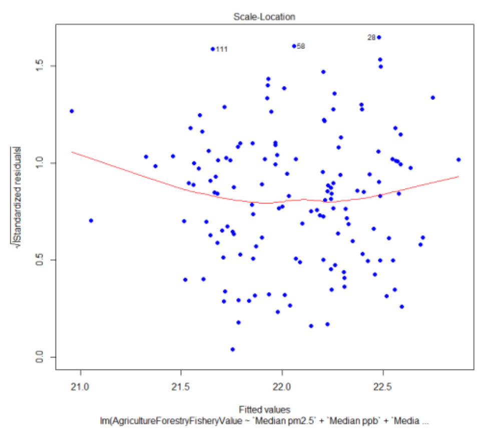


Figure 6(ii) – scale vs location plot

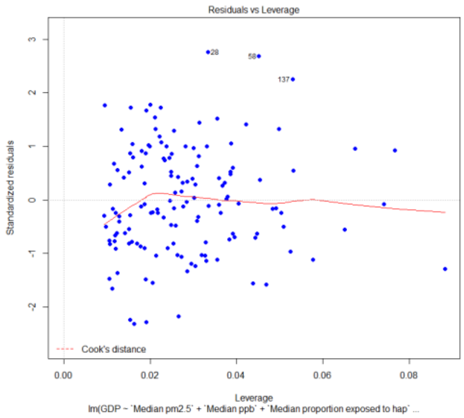


Figure 6(iii) - residuals vs leverage plot

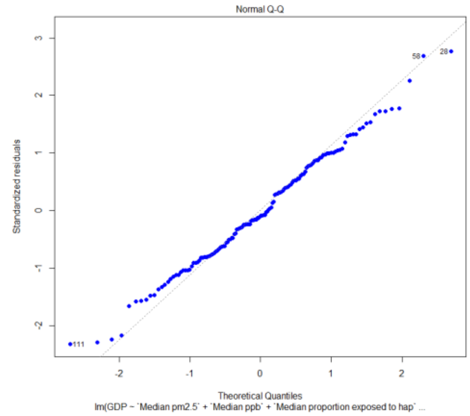


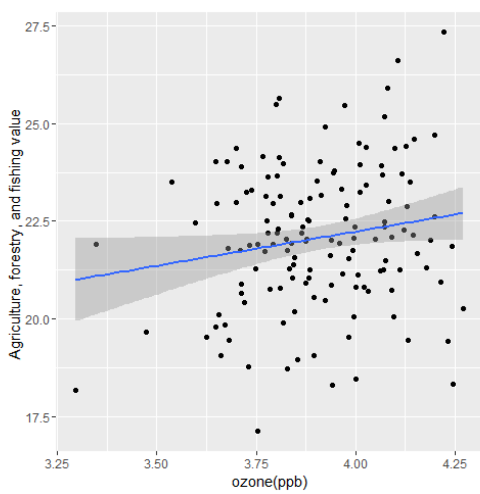
Figure 6(iv) - normal q-q plot

Figure 6 – GDP vs air pollutants summary plots

<u>5.1.2 Agriculture, forestry, and fishing value and Air-pollutants:</u>

For the linear regression between agriculture, forestry, and fishing value and air-pollutants, ozone is very close to 0.05 significance level with p-value 0.052037 so we cannot say with certainty that if it is significant or not. Population proportion exposed to HAP and PM2.5 are not significant. Table.1 below summarized the p-values.

Model selection showed that ozone and population proportion exposed to HAP have the lowest AIC value, which means it was a relatively better model. The linear regression of ozone with Agriculture, forestry, and fishing value, as shown in figure 7, shows that as median ozone parts per billion increases, the contribution of agriculture, forestry, and fishing in GDP also increases. The linear regression of population proportion exposed to HAP with GDP, as shown in figure 8, shows that the contribution of agriculture, forestry, and fishing in GDP remains almost constant with the increase in population proportion is exposed to HAP.



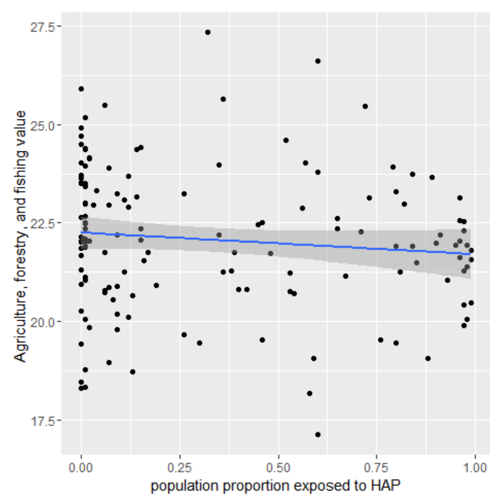


Figure 7 – Regression prediction of ozone vs agriculture, forestry, and fishing value

Figure 8 – Regression prediction of ozone vs agriculture, forestry, and fishing value

We can see from figure 9(i) that residual vs fitted model there are no increasing or decreasing patterns. Similarly, the scale location plot as seen in figure 9(ii) seems to be random hence homoscedasticity holds. Figure 9(iii) i.e. residuals vs leverage plot, shows there are no outliers. Finally, figure 9(iv), the normal q-q plot, shows that the distribution is normal.

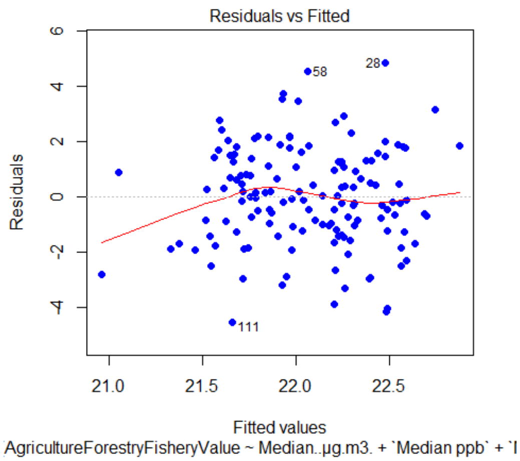


Figure 9(i) - residual vs fitted plot

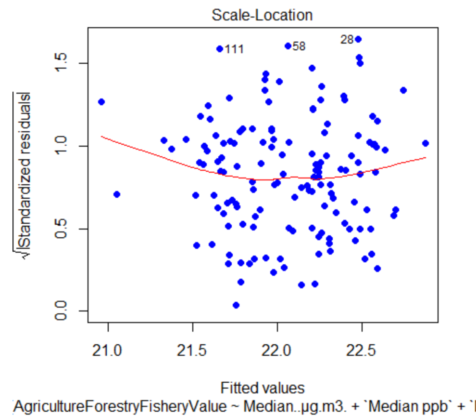


Figure 9(ii) – scale vs location plot

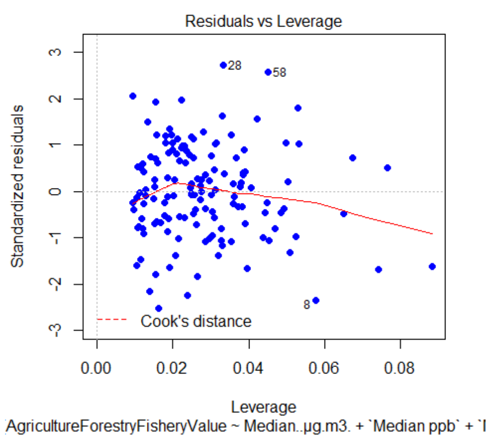


Figure 9(iii) - residuals vs leverage plot

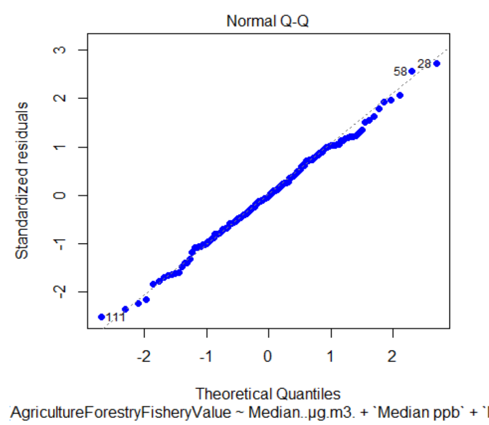


Figure 9(iv) - normal q-q plot

Figure 9 – Agriculture, forestry, and fishing value vs air pollutants summary plots

5.1.3 Health Expenditure and Air-pollutants:

For the linear regression between health expenditure and air-pollutants, all explanatory variables i.e. PM2. 5, ozone and population proportion exposed to HAP are significant at 0.05 level.

Table.1 below summarized the p-values. Model selection showed that all three variables i.e. PM2.5, ozone and population proportion exposed to HAP have the lowest AIC value, which means it was a relatively better model. The linear regression of ozone with health expenditure, as shown in figure 10, shows that as median ozone parts per billion increases, the amount that countries spends on health per person also increases. The linear regression of population proportion exposed to HAP with health expenditure, as shown in figure 11, shows that health expenditure per capita globally decreases very sharply with the increase in population proportion is exposed to HAP. The linear regression of PM2.5 with health expenditure shows, as seen in figure 12, that with the increase in PM2.5 the amount that countries spends on health per person also decreases drastically.

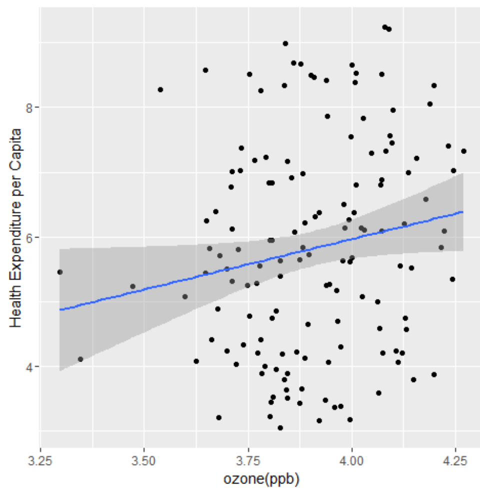


Figure 10 – Regression prediction of ozone vs health expenditure per capita

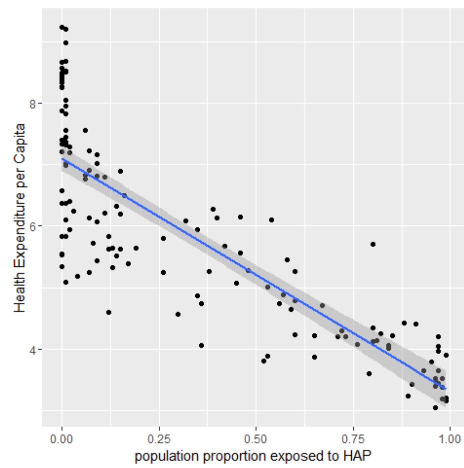


Figure 11 – Regression prediction of ozone vs health expenditure per capita

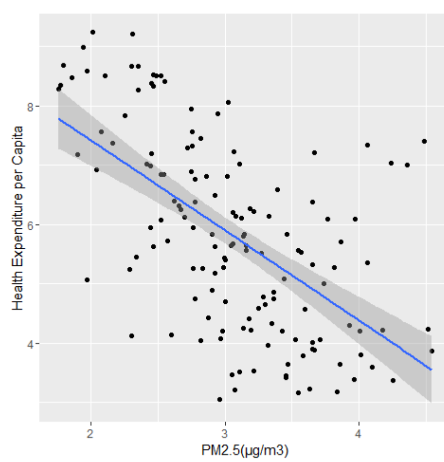


Figure 12 – Regression prediction of PM2.5 vs health expenditure per capita

We can see from figure 13(i) that residual vs fitted model there are no increasing or decreasing patterns. However, there is a slight increase in values towards the end. Similarly, the scale location plot as seen in figure 13(ii) seems to be random hence homoscedasticity holds. Figure 13(iii) i.e. residuals vs leverage plot, shows there are no outliers. Finally, figure 13(iv), the normal q-q plot, shows that the distribution is normal.

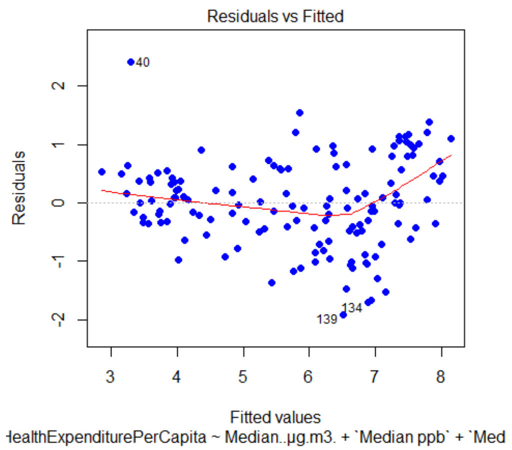


Figure 13(i) - residual vs fitted plot

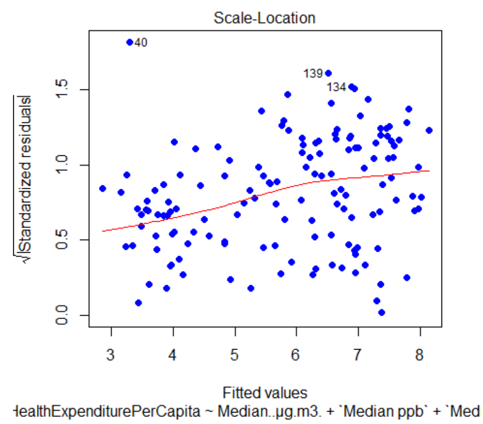


Figure 13(ii) – scale vs location plot

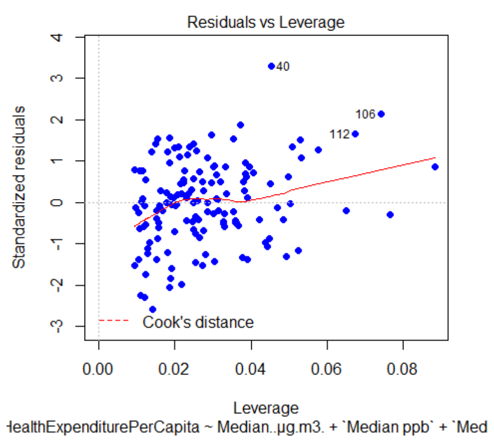


Figure 13(iii) - residuals vs leverage plot

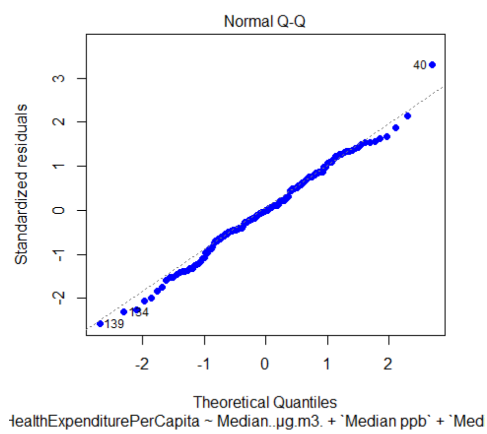


Figure 13(iv) - normal q-q plot

Figure 13 – Health expenditure vs air pollutants summary plots

Explanatory variables	GDP and air-pollutant	Agriculture, forestry, and fishing value and air-pollutant	Health expenditure and air-pollutant
PM2. 5	0.059547	0.407481	4.38e-12
ozone	0.000625	0.052037	0.000192

population proportion exposed to HAP	1.65e-05	0.855923	< 2e-16

Table 1 - significance p-values for regression analysis.

5.1 Principal Component Analysis:

We obtained 7 principal components, where 92.96% variation is explained by the first 3 components. Variation across all variables can be seen in Table 2.

	PC.1	PC.2	PC.3	PC.4	PC.5	PC.6	PC.7
Eigen Value	3.0577	2.2401	1.2097	0.27483	0.15348	0.053990	0.01022
Proportion of Variance	0.4368	0.3200	0.1728	0.03926	0.02193	0.007713	0.00146

Table 2- Eigen values and variation of PC (principal components)

Guttman-Kaiser criterion and broken stick model both shows, as seen in figure 14, that principal component 1, 2 and 3 are components of importance as they are the axes that are greater than average Eigen-value.

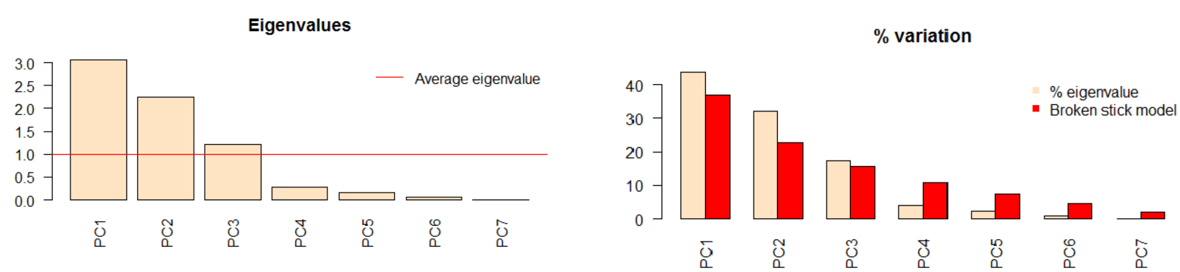


Figure 14 - Guttman-Kaiser criterion & Broken stick model

The plot of principal components PCA 1 and PCA 2 shows, as seen in figure 15 that agriculture forest and fishery value, GDP, health expenditure make a higher contribution as their vectors are longer than circle radius. Being in opposed quadrants, health expenditure is negatively correlated with population exposed to HAP and PM2.5. GDP, agriculture forest and fishery value and ozone are positively co-related.

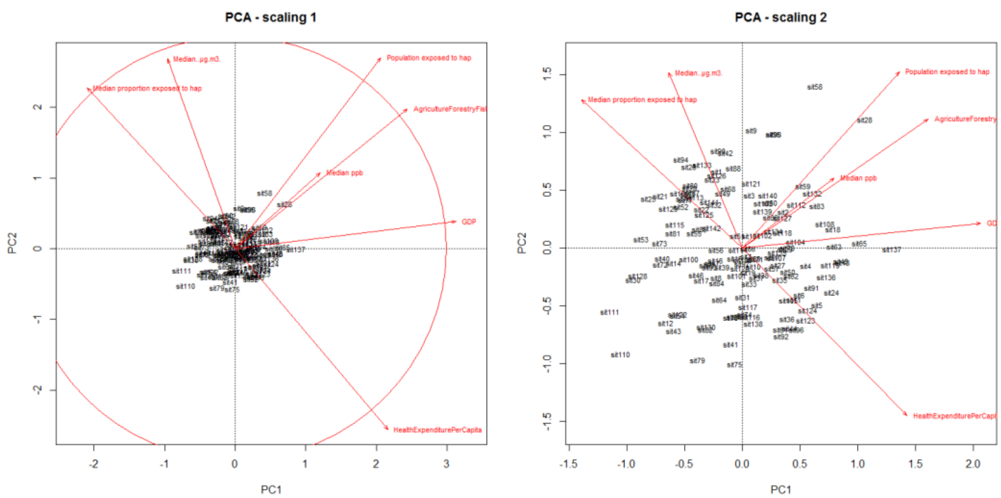


Figure 15 - Plot of principal components 1 and 2 (i) scaling = 1 (ii) scaling = 2

The plot of principal components PCA 2 and PCA 3 shows, as seen in figure 15 that ozone and PM2.5 make a higher contribution as their vectors are longer than circle radius. Being in opposed quadrants, health expenditure is negatively correlated with population exposed to HAP and PM2.5. Agriculture forest and fishery value is positively correlated with population proportion exposed to HAP.

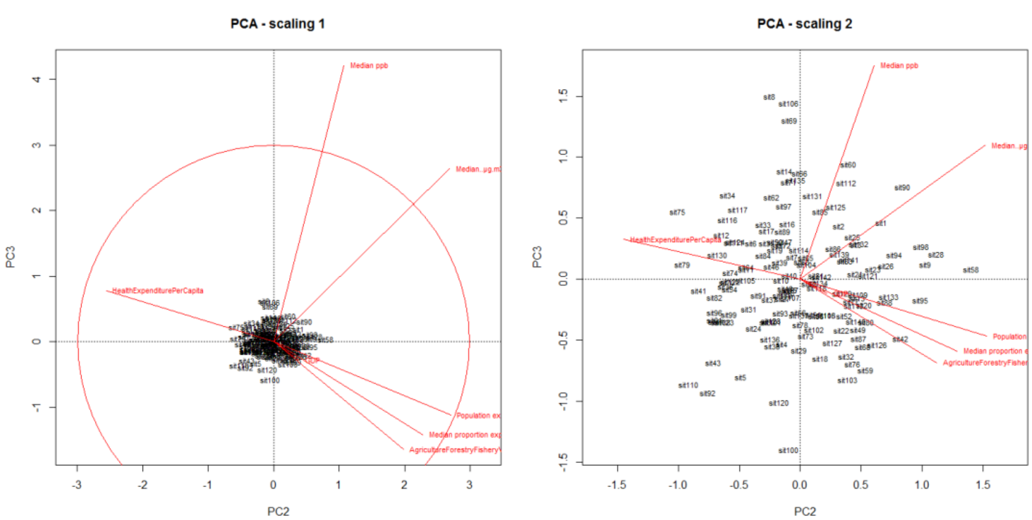


Figure 16 - Plot of principal components 2 and 3 (i) scaling = 1 (ii) scaling = 2

Figure 17 shows the cluster analysis. We can see that there are groups of four clusters are being formed and the points are very close together.

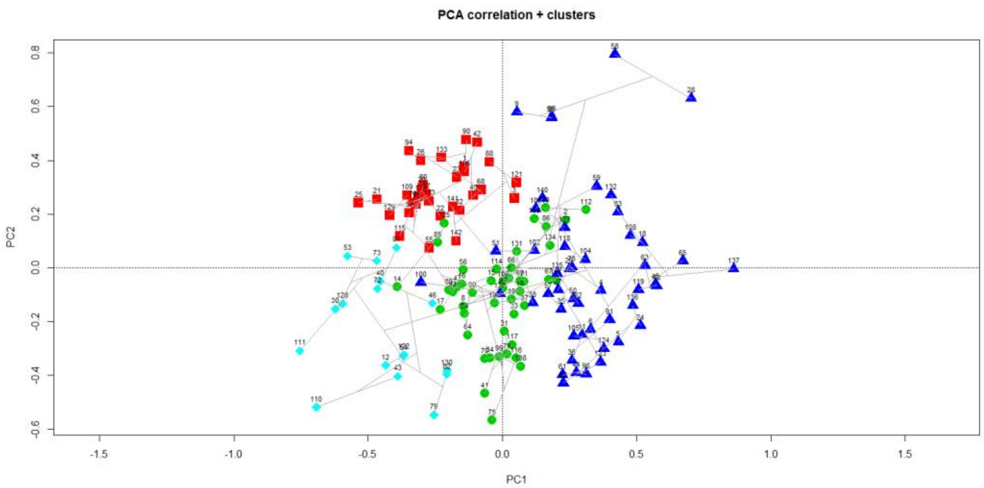


Figure 17 – clustering analysis

<u>6. Discussion:</u>

Evidence on the relationship between air pollution and economic development is mixed and not very clear. Our hypothesis that air pollution negatively impacts the value added by agriculture, forestry and fishing globally in the economic development does not hold as value added by agriculture, forestry and fishing increases with increases in ozone and remains same (with slight decrease) with increase in increase in population proportion exposed to house air pollution. Our hypothesis that air pollution results in increased health expenditures of total economy of a country does not hold completely as with the increase of PM2.5 and population proportion exposed to house air pollution, the per-capita amount spent on health expenditure drastically decreases. However, the hypothesis hold only if the considered air-pollutant is ozone. Our hypothesis that air pollution impacts economic activities negatively and reduces economic development also does not completely hold as while the increase in population proportion exposed to house air pollution has negative impacts of global GDP, increase in ozone is found to increase global GDP. Our analysis shows that population exposed to HAP, agriculture forest and fishery value, GDP, and health expenditure are much more significant therefore focusing on these variables only may have yield better results. Our principal component analysis shows the presence of clusters which means that there is a presence of varying correlations between clustered groups of different countries.

Our investigations inform that the debate that stringent policies on environmental control are a trade-off between protecting human health and environment, and between economic developments do not completely hold. We can also say that the economic benefits of emissions reductions are not completely related; therefore, stronger air quality regulations may or may not warranty economic benefits. Our data is focused on 142 countries only so we were not able to provide complete geographic coverage. Above all, our analysis is based on a particular year, lacking temporal coverage. This means that any external economic or environmental impacts faced by a country in 2017 will impact and bias the results.

There are several reasons to suspect that the available data and evidence does not correctly either affirm or negate the hypothesis. Firstly, estimating the causal effect of air pollution on economic outcomes at an aggregate level is challenging because of the potential for reverse causality. Not only might air pollution impact economic output and productivity (the effects we investigates), but economic activity clearly also affects pollution emissions through a number of potential channels. Therefore, there is a lack of control variables. Moreover, we focus on 3 most common air-pollutants, some of these air pollutants may be typically correlated with one another. Some ambient air pollutants share many sources in common – in particular they may be released by a common product of combustion or activity. Furthermore, there is a strong direct relationship between temperature and air pollution (Kalisa, Fadlallah, Amani and Nahayo, 2018), but we assessed all countries, which have different climates (and temperatures) without any control variables which contributed to bias.

7. References:

Close, B., & Phaneuf, D. J. (2017). Valuation of Local Public Goods: Migration as Revealed Preference for Place.

Dechezleprêtre, A., Rivers, N., & Stadler, B. (2019). The economic cost of air pollution: Evidence from Europe.

Graff Zivin, J., & Neidell, M. (2012). The impact of pollution on worker productivity. *American Economic Review*, 102(7), 3652-73.

Kalisa, E., Fadlallah, S., Amani, M., Nahayo, L., & Habiyaremye, G. (2018). Temperature and air pollution relationship during heatwaves in Birmingham, UK. *Sustainable cities and society*, 43, 111-120.

Morgenstern, R. D., Pizer, W. A., & Shih, J. S. (2002). Jobs versus the environment: an industry-level perspective. *Journal of environmental economics and management*, 43(3), 412-436.

OECD (2016), The Economic Consequences of Outdoor Air Pollution, *OECD Publishing*, Paris. http://dx.doi.org/10.1787/9789264257474-en