SEEDS
Sustainability Environmental Economics and Dynamics Studies

Working Paper Series

The relationship between air quality, wealth, and COVID-19 diffusion and mortality across countries

by

Roberto Antonietti, Paolo Falbo, Fulvio Fontini

08/2020

SEEDS Working Paper 08/2020
May 2020
By Roberto Antonietti, Paolo Falbo, Fulvio Fontini

The relationship between air quality, wealth, and COVID-19 diffusion and mortality across countries

Roberto Antonietti*

Paolo Falbo**

Fulvio Fontini*

* “Marco Fanno” Department of Economics and Management, University of Padova, Italy

** Department of Economics, University of Brescia, Italy

May 2020

Abstract

This study concerns the relationship between economic wealth, air quality and COVID-19 diffusion and mortality around the world. We show that the level of air quality, in terms of particulate (PM 2.5) concentrations, does not significantly contribute to explaining the diffusion of COVID-19 and the related mortality after accounting for socioeconomic factors, especially per capita GDP. This latter variable significantly correlates with the diffusion of COVID-10 and related mortality, and the result holds for different times when COVID-19 infections and deaths are counted. When we cluster countries by level of wealth, economic openness, macroeconomic structure, CO2 emissions, and climate conditions, we find that higher concentrations of PM 2.5 coincide with more infections and deaths, but only holds in high-income countries.

Keywords: COVID-19, pollution, PM2.5, wealth, cross-country analysis

J.E.L. Classification Codes: I10, Q50, Q53

1. Introduction

At the end of 2019 an unknown virus spread throughout Wuhan city and Hubei province in China, causing a disease with severe respiratory symptoms that resembled SARS, and had fatal consequences in numerous cases. The virus was later identified as belonging to the Corona virus family and termed SARS-CoV-2, or COVID-19. At the end of January 2020, a Public Health Emergency of International Concern was declared as the virus rapidly spread around the world. On March 11th, the World Health Organization officially declared the COVID-19 emergency a pandemic. By April 21st, 178 countries had confirmed cases of COVID-19 infection.[1] The total count of reported cases and casualties was still rising at the time of writing this paper.

The pandemic is having a huge social and economic impact. Several countries have adopted more or less stringent emergency measures in an effort to contain the diffusion of the virus and its consequences. These measures include social distancing and lockdowns, which have severely limited industrial, commercial and transportation activities. There have been huge losses on stock markets: S&P fell by 30% in the month from 20 February to 19 March as the epidemic continued to spread, and is still 12% below its quotation of the start of the year. Much the same happened to the Nikkei 225 (30% and 18%, respectively), and Euro STOXX500 (38% and 22%, respectively). Even commodity prices crashed. As an example, the price of oil (WTI spot, quoted at NYMEX) dropped in two months from 53.7$ a barrel to a nil, and its May futures price reached a record-breaking negative value of -38$ a barrel on April 20th. Current GDP estimates indicate a fall in economic production for several developed countries, such as the USA (-4.8%, source: BEA[2]) or the Euro area (-5.5%, source: ECB[3]).

On the other hand, lockdown measures are having positive effects on the environment in general, such as better air and water quality, less pollution, and a lower anthropic pressure on several animal species (EEA, 2020). The relationship between the environment and the COVID-19 pandemic has also attracted attention because it was notable that the areas being hardest hit by the virus were also among the most polluted of the planet. Wuhan and the province of Hubei, where the outbreak began, the Lombardy region in Italy and the Madrid area in Spain, which have all suffered particularly badly from viral infection, are in regions with a normally very poor air quality.

There is a twofold rationale behind the link identified between air pollution and the COVID-19 pandemic. First, it has been argued that poor air quality correlates with a greater diffusion of COVID-19 because atmospheric conditions favoring the permanence of airborne pollutants, such as particulate matter (PM), would also facilitate the spread of the virus conveyed in the aerosols from human saliva, which seems to be one of the main sources of contagion. PM could serve as a carrier of COVID-19 virus. Second, there may be a relationship between air quality and mortality due to COVID-19 infection because chronic exposure to environmental pollution in general, and poor air quality in particular, have a debilitating effect on the body, increasing its exposure to other respiratory diseases, and reducing the immune system’s response to infections. All these effects can increase the mortality risk associated with COVID-19.

Research into these aspects is ongoing. At the time of writing, a few published studies seem to confirm the above-mentioned links between air quality and coronavirus diffusion. Xiao et al. (2020) estimated an 8% increase in the COVID-19 death rate associated with a rise of 1 mg/m3 in PM2.5 levels in parts of the US. Ogen (2020) found a positive correlation between NOx exposure and

[1] Data obtained from the World Health Organization website: https://www.who.int

[2] https://www.bea.gov

[3] https://www.ecb.europa.eu/stats/ecb_surveys/survey_of_professional_forecasters/html/table_hist_rgdp.en.html

COVID-19-related mortality in 66 administrative regions in Italy, Spain, France and Germany. Setti et al. (2020) found evidence of COVID-19 on outdoor PM in samples tested in the province of Bergamo (Lombardy, Italy), which experienced the highest diffusion and mortality rates in Italy (and among the highest worldwide).

We suspect, however, that the evidence gained to date might not be telling the whole story about the link between air quality and the diffusion of COVID-19 and associated mortality. Hence our study to check whether the macro-economic structure of countries has a part to play in explaining infection and death rates relating to COVID-19, as well as more direct factors like average air pollution. The potential effects of socio-economic factors might be extremely diversified, and more or less readily identifiable. It is common knowledge that wealthier countries are more likely to have populations with a higher proportion of elderly citizens, who would be more at risk, and more severely affected by the pandemic. Countries with significant manufacturing sectors (where people are less able to work remotely) can have complex supply chains characterized by a greater degree of social proximity than supply chains for services, for instance. Such countries may also come under greater pressure from industrial lobbies to limit or to delay policies to prevent the spread of the contagion. The moment when the virus started to spread in a given country was probably rather random, but there were more opportunities for it to do so in countries with strong and frequent international trade links. Countries' different levels of general wealth is also associated with more or less developed health care systems, in terms of facilities, personnel, and organization. Wealthier countries probably had a better chance of taking care of infected people and testing larger proportions of the population for contagion. This last aspect might also be a factor introducing a significant measurement bias in the way COVID-19-related hospitalizations and deaths have been counted.

At the same time, economic wealth interacts with air pollution levels. We know that a combination of less efficient production and transport systems, particularly in less developed countries, and a lower quality of energy consumption coincide with high environmental externalities (Sovacol, 2012). The cross-country data we use here confirm as much, showing a negative relationship between real per capita GDP and air pollution, as measured from the concentrations of small particulate (PM 2.5).

The role of agriculture needs to be considered as well, not just for its contribution to GDP, but also because of how it relates to air pollution. At first glance, countries based largely on agriculture might be expected to be less exposed to air pollution, but high-tech and intensive animal breeding is associated with the extensive use of manure for fertilization, which is in turn associated with large particulate formation. Our analysis on the COVID-19 pandemic included both economic and environmental factors, so their interactions were tested too.

We have no intention of conducting an epidemiological analysis, in the sense of assessing how and why any of the previously-mentioned factors favored or hindered the spread of COVID-19. What we do consider well worth assessing is the body of data and evidence concerning the possible influence of socio-economic and environmental factors with a view to improving our understanding of how the COVID-19 pandemic came about.

To develop our analysis, we merge data on COVID-19 infections and deaths provided by the European Centre for Disease Prevention and Control (ECDC) with country-level macro-economic data collected by the World Development Indicators of the World Bank group. We consider data for 142 countries. We first run a cross-sectional regression of two dependent variables, namely COVID-19 infections and deaths, on the populations of countries to check for the main effects of several macro-economic variables (economic structure), average exposure to PM2.5 and population size. Then, to disentangle the interaction between exposure to air pollution and economic structure, we group the countries by economic similarity and test the impact of air pollution on COVID-19 infections and mortality within each group.

The paper is arranged as follows: Section 2 presents the data and the variables used in the empirical analysis, and their main descriptive statistics; Section 3 outlines the econometric strategy; Section 4 describes and discusses the main findings; Section 5 focuses on the cluster analysis and the results of the corresponding estimations; and Section 6 concludes. Additional estimates and robustness tests are provided in the Appendix.

2. Data

To build the dataset we merge information from two sources. Data on COVID-19 infections (variable: INFECTIONS) and deaths (variable: DEATHS) are used as the dependent variables. They are drawn from the ECDC, an EU agency for the protection of European citizens against infectious diseases and pandemics. The data on the distribution of COVID-19 worldwide are updated on a daily basis by the ECDC's Epidemic Intelligence team, based on reports provided by national health authorities[4].

Data for these two variables were collected for four dates: 24 March, 7 April, 14 April and 21 April 2020. The COVID-19 outbreak did not happen everywhere at once, and national authorities have adopted different strategies and policies to deal with the pandemic. The diffusion of COVID-19 has taken a certain amount of time, incurring far from negligible costs, and countries have reacted with lockdown and other measures in different ways. All this has affected the measurement of the effects of stock variables, and that is why we measure the effect of wealth and pollution on the diffusion and mortality of COVID-19 in different periods, from right after the start of the outbreak up until the moment when the strictest lockdown measurers started to be lifted in the European countries hit earliest and most severely (Italy and Spain). At the beginning of our observation period, the relationship might have been influenced by the different pace at which COVID-19 was spreading around the globe. By the end of April 2020, lockdown measures were having an effect on the phenomenon. We nonetheless show stable, significant results across the dates selected, which means that our findings are robust to the timing of the virus's diffusion and to the heterogeneity of the policies adopted.

Data on infections and deaths are merged with macro-economic information provided by the World Development Indicators of the World Bank on:

- PM2.5: mean annual exposure to PM 2.5 (micrograms per cubic meter).
- GDPPC: real per capita GDP (in 2010 US$ at PPP)
- POPULATION: total resident population
- IMPORT/GDP: import intensity (i.e. import value as a proportion of domestic GDP)
- AGRVA/GDP: agriculture value added as a proportion of GDP
- MANVA/GDP: manufacturing value added as a proportion of GDP[5]
- CO2: CO_2 emissions (metric tons per capita),
- TEMP: average temperature in March (in °C).

[4] For more information, see: https://www.ecdc.europa.eu/en/covid-19/data-collection.

[5] We have omitted the share of services as a proportion of GDP (SERV/GDP) as an explanatory variable because it is collinear with AGRVA/GDP and MANVA/GDP.

Since the World Bank provides information on PM2.5 exposure up until 2017, we measure all the explanatory variables in the same year. The CO_2 variable has been included as a measure of the intensity and efficiency with which primary energy sources are used in a country to generate the aggregate output. The final sample consists of a cross-section of 142 countries. Table 1 shows the main summary statistics. Table 2 shows the pairwise correlations among the explanatory variables. Interestingly, we find PM 2.5 and GDPPC strongly, but negatively, correlated.

Table 3 shows the pairwise correlations of our two dependent variables, i.e. the total number of COVID-19 infections, and the total number of COVID-19-related deaths, with PM 2.5 exposure and real per capita GDP on five different days between March and April 2020. Unlike the literature on air pollution and coronavirus diffusion, we find a negative correlation between the two, whereas the correlation between coronavirus (both infections and deaths) and wealth is positive[6]. For the number of deaths, we also note that all correlations become stronger and more significant towards the end of April.

Table 1. Summary statistics

Variable	Mean	Std.dev.	Min	Max
INFECTIONS 24/03	2648.04	10336.8	1	81553
INFECTIONS 14/04	13142.78	54833.8	4	582594
INFECTIONS 21/04	17028.93	72148.47	6	787752
DEATHS 24/03	114.98	625.48	0	6077
DEATHS 14/04	835.27	3378.06	0	23649
DEATHS 21/04	1196.82	4956.5	0	42539
PM2.5	28.43	20.32	5.861	99.73
GDPPC	15661.6	20679.9	370.74	109453
POPULATION (mln)	50.554	165.51	0.072	1386.4
IMPORT/GDP	0.458	0.250	0.116	1.825
AGRVA/GDP	0.099	0.098	0.0003	0.486
MANVA/GDP	0.128	0.063	0.010	0.374
CO_2 per capita	4.954	6.168	0.053	43.86
TEMPERATURE (March, °C)	14.83	11.47	-18.72	30.63

Table 2. Correlation matrix

	1	2	3	4	5	6	7	8
1. PM 2.5	1							
2. GDPPC	-0.344***	1						
3. POPULATION	0.266***	-0.071	1					
4. IMPORT/GDP	-0.187**	0.300***	-0.236***	1				
5. AGRVA/GDP	0.430***	-0.553***	0.040	-0.165**	1			
6. MANVA/GDP	-0.051	0.070	0.209**	-0.003	-0.277***	1		
7. CO_2 p.c.	0.041	0.595***	-0.011	0.144*	-0.519***	0.115	1	
8. TEMP	0.341***	-0.447***	-0.022	-0.170**	0.439***	-0.111	-0.277***	1

*** Significant at 1% level; ** significant at 5% level; * significant at 10% level.

[6] We should stress that the correlation between PM 2.5 and COVID-19 infections or deaths is at country level, or *between* countries. It may be that, *within* countries, there is a higher level of contagion or mortality in regions where air quality is lower.

Table 3. Correlations of COVID-19 infections and related deaths with PM 2.5 and real per capita GDP

	Infections				
	24/03	31/03	07/04	14/04	21/04
PM 2.5	-0.059	-0.132	-0.145*	-0.143*	-0.139*
GDPPC	0.200**	0.275***	0.276***	0.269***	0.261***
	Deaths				
	24/03	31/03	07/04	14/04	21/04
PM 2.5	-0.032	-0.100	-0.148*	-0.168**	-0.170**
GDPPC	0.107	0.177**	0.243***	0.274***	0.280***

*** Significant at 1% level; ** significant at 5% level; * significant at 10% level.

3. Econometric analysis

Our empirical approach aims to test whether the correlations in Table 3 hold in a multivariate setting. To do so, we estimate the following equation:

$$Y_{iT} = \beta_0 + \beta_1 PM2.5_i + \beta_2 POPULATION_i + \mathbf{X}_i'\beta_3 + \epsilon_{iT} \qquad [1]$$

where *i* is the country, *Y* is respectively the total (cumulative) number of COVID-19 infections, and the total (cumulative) number of COVID-19-related deaths, recorded on day *T* (21 April, 14 April, 7 April and 24 March); **X** is the vector of additional macro-economic and climate-related variables (GDPPC, IMPORT/GDP, AGRVA/GDP, MANVA/GDP, CO_2 and TEMP); and ε is the error term.

We proceed as follows: first, we estimate a model with only PM2.5 and POPULATION as regressors to check whether the coefficient of the former is statistically significant. Then, we start adding the real per capita GDP to check whether the estimated coefficient for PM 2.5 loses its statistical significance, revealing a spurious relationship with COVID-19 infections or related deaths. Third, we add the other controls to check for the stability of the previous estimates. Finally, we add a set of five macro-regional controls (one for each continent) to test for the presence of common geographical trends.

Since our dependent variables are the discrete, and non-negative, number of infections or deaths in each country, we cannot estimate Equation 1 using Ordinary Least Squares (OLS). To cope with this issue, we use a count data model, i.e. the negative binomial. This also enables us to solve the problem of data overdispersion that arises when the variance of the observed distribution of the count variable is larger than the mean, as imposed in the Poisson regression model.

To check for potential multicollinearity among regressors, we also estimate Equation 1 using OLS, and we consider the Variance Inflation Factor test, where we look at the mean and the maximum value of the corresponding statistic. As a rule of thumb, multicollinearity becomes an issue for values of the VIF statistic higher than 5.

4. Results

Tables 4 and 5 show the results of the negative binomial estimates of Equation 1. Table 4 refers to the number of COVID-19 infections on two different days, 21 April and 24 March, to test whether the results are robust across one month of the virus's diffusion. Columns 1 and 5 refer to the estimates of Equation 1 where only PM 2.5 and POPULATION are included as regressors. We find that a higher exposure to PM 2.5, i.e. a lower air quality, related to a lower number of infections, the coefficient being slightly higher on 24 March than a month later. This result confirms the negative pairwise correlation between PM 2.5 and the number of infections described in Table 3.

Table 4. Infections. Relationship with pollution and wealth

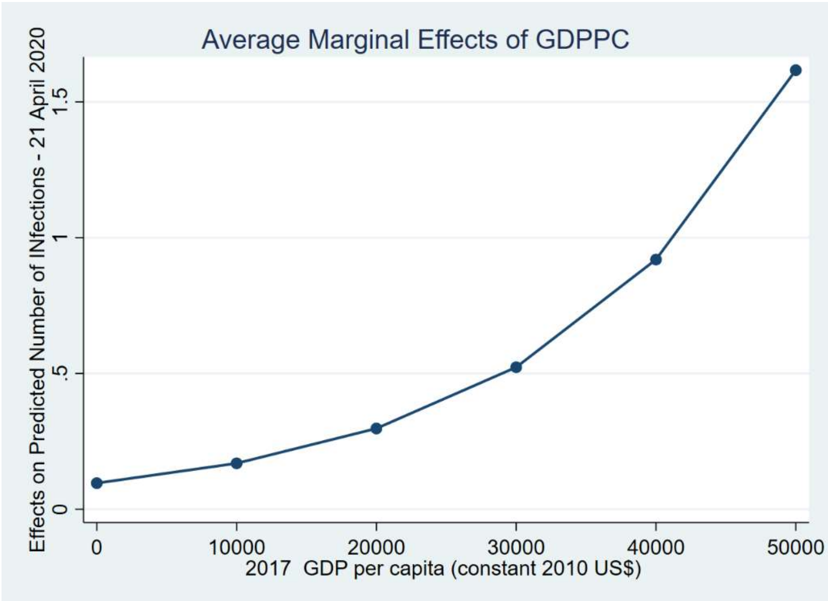
	21 April 2020				24 March 2020			
NEG BIN	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PM 2.5	-0.028***	-0.012	-0.004	-0.004	-0.040***	-0.014	-0.016*	-0.014
	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	(0.013)	(0.010)	(0.011)
GDPPC		0.00006***	0.00005***	0.00005***		0.00008***	0.00005***	0.00006***
		(0.00001)	(0.00001)	(0.00001)		(0.00002)	(0.00002)	(0.00001)
POPULATION	0.011*	0.012	0.004	0.003	0.010	0.016	0.005	0.003*
	(0.006)	(0.008)	(0.004)	(0.004)	(0.010)	(0.016)	(0.007)	(0.002)
IMPORT/GDP			-3.289***	-3.635***			-3.618***	-4.500***
			(0.716)	(0.723)			(0.717)	(0.583)
AGRVA/GDP			-4.734**	-2.989			-5.241**	-5.761**
			(1.966)	(2.117)			(2.097)	(2.207)
MANVA/GDP			5.463**	4.464*			6.893**	3.802
			(2.654)	(2.694)			(3.078)	(2.854)
CO_2			0.016	0.014			0.081	0.024
			(0.048)	(0.046)			(0.097)	(0.070)
TEMP			-0.042***	-0.015			-0.045**	0.010
			(0.013)	(0.017)			(0.019)	(0.020)
Area dummy	NO	NO	NO	YES	NO	NO	NO	YES
N	142	142	142	142	142	142	142	142
Pseudo R^2	0.020	0.038	0.062	0.066	0.022	0.050	0.083	0.100
Wald $\chi 2$	24.02***	36.71***	187.9***	263.9***	29.09***	23.37***	169.4***	311.98***
Alpha	2.941***	2.350***	1.719***	1.618***	3.491***	2.753***	1.976***	1.640***
Max VIF	1.08	1.22	2.21	4.44	1.08	1.22	2.21	4.44
Mean VIF	1.08	1.14	1.65	2.15	1.08	1.14	1.65	2.15
AIC	2695.92	2648.95	2593.84	2589.7	2064.46	2013.67	1954.03	1925.51

Robust standard errors in brackets. Each estimate includes a constant term. *** Significant at 1% level; ** significant at 5% level; * significant at 10% level.

Instead, Columns 2 and 6 show that the estimated coefficient for PM 2.5 is no longer statistically significant when we add GDPPC as a regressor. After controlling for a country's level of wealth, the role of air pollution vanishes, making its relationship with COVID-19 spurious. Columns 3 and 7 confirm this result when further controls are added to GDPPC: we find that, *ceteris paribus,* the number of infections is higher the lower the country's openness to imports, the lower the contribution of agriculture to domestic GDP, and the higher that of manufacturing. As expected, we also find that a lower temperature in March is associated with a higher rate of contagion. These results remain much the same when we further add continent-specific dummies in Columns 4 and 8. The AIC statistics show that this is the specification that fits the data best, while the VIF statistics confirm that multicollinearity is not an issue[7]. Table A1 in the Appendix shows that these results hold when we look at the number of infections on two intermediate days, i.e. 14 and 7 April 2020.

Observing the values in Column 4, we calculated an average marginal effect of GDPPC (given by $\beta_{GDPPC}e^{(\mathbf{x}'\beta_{GDPPC})}$) of 2.038, and a marginal effect at the mean (i.e. 15662 US$) of 0.179. This amounts to a 1790-unit increase in the average number of COVID-19 infections in response to a 10,000 US$ increase over the average GDPPC as at 21 April. Looking at the values a month earlier (in Column 8) we find that the average marginal effect of GDPPC is rather stable. Figure 1 shows that the marginal effect of GDPPC increases exponentially with the level of GDPPC: the marginal effect of the richest countries - those belonging to the 90th percentile of the GDPPC distribution (10.28) - is almost thirteen times larger than that of the poorest (0.797) belonging to the 10th percentile.

Figure 1. The trend of the average marginal effect of GDPPC: infections



Source: authors' elaborations

Turning to the number of deaths attributed to COVID-19, the results as at 21 April are shown in Table 5. As in the case of infections in Table 4, Column 1 shows a negative and strongly significant coefficient for PM 2.5: a higher level of air pollution correlates with a lower number of deaths. Here again, however, the same coefficient becomes much lower and no longer statistically significant in Column 2, when we add GDPPC as a regressor. In line with our hypothesis, the correlation between air quality and COVID-19-related deaths looks spurious. Instead, we find a positive and highly significant correlation with countries' wealth. This latter correlation is also robust to the inclusion of additional controls (Column 3) and continent-specific dummies (Column 4). As a further robustness check, we drop sixteen countries not reporting any deaths on April 21st (Column 5), but

[7] Multicollinearity becomes an issue when we include geographical dummies for regions smaller than continents: the higher the level of geographical disaggregation that we consider to generate region-specific dummies, the higher the correlation with the other regressors. We consequently choose the widest geographical level, i.e. continents, to control for common specific geographical characteristics across countries.

the results do not change. Finally, Tables A2 and A3 in the Appendix show that these results do not change when we measure the number of deaths as at 14 and 7 April 2020.

Table 5. Deaths. Correlation with pollution and wealth. 21 April 2020

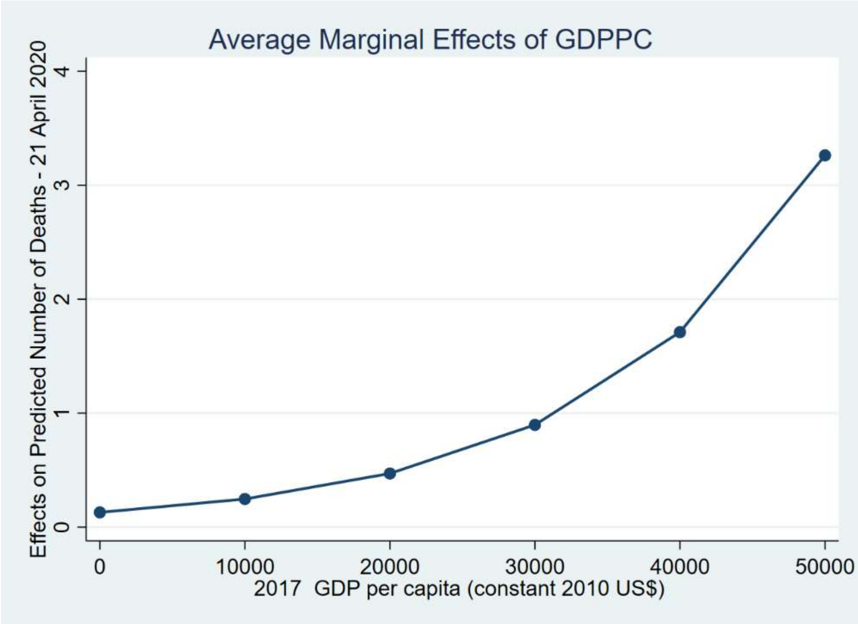
NEG BIN	(1)	(2)	(3)	(4)	(5)
PM2.5	-0.056***	-0.019	-0.015	-0.011	-0.008
	(0.008)	(0.012)	(0.009)	(0.011)	(0.011)
GDPPC		0.00008***	0.00007***	0.00006***	0.00007***
		(0.00002)	(0.00002)	(0.00002)	(0.00002)
POPULATION	0.017	0.017	0.004	0.005	0.004
	(0.015)	(0.014)	(0.006)	(0.007)	(0.005)
IMPORT/GDP			-4.978***	-5.286***	-5.241***
			(1.163)	(1.231)	(1.029)
AGRVA/GDP			-3.117	-2.437	-1.483
			(3.176)	(3.273)	(3.302)
MANVA/GDP			3.147	3.147	1.852
			(3.192)	(3.192)	(3.350)
CO_2			-0.050	-0.026	-0.060
			(0.072)	(0.074)	(0.064)
TEMP			-0.067***	-0.021	-0.026
			(0.022)	(0.024)	(0.025)
Area dummy	NO	NO	NO	YES	YES
N	142	142	142	142	126
Pseudo R^2	0.029	0.053	0.081	0.088	0.090
Wald χ^2	53.34***	41.92***	137.96***	217.08***	216.39***
Alpha	4.404***	3.585***	2.573***	2.567***	2.047***
Max VIF	1.08	1.22	2.21	3.02	4.44
Mean VIF	1.08	1.14	1.65	1.82	2.15
AIC	1701.05	1662.30	1622.71	1619.55	1564.85

Robust standard errors in brackets. Each estimate includes a constant term. *** Significant at 1% level; ** significant at 5% level; * significant at 10% level.

Based on the estimates in Column 4, we compute an average marginal effect of 0.338, and a marginal effect at the mean of GDPPC of 0.009[8]: a 10,000 US$ increase in GDPPC above the mean coincides with an average increase of 90 COVID-19-related deaths as at April 21st. Figure 2 plots the marginal effect with respect to different levels of GDPPC. As in the case of infections, it increases exponentially with the level of countries' wealth: the marginal effect of the wealthiest countries (2.950) is almost twenty-two times larger than that of the poorest (0.136).

[8] The corresponding average marginal effect and marginal effect at the mean from Column 5 are, respectively, 0.271 and 0.015.

Figure 2. The trend of the average marginal effect of GDPPC: deaths



Source: authors' elaborations

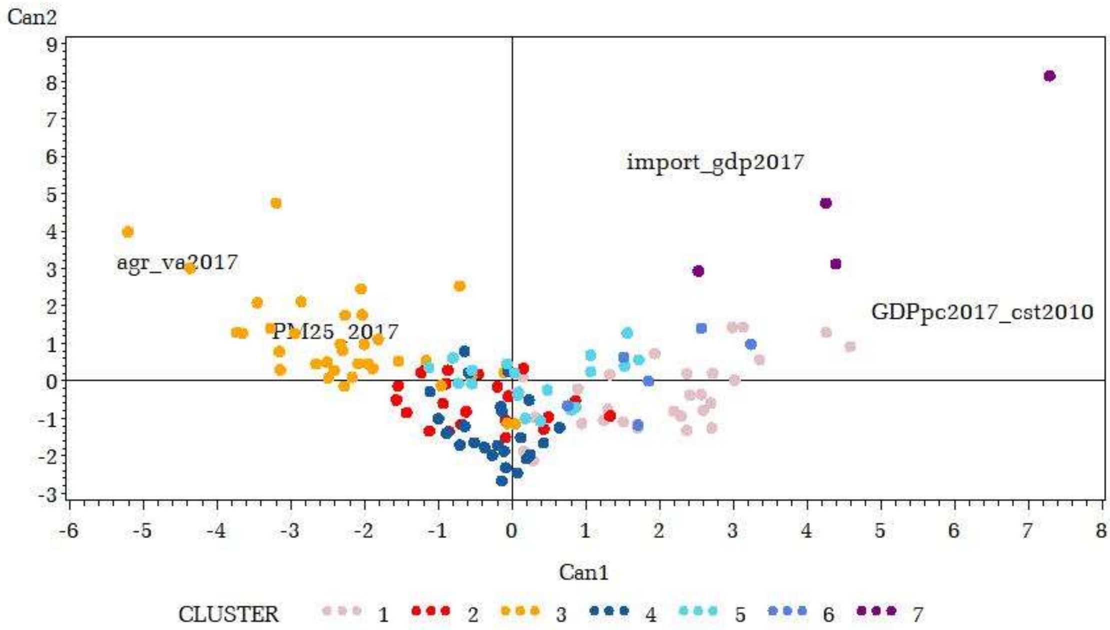
5. Cluster analysis

Although the impact of PM 2.5 on infections and deaths disappears once the level of countries' wealth has been taken into account, its negative sign deserves further examination. In this section, we check whether the association between air quality and COVID-19 outcomes changes across different areas of the world with respect to the economic structure and to climate-related variables. Since air pollution is the dependent variable in the regression model, it is not included in the cluster analysis to avoid endogeneity problems. The 142 countries are grouped using Ward's method, a well-known hierarchical approach to grouping observations (see Blashfield, 1980, for example). We identify seven clusters based on all the regressors in Equation 1 except for PM2.5 and population, the two variables that we use in the baseline specification of Tables 4 and 5, Column 1. Table A4 in the Appendix shows the eigenvectors of the correlation matrix, while the corresponding eigenvalues are shown in Table A5. Table A6 shows the mean values of each item in the clusters.

The composition of each cluster is shown in Table 6. Cluster 1 is the group whose variance explains the largest amount of the total variance. It includes many European countries and the US: these countries share a high-income level, a large share of services as a proportion of their GDP, a high exposure to CO_2 emissions per capita, a small share of agriculture, and a low temperature in March. Cluster 2 mainly comprises Eastern Asian and Northern African countries, which share a low weight of manufacturing as a proportion of domestic GDP and a moderately high import propensity. Cluster 3 is essentially made up of West Asian and Sub-Saharan countries, sharing a high openness to imports. Cluster 4 is a mix of countries sharing high CO_2 emissions and a high average temperature in March, e.g. countries below the Equator. Cluster 5 includes service economies sharing a low temperature in March. Cluster 6 contains high-income countries specializing in natural resource extraction. Cluster 7 pools four small open economies (three islands) with a large share of services and agriculture, and a moderately high level of CO_2 emissions.

We interpret the clusters using a linear discriminant analysis. We thus obtain linear combinations of the variables (the so-called canonical discriminant axes) that maximize the separation between the different classes/clusters. These axes are calculated to respect reciprocal orthogonality, so they can be used to plot individual data on a Cartesian space, to enable a visual inspection of the bivariate distribution of the clusters. Figure 3 below shows the distribution of the 142 countries, grouped into the seven clusters (using a different color for each cluster). Four clusters tend to stand out quite clearly. Clusters 7 and 1, on the right, denote high GDP levels. The countries they contain exhibit a high share of services as a proportion of GDP, and high levels of CO_2 emissions per capita, a low share of agriculture, and a low temperature in March. Cluster 7 is also characterized by high import levels. On the left, we see cluster 3, which is characterized by a high share of agriculture. In the middle, and lower down, we find cluster 4, with the lowest average import propensity. The remaining clusters 2, 5 and 6 are not neatly separable on Figure 3, but Table A6 in the Appendix shows the average values of each cluster for the variables used in the cluster analysis.

Figure 3. Distribution of countries and clusters on the first two canonical axes



Source: authors' elaborations

For each cluster, we adopt seven dummies, which take the value of 1 when a country belongs to the corresponding cluster. Then, we split our air pollution variable into seven new variables multiplying PM 2.5 levels by each cluster dummy (PM2.5*cluster$_i$). As a final step, we estimate the following equation, one for COVID-19 infections and one for related deaths (both as at 21 April 2020), using a negative binomial regression model:

$$Y_{iT} = \gamma_0 + \sum_{j=1}^{7} \gamma_j PM2.5 * cluster_j + \beta_2 POP_i + u_{iT}, \qquad [2]$$

where T refers to the 21st, 14th and 7th of April, respectively.

Table 6. Clusters of countries

Cluster	Countries
1	Australia, Austria, Belgium, Canada, Cyprus, Denmark, Spain, Estonia Finland, France, Germany, Georgia, Greece, Iceland, Israel, Italy, Japan, Kazakhstan, Lebanon, Latvia, Montenegro, Netherlands, Norway, New Zealand, Portugal, Russian Federation, Sweden, Switzerland, United Kingdom, United States.
2	Bangladesh, China, Cameroon, Algeria, Gabon, Equatorial Guinea, Guatemala, Honduras, Indonesia, India, Jordan, Korea, Rep., Morocco, Mexico, Malaysia, Nicaragua, Oman, Philippines, Paraguay, Senegal, El Salvador, Thailand, Tunisia.
3	Afghanistan, Albania, Azerbaijan, Benin, Burkina Faso, Bhutan, Central African Republic, Cote d'Ivoire, Congo, Dem. Rep., Congo, Rep., Ethiopia, Ghana, Guinea, Gambia, Guyana, Iraq, Kenya, Cambodia, Liberia, Madagascar, Myanmar, Mozambique, Mauritania, Niger, Nigeria, Nepal, Pakistan, Rwanda, Chad, Togo, Timor-Leste, Tanzania, Uganda, Uzbekistan, Vietnam.
4	Angola, Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Cabo Verde, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, Egypt, Grenada, Iran, Jamaica, Sri Lanka, Namibia, Panama, Peru, Uruguay, South Africa, Zambia, Zimbabwe.
5	Armenia, Bulgaria, Bosnia and Herzegovina, Belarus, Czech Republic, Croatia, Hungary, Kyrgyz Republic, Lithuania, Moldova, North Macedonia, Mongolia, Poland, Romania, Serbia, Slovak Republic, Slovenia, Turkey, Ukraine.
6	United Arab Emirates, Bahrain, Brunei Darussalam, Kuwait, Qatar, Saudi Arabia.
7	Ireland, Luxembourg, Malta, Singapore.

Table 7 shows the results. In one case, namely cluster 1 (the wealthiest economies in the world) higher PM 2.5 concentrations (strongly) correlate with higher rates of infection and death (at each date). We also find evidence of a (weak) positive correlation between PM2.5 and COVID-19 infections in cluster 5. In cluster 3, the association between PM2.5 and COVID-19 outcomes is negative and strongly significant: these low-income countries are mainly in Africa and East Asia. Such a negative and significant estimated coefficient is nevertheless roughly ten times smaller than the positive coefficient of PM*cluster1. The same order of magnitude holds for the marginal effects at the mean: on April 21st, a rise of 10 micrograms in PM2.5 levels per cubic meter corresponds, on average, to 9850 more infections and 608 more deaths in cluster 1 countries, and to 1430 fewer cases of infection and 64 less deaths in cluster 3 countries.

Table 7. Correlation between pollution, wealth and COVID-19 infections and related deaths, by cluster

	Deaths			Infections		
NEG BIN	21 April	14 April	7 April	21 April	14 April	7 April
PM2.5*cluster1	0.304***	0.313***	0.333***	0.211***	0.223***	0.234***
	(0.053)	(0.056)	(0.058)	(0.045)	(0.046)	(0.048)
PM2.5*cluster2	-0.013	-0.011	-0.006	-0.020**	-0.020*	-0.016
	(0.013)	(0.018)	(0.019)	(0.010)	(0.010)	(0.012)
PM2.5*cluster3	-0.032***	-0.031***	-0.025***	-0.031***	-0.030***	-0.032***
	(0.009)	(0.009)	(0.010)	(0.007)	(0.007)	(0.008)
PM2.5*cluster4	0.038	0.043	0.053*	0.028	0.029	0.030
	(0.025)	(0.026)	(0.027)	(0.022)	(0.023)	(0.025)
PM2.5*cluster5	0.033*	0.031	0.031	0.039**	0.036**	0.031*
	(0.019)	(0.020)	(0.020)	(0.017)	(0.018)	(0.018)
PM2.5*cluster6	-0.010	-0.011	-0.010	0.013**	0.009	0.005
	(0.008)	(0.008)	(0.009)	(0.005)	(0.006)	(0.006)
PM2.5*cluster7	0.085	0.064	0.056	0.097*	0.082	0.067
	(0.117)	(0.107)	(0.102)	(0.055)	(0.064)	(0.061)
POPULATION	0.011**	0.010**	0.009*	0.010***	0.010**	0.010***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
N	142	142	142	142	142	142
Pseudo R2	0.068	0.072	0.078	0.047	0.049	0.051
Wald χ2	128.1***	114.4***	95.96***	146.1***	127.8***	123.1***
Alpha	3.121***	3.138***	3.157***	2.081***	2.151***	2.151***
Max VIF	1.91	1.91	1.91	1.91	1.91	1.91
Mean VIF	1.45	1.45	1.45	1.45	1.45	1.45

Robust standard errors in brackets. Each estimate includes a constant term. *** Significant at 1% level; ** significant at 5% level; * significant at 10% level. Clusters are identified using the following variables: GDPPC, IMPORT/GDP, AGRVA/GDP, MANVA/GDP, SERVVA/GDP (services value added on GDP), CO_2 per capita and TEMP.

6. Conclusions

This analysis focuses on the relationship between air pollution, wealth and COVID-19 infections and related deaths. We consider a sample of 142 countries and information on COVID-19 diffusion and mortality at different times in March and April 2020, coupled with environmental and socio-economic variables. We find that the association between air quality and COVID-19 diffusion and mortality does not hold after controlling for countries' environmental and economic characteristics.

Instead, the level of wealth measured in terms of per capita GDP correlates strongly with both COVID cases and related deaths. A 10,000 US$ increase in GDPPC over its mean corresponds to an average 1790 more COVID-19 cases and 90 more COVID-related deaths on April 21st. Our analysis also shows that the share of manufacturing (as a proportion of GDP) correlates positively with the spread of the contagion. We mentioned earlier that lobbying by the manufacturing sectors in some countries might have delayed or loosened restrictions on people's movements. Our data seem to indicate that, under these circumstances, this had a significant impact on the diffusion of COVID-19.

Finally, our analysis suggests that temperature has played a part in the rate of contagion and deaths due to COVID-19. The southern hemisphere (where it was late summer and early autumn in March and April) clearly shows lower numbers of contagions and deaths.

To disentangle the impact of economic characteristics from physical factors such as average exposure to PM2.5, we grouped countries by economic and climate-related variables. Focusing on some single clusters, and particularly the clusters of wealthier and more advanced countries, we find a positive relationship between air pollution and COVID diffusion and related death: a rise of 1 mg in PM2.5 levels per cubic meter coincides with an average 233 more infections and 333 more deaths due to COVID-19 on April 21st.

At the same time, there is a limited negative relationship between air pollution and COVID infections and related deaths for countries in cluster 3, which are mostly in Sub-Saharan regions (the poorest economies, largely based on agriculture) for which we cannot advance a plausible explanation. This puzzle might relate to data quality issues, especially with organizational difficulties and the costs of testing for the infection on large samples of the population.

There are several factors to consider regarding the quality of available data on COVID-19. A first aspect concerns the homogeneity of the data collection process. Apart from the above-mentioned costs and organizational problems, different policies have been adopted around the world concerning the use of testing for the infection and mitigation measures. There has been a generalized scarcity of test kits, which has influenced how the phenomenon has been measured. Overall, it is safe to assume that the official COVID counts fall abundantly short of the real number of infections around the world.

This may be true of the real number of deaths as well. There are non-trivial problems with certifying a death as being due to COVID-19. It preliminarily demands doing a test. Many of the elderly people infected with COVID-19 have been treated outside hospitals, and died in nursing homes, adding to the difficulty of applying the test and establishing the cause of death[9]. Besides, a large proportion of the people dying with the infection are elderly and have underlying medical conditions, including cardiocirculatory and respiratory problems. In such cases, definitively establishing the ultimate cause of death is not always easy, and can be costly and time-consuming.

Overall, there is a large degree of uncertainty regarding the data describing the COVID-19 pandemic. The distribution of the accounting noise could depend on countries' level of wealth, since wealthier countries may be in a better position to cover the monitoring costs. Our analysis could not account for possible accounting bias across the countries. In a relatively near future, once general mortality data have become available for all countries, it will be possible to refine the official estimates more consistently. Comparing total deaths (due to whatever cause) in the period from January to April 2020 with the figures for previous years will identify the excess mortality, which will provide a reliable estimate of how many people have really died as a result of contracting COVID-19. Meanwhile, our analysis gives an account of the impact of air pollution and economic and environmental variables on the COVID-19 pandemic around the world. The study of this phenomenon is just beginning and we welcome future analyses that include local and global factors to help explain the relationship between air pollution and COVID-19 pandemic, as done in this work.

[9] In Italy, for instance, the classification protocol states that only people who die after officially testing positive in hospitals can be classified as COVID-19 victims. Some reports (e.g. Gabanelli and Ravizza, 2020) show that in several EU countries (e.g. Netherlands, Belgium, among others) the mortality rate due to coronavirus remains particularly low, but in the first four months of 2020 these countries have had more than double the mortality rates of the same period in 2019. It is not clear why different countries count COVID-19-related deaths differently.

References

EEA (2020), Air quality and COVID-19, European Environmental Agency, available at https://www.eea.europa.eu/themes/air/air-quality-and-covid19 (lasst access 2 May 2020)

Gabanelli M., Ravizza S., (2020), "Morti COVID, tutte le bugie in Europa. Ecco i dati reali", Corriere della Sera, Dataroom, at https://www.corriere.it/dataroom-milena-gabanelli/morti-covid-tutte-bugie-dell-europa-ecco-dati-reali/1c28ca00-88b3-11ea-96e3-c7b28bb4a705-va.shtml (in italian)

Ogen, Y, (2020), Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality, Science of The Total Environment, 726, https://doi.org/10.1016/j.scitotenv.2020.138605

Roger K. Blashfield (1980), The Growth Of Cluster Analysis: Tryon, Ward, And Johnson, Multivariate Behavioral Research, 15:4, 439-458, DOI: 10.1207/s15327906mbr1504_4

Setti, L., Passarini, F., De Gennaro, G., Baribieri, P., Perrone, M. G., Borelli, M., Palmisani, J., Di Gilio, A., Torboli, V., Pallavicini, A. Ruscio, M., Piscitelli, R., Miani, A., 2020, COVID-19 RNA Found on Particulate Matter of Bergamo in Northern Italy: First Preliminary Evidence, MedRXiv, https://doi.org/10.1101/2020.04.15.20065995

Sovacol, B. J. 2012, The political economy of energy poverty: A review of key challenges, Energy for Sustainable Development 16, 272-282

Wu, X, Nethery, R.C., Sabath, B., Braun, D, Dominici, F. (2020), Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. medRxiv 2020.04.05.20054502; doi: https://doi.org/10.1101/2020.04.05.20054502

Appendix

Table A1. Corrrelations between pollution, wealth and total COVID-19 infections

	14 April 2020				7 April 2020			
NEG BIN	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PM2.5	-0.033***	-0.012	-0.006	-0.005	-0.038***	-0.013	-0.005	-0.006
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
GDPPC		0.0001***	0.0001***	0.0001***		0.0001***	0.0001***	0.0001***
		(0.00001)	(0.00001)	(0.00001)		(0.00002)	(0.00001)	(0.00001)
POPULATION	0.011*	0.012	0.003	0.003	0.011	0.013	0.003	0.003
	(0.006)	(0.010)	(0.003)	(0.003)	(0.007)	(0.011)	(0.004)	(0.003)
IMPORT/GDP			-3.676***	-4.089***			-3.918***	-4.472***
			(0.596)	(0.606)			(0.593)	(0.565)
AGRVA/GDP			-4.080**	-2.435			-4.845**	-3.771*
			(1.990)	(2.185)			(1.906)	(2.078)
MANVA/GDP			5.483**	4.127			6.119**	4.053
			(2.561)	(2.613)			(2.639)	(2.668)
CO_2			0.021	0.014			0.023	0.008
			(0.056)	(0.050)			(0.068)	(0.057)
TEMP			-0.046***	-0.014			-0.047***	-0.010
			(0.013)	(0.017)			(0.014)	(0.017)
Area dummy	NO	NO	NO	YES	NO	NO	NO	YES
N	142	142	142	142	142	142	142	142
Pseudo R^2	0.021	0.040	0.066	0.071	0.022	0.043	0.071	0.078
Wald $\chi 2$	31.54***	34.63***	190.8***	276.41***	35.54***	31.12***	217.9***	342.9***
Alpha	3.015***	2.394***	1.735***	1.609***	3.111***	2.455***	1.751***	1.585***
Max VIF	1.08	1.22	2.21	4.44	1.08	1.22	2.21	4.44
Mean VIF	1.08	1.14	1.65	2.15	1.08	1.14	1.65	2.15
AIC	2601.94	2553.44	2496.35	2489.28	2481.63	2431.55	2371.18	2359.32

Robust standard errors in brackets. Each estimate includes a constant term. *** Significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A2. Correlations between pollution, wealth and total COVID-19 deaths, 14 April 2020

NEG BIN	(1)	(2)	(3)	(4)	(5)
PM2.5	-0.058***	-0.020	-0.016	-0.014	-0.012
	(0.008)	(0.014)	(0.010)	(0.013)	(0.011)
GDPPC		0.00008***	0.00008***	0.00006***	0.00007***
		(0.00002)	(0.00002)	(0.00002)	(0.00002)
POPULATION	0.017	0.018	0.004	0.004	0.003
	(0.018)	(0.017)	(0.005)	(0.005)	(0.003)
IMPORT/GDP			-5.163***	-5.528***	-5.443***
			(1.045)	(1.043)	(0.863)
AGRVA/GDP			-2.881	-3.135	-2.123
			(3.255)	(3.371)	(3.352)
MANVA/GDP			3.133	2.499	1.247
			(3.178)	(3.465)	(3.406)
CO_2			-0.032	-0.018	-0.056
			(0.084)	(0.086)	(0.073)
TEMP			-0.069***	-0.018	-0.023
			(0.023)	(0.027)	(0.028)
Area dummy	NO	NO	NO	YES	YES
N	142	142	142	142	125
Pseudo R^2	0.030	0.053	0.083	0.091	0.091
Wald χ^2	53.61***	36.40***	127.4***	194.1***	197.8***
Alpha	4.508***	3.699***	2.802***	2.611***	2.119***
Max VIF	1.08	1.22	2.21	4.44	4.44
Mean VIF	1.08	1.14	1.65	2.15	2.15
AIC	1604.09	1567.13	1526.67	1522.66	1470.73

Robust standard errors in brackets. Each estimate includes a constant term. *** Significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A3. Correlations between pollution, wealth and total COVID-19 deaths, 7 April 2020

NEG BIN	(1)	(2)	(3)	(4)	(5)
PM2.5	-0.058***	-0.019	-0.013	-0.012	-0.012
	(0.008)	(0.018)	(0.012)	(0.015)	(0.012)
GDPPC		0.00009***	0.00007***	0.00007***	0.00007***
		(0.00003)	(0.00002)	(0.00002)	(0.00002)
POPULATION	0.019	0.020	0.003	0.003	0.003
	(0.021)	(0.020)	(0.003)	(0.002)	(0.002)
IMPORT/GDP			-5.740***	-6.051***	-5.703***
			(0.975)	(0.877)	(0.886)
AGRVA/GDP			-3.333	-3.972	-2.844
			(3.511)	(3.590)	(3.306)
MANVA/GDP			3.183	2.360	0.490
			(3.330)	(3.556)	(3.409)
CO_2			-0.012	-0.005	-0.050
			(0.103)	(0.109)	(0.092)
TEMP			-0.066***	-0.007	-0.016
			(0.023)	(0.031)	(0.034)
Area dummy	NO	NO	NO	YES	YES
N	142	142	142	142	124
Pseudo R^2	0.030	0.053	0.090	0.100	0.097
Wald χ^2	51.71***	30.39***	122.4***	192.3***	194.9***
Alpha	4.639***	3.862***	2.839***	2.590***	2.173***
Max VIF	1.08	1.22	2.21	4.44	4.44
Mean VIF	1.08	1.14	1.65	2.15	2.15
AIC	1471.02	1437.10	1392.26	1385.05	1338.66

Robust standard errors in brackets. Each estimate includes a constant term. *** Significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A4. Eigenvectors of the correlation matrix

Cluster	Eigenvalue	Difference	Proportion	Cumulative
1	3.0519	1.9871	0.4360	0.4360
2	1.0648	0.1769	0.1521	0.5881
3	0.8879	0.1044	0.1268	0.7150
4	0.7835	0.1295	0.1119	0.8269
5	0.6540	0.2849	0.0934	0.9203
6	0.3692	0.1806	0.0527	0.9731
7	0.1886	-	0.0269	1

Table A5. Eigenvalues of the correlation matrix

Cluster →	1	2	3	4	5	6	7
GDPPC	0.480	0.147	-0.014	0.228	-0.045	0.747	-0.370
AGRVA/GDP	-0.484	0.199	0.125	0.091	-0.297	0.520	0.587
MANVA/GDP	0.124	-0.835	0.443	-0.120	0.078	0.213	0.161
SERVA/GDP	0.436	0.147	-0.291	-0.373	0.476	0.111	0.572
IMPORT/GDP	0.206	0.463	0.836	-0.076	0.108	-0.158	0.040
CO_2	0.390	-0.079	-0.033	0.757	-0.174	-0.280	0.399
TEMP	-0.368	-0.004	0.056	0.456	0.797	0.109	-0.078

Table A6. Average values of clusters

Cluster	Freq.	GDPPC	AGRVA/GDP	MANVA/GDP	SERVA/GDP	CO_2	IMPORT/GDP	TEMP
1	30	39074	0.03	0.12	0.66	7.9	0.42	2.8
2	23	6217	0.09	0.20	0.53	3.5	0.39	21.1
3	35	1615	0.23	0.09	0.43	0.8	0.43	22.1
4	25	6758	0.07	0.10	0.58	2.6	0.33	22.1
5	19	10809	0.06	0.16	0.54	5.0	0.61	2.0
6	6	35377	0.01	0.12	0.52	26.3	0.49	21.8
7	4	66405	0.01	0.16	0.70	10.1	1.39	10.6