The Political Risk Factors of COVID-19

Vera-Valdés, J. Eduardo

Published in:
International Review of Applied Economics

DOI (link to publication from Publisher):
10.1080/02692171.2020.1866973

Publication date:
2021

Document Version
Early version, also known as pre-print

RESEARCH ARTICLE

The Political Risk Factors of COVID-19.

J. Eduardo Vera-Valdés*

*Aalborg University, Department of Mathematical Sciences, Skjernvej 4A, Aalborg, 9220, Denmark

ARTICLE HISTORY
Compiled October 21, 2020

ABSTRACT
This paper analyses a broad range of macro variables to assess the effects that they have on the number of cases and deaths due to COVID-19. We consider 23 explanatory variables on health, political, and economic factors for 94 countries.

Given the large number of explanatory variables analysed, the paper employs advanced statistical tools for the analysis. We use regularised regression and dimension reduction methods to increase estimation efficiency.

We find that alcohol drinking is associated with an increase in the number of cases and deaths due to COVID-19. In this regard, our results support the World Health Organization's recommendation of reducing alcohol drinking during the pandemic.

Furthermore, our results show that the level of trust inside the society is associated with both the number of cases and deaths. A higher level of trust in medical personnel is associated with fewer cases, while a higher level of trust in the government is associated with fewer deaths due to COVID-19.

Finally, hospital beds per thousand inhabitants are a statistically significant factor in reducing the number of deaths.

Our results are robust to the estimation method, and they are of interest to governments and authorities responsible for the control of the pandemic.

KEYWORDS
Risk Factor, Morbidity, COVID-19, Political Economy, Trust.

1. Introduction

The COVID-19 pandemic has resulted in a major health crisis costing hundreds of thousands of lives around the World. Furthermore, as a way to mitigate the spread of the virus and reduce the number of cases and deaths, governments had to impose several restrictions on movement and commerce. These restrictions resulted in significant economic downturns whose total effects will not be known for several years. Even though the pandemic is still ongoing, a lot of knowledge has been gained in the past few months.

We are starting to isolate some of the potential risk factors on an individual level. In particular, several articles have found evidence for some of the comorbidities of COVID-19. The most prevalent comorbidities seem to be hypertension, diabetes, cardiovascular disease, and respiratory system disease; see Atkins et al. (2020); Chudasama et al. (2020); Yang et al. (2020). Moreover, some evidence has been obtained

Email: eduardo@math.aau.dk ORCID iD: 0000-0002-0337-8055

regarding the effect that pollution and human habits like smoking and alcohol drinking have on morbidity; see Alqahtani et al. (2020); Gupta et al. (2020); Hendryx and Luo (2020); Zoran, Savastru, Savastru, and Tautan (2020).

This paper adds to the literature on risk factors of COVID-19 on a macro-level. The data shows notable differences in the number of deaths per million inhabitants between countries. Nonetheless, the difference does not seem to be explained due to development, as the cases of Belgium, the United Kingdom, and the United States show. Yet, there may be other political or economic variables that may explain the difference in the number of deaths between countries. The goal of this paper is to identify factors that can be of use to design policies aimed at mitigating the number of cases and deaths due to COVID-19.

To achieve this goal, this paper analyses data on health-related variables, political conditions, pollution levels, economic variables, and trust-related variables. We consider 23 variables for 94 countries. The high number of regressors for a macro analysis may result in a lack of degrees of freedom in the estimation and less efficient estimates. Thus, we use advanced statistical tools to alleviate these concerns.

The surge of machine learning has resurfaced several statistical tools to deal with the problem of not enough degrees of freedom. We use two of the most popular statistical tools to deal with this problem: i) regularised regression via the lasso estimator, see Hastie, Tibshirani, and Wainwright (2016); Tibshirani (1996), and ii) dimension reduction via principal component regression, see Jolliffe (1982); Park (1981).

We find that a higher level of alcohol consumption in a country is associated with a higher number of cases and deaths due to COVID-19. Moreover, our results show the importance that trust has on controlling the effects of the pandemic. In particular, we find that trust in medical personnel is negatively associated with the number of cases, while trust in government is negatively related to the number of deaths. In this regard, our results show that collective, coordinated actions are needed to slow the spread of the virus.

This article proceeds as follows. The next section presents the data used in this study. Section 3 presents the results from the analysis, while Section 4 concludes.

2. Data

All data used in this paper was obtained from 'Our World in Data'. The website is a collaborative effort between the researchers of the Oxford Martin Programme on Global Development at the University of Oxford, and the non-profit organisation Global Change Data Lab. We use data on a broad range of health, political, economic, pollution and trust variables.

Table 1 presents an overview of the data considered in this study.

Name of variable	Description
$Pollution$	Population-weighted average level of exposure to concentrations of suspended particles measuring less than 2.5 microns in diameter. ($\mu g/m^3$).
$DeathsPollution$	Number of deaths per 100,000 population from both outdoor and indoor air pollution. Age-standardized.
$SmokeDaily$	Estimates of the prevalence of daily smoking, defined as the percentage of men and women, of all ages, who smoke daily.

$Drinking$	Share of adults aged 15 and older who drank any form of alcohol within the previous 12 months.
$UnsafeWater$	Share of deaths from unsafe water sources.
$Sanitation$	Death rates from unsafe sanitation measured as the number of deaths per 100,000 individuals.
$Overweigth$	Share of adults that are overweight or obese.
$Cardiovascular$	Annual number of deaths per 100,000 people from cardiovascular disease.
$Diabetes$	Diabetes prevalence (% of population aged 20 to 79).
$Aged65$	Share of the population that is 65 years and older.
$HospBeds$	Hospital beds per 1,000 people (OECD, Eurostat, World Bank, national government records and other sources).
$Corruption$	Transparency International's Corruption Perception Index. Scores are on a scale of 0-100, where 0 means that a country is perceived as highly corrupt.
$TrustShare$	Share of respondents who answered 'a lot' or 'some' to the question: 'How much do you trust your national government?'
$TrustMedics$	Share of people who trust doctors and nurses in their country.
$Literacy$	Estimates of the share of the population older than 14 years that is able to read and write.
$HumanRights$	Degree to which governments protect and respect human rights. The values range from -3.8 to around 5.4 (the higher the better).
$PoliticalRegime$	The scale goes from -10 (full autocracy) to 10 (full democracy).
$GiniIndex$	Gini Index. World Bank inequality data. A higher Gini index indicates higher inequality.
$EconomicFreedom$	Calculated by the Fraser Institute. Measures the degree to which individuals are free to choose, trade, and cooperate with others. Scores are on a scale of 0-10, where 10 represents maximum economic freedom.
$HealthShare$	Public health expenditure (%GDP).
$PopDensity$	Number of people divided by land area, measured in square kilometers.
$GDPpcp$	Gross domestic product at purchasing power parity (constant 2011 international dollars).
$Poverty$	Share of the population living in extreme poverty, most recent year available since 2010.
$TotCases$	Total confirmed cases of COVID-19 per 1,000,000 people as of August 7, 2020.
$TotDeaths$	Total deaths attributed to COVID-19 per 1,000,000 people as of August 7, 2020.

Table 1.: Data considered. Source: Our World in Data.

Besides the health-related measurements, the selected variables capture the distinct political and economic circumstances at each country before the start of the pandemic. We are interested in assessing if a more open and trusting society can cope better with the pandemic. Moreover, following recent results at the micro-level (see Gupta et al. (2020); Hendryx and Luo (2020); Zoran et al. (2020)), the dataset also allows us to test the comorbidity of pollution at the macro-level.

We first clean the data by removing countries with missing values in any of the variables in Table 1. Then, we consider the last available observation for each country. Even though there is an ongoing debate regarding a probable under-counting of COVID-19 cases and deaths, the statistical tools used for the analysis will not produce biased results as long as the under-counting is proportionally similar for all countries, allowing for some random variation between them. Notwithstanding that the pandemic is not over, we believe that the results presented in this paper can be of use in the design and implementation of policies aimed at mitigating the effects of the current pandemic, and help societies be better prepared for the next one.

The dataset contains 94 countries once we remove missing observations. As such, the analysis covers a broad range of countries with heterogeneous characteristics. The list of countries included in the dataset is presented in Appendix A.1, while summary statistics are presented in Appendix A.2. The cleaned data is available at `https://figshare.com/s/15c09557f9aa4c33da46`.

2.1. Standardised data

It is a well-known result that using raw data in a regularised regression can negatively affect the results. Given that regularised regressions penalise the size of the estimates, they are no longer free to take large values that may be associated to variables measured in a reduced scale in comparison to the independent variable, see Hastie, Tibshirani, and Friedman (2009). To alleviate these concerns, we standardise the data for all variables. Nonetheless, normalising the data does not qualitatively change the parametric estimation by ordinary least squares. The estimated coefficients adjust to the standardisation, given that they are free to increase with the scaling. In this regard, standardising the data makes it easier to compare estimators.

Moreover, standardised data is easier to analyse and present graphically. Figure 1 presents boxplots for the standardised data. The boxplots are a nonparametric graphical representation of the distribution of the data. They provide a clear representation of the data dispersion, which is preserved after standardisation.

As the figure shows, there is considerable heterogeneity between countries. Furthermore, there seem to be some extreme values for some of the variables, particularly in exposure to pollution and population density. This further shows the broad range of countries considered in the analysis.

2.2. Principal component analysis

As Table 1 and Figure 1 show, we have a large number of regressors for a macro-level analysis. A regression considering all of the regressors could potentially suffer from having too few degrees of freedom and thus lack efficiency.

An alternative to regularised regression for dealing with too few degrees of freedom is principal component regression. The idea behind principal component regression is to reduce the dimension of the space spanned by the regressors. The method first obtains the principal components for the regressors and uses only the ones that capture most of the variance. Typically, only a few principal components are needed to explain most of the variation in the data, thus the dimension reduction in the regression and the increase in degrees of freedom.

To estimate principal components regression, we first divide the regressors in five

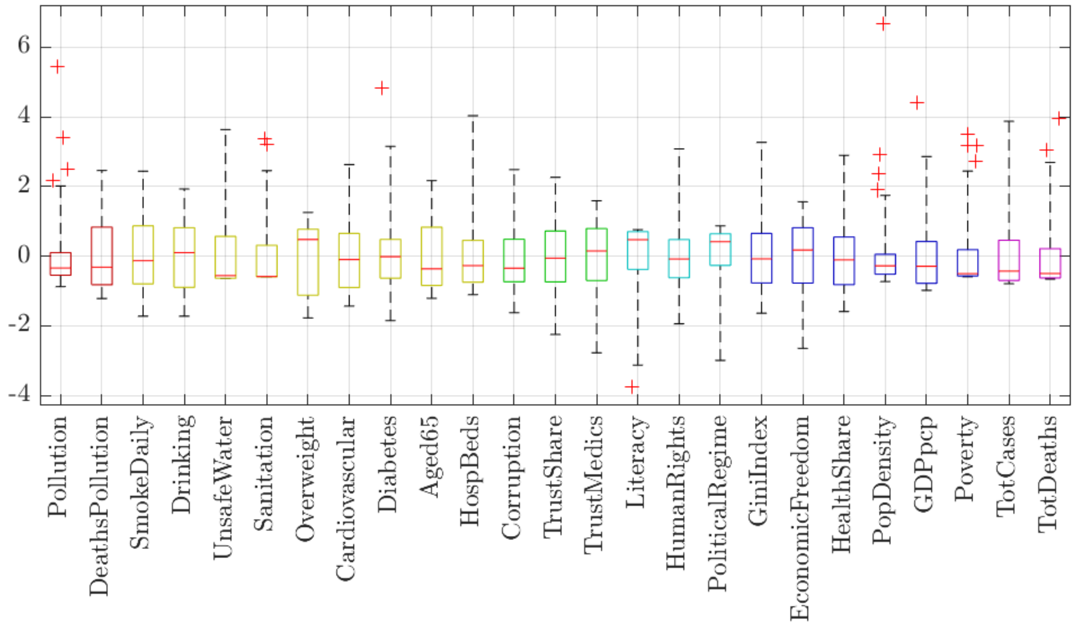

Figure 1. Boxplot of standardised data in Table 1. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually.

categories.

- ***PollutionVars***: *Pollution*, *DeathsPollution*.
- ***HealthVars***: *SmokeDaily*, *Drinking*, *UnsafeWater*, *Sanitation*, *Overweight*, *Cardiovascular*, *Diabetes*, *Aged65*, *HospBeds*.
- ***TrustVars***: *Corruption*, *TrustShare*, *TrustMedics*.
- ***PoliticalVars***: *Literacy*, *HumanRights*, *PoliticalRegime*.
- ***EconomicVars***: *GiniIndex*, *EconomicFreedom*, *HealthShare*, *PopDensity*, *GDPpcp*, *Poverty*.

Figure 2 presents the biplots, a graphical representation of the magnitude and sign of each variable's contribution to the first two principal components, for the five categories. Each observation in terms of those components is shown inside each biplot.

As can be seen from the figure, the biplots suggest that the first principal component captures almost all of the variance for each of the categories. Indeed, the first principal component captures more than 80% of the variance for the $PollutionVars$ (84%), $HealthVars$ (90%), $PoliticalVars$ (95%), and $EconomicVars$ (99%) categories. The only exception is the variance explained by the first component of the $TrustVars$ category that captures (58%) of the variance. These results suggest using only one principal component for the first four categories, while the first and second principal component for the $TrustVars$ category may be needed in the analysis.

3. Results

This section presents the results from the statistical analysis of the data presented in 2. We use modern statistical tools to get a better understanding of the effects of the

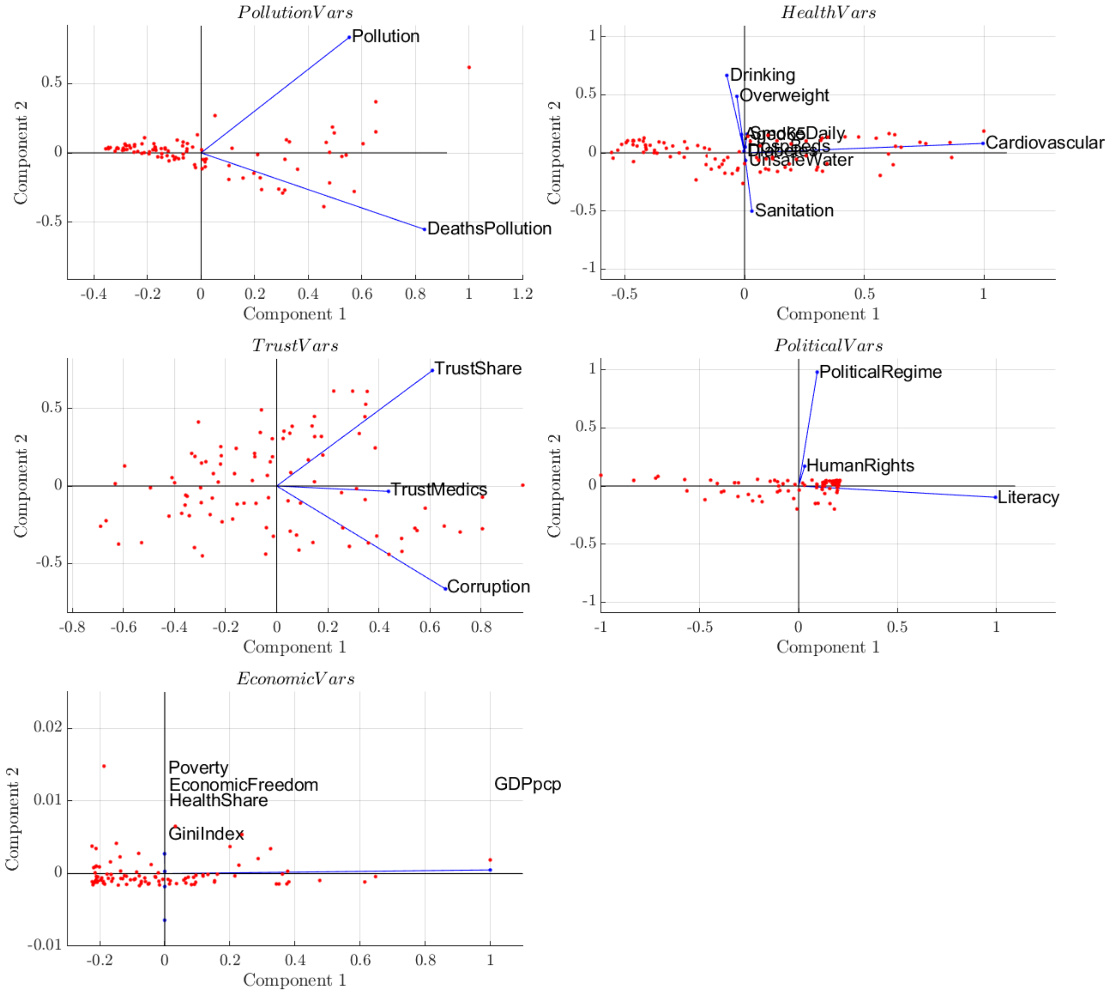

Figure 2. Principal components analysis.

political and economic variables on the impact of the pandemic at the macro-level. To ease the exposition, we present the results separately from the confirmed number of cases and the total number of deaths, both measured per million of inhabitants.

3.1. Confirmed cases of COVID-19 per million inhabitants

We consider a linear specification for the total number of cases of COVID-19 per million of inhabitants explained by the health, political, and economic-related macro variables to get a sense of their effects. The model is thus given by

$$TotCases = \alpha + X\beta + \epsilon, \tag{1}$$

where X is a matrix containing the 23 regressors presented in Table 1.

Table 2 presents the results from estimating Equation (1) using three distinct statistical tools.

Ordinary least squares estimation is considered by using all of the regressors in the specification. The estimation is thus simple to implement, and standard errors and

	Ordinary least squares			Forward stepwise selec.			Lasso
Variable	Est.	SE	pVal.	Est.	SE	pVal.	Est.
$(Intercept)$	0.000	0.088	1.000	0.000	0.083	1.000	–
$Pollution$	-0.051	0.156	0.744	–	–	–	–
$DeathsPollution$	-0.125	0.341	0.715	–	–	–	–
$SmokeDaily$	0.140	0.165	0.400	–	–	–	–
$Drinking$	0.400	0.202	0.051*	0.393	0.157	0.014**	0.165
$UnsafeWater$	-0.228	0.561	0.686	–	–	–	–
$Sanitation$	0.476	0.794	0.551	–	–	–	–
$Overweight$	0.245	0.274	0.374	0.272	0.180	0.135	0.226
$Cardiovascular$	-0.003	0.193	0.989	–	–	–	-0.085
$Diabetes$	-0.109	0.124	0.381	–	–	–	–
$Aged65$	-0.359	0.253	0.160	-0.224	0.192	0.246	-0.058
$HospBeds$	-0.222	0.169	0.194	-0.145	0.121	0.234	-0.006
$Corruption$	0.025	0.247	0.919	–	–	–	–
$TrustShare$	0.115	0.170	0.501	0.162	0.146	0.270	–
$TrustMedics$	-0.306	0.167	0.071*	-0.266	0.141	0.063*	-0.089
$Literacy$	0.033	0.240	0.891	–	–	–	–
$HumanRights$	-0.297	0.180	0.102	-0.267	0.137	0.055*	-0.023
$PoliticalRegime$	0.143	0.128	0.269	0.165	0.110	0.140	0.022
$GiniIndex$	0.199	0.136	0.146	0.206	0.108	0.060*	0.200
$EconomicFreedom$	0.003	0.142	0.981	–	–	–	–
$HealthShare$	0.064	0.177	0.719	–	–	–	–
$PopDensity$	0.099	0.117	0.402	–	–	–	–
$GDPpcp$	0.332	0.206	0.111	0.365	0.159	0.024**	0.139
$Poverty$	-0.440	0.320	0.173	-0.221	0.142	0.123	-0.109
$RMSE$	0.854			0.804			0.796
DoF	70			82			83
Adj. R^2	0.27			0.354			–

Table 2. Results from parametric estimators for number of cases of COVID-19 per million inhabitants. Est., pVal, $RMSE$, DoF, *Adj.* R^2 stand for estimates, p-values, root-mean squared error, degrees of freedom, and adjusted R-squared, respectively.

p-values are easily obtained. Nonetheless, as previously discussed, including such a large set of regressors may significantly reduce the degrees of freedom and thus reduce the efficiency of the estimates.

The results from ordinary least squares estimation show that *Drinking* seems to be positively associated with an increase in the number of cases of COVID-19. This result may point to the fact that gatherings of people increase the spread of the virus, particularly in closed environments like nightclubs, pubs, and bars. Thus, policies aimed at controlling the opening and functioning of indoor alcohol drinking places may play a major role in controlling the pandemic, as recent experience in the United Kingdom and Florida in the United States show. In particular, Florida and the United Kingdom have discussed or introduced bans on the selling of alcohol at several episodes during the pandemic.

Furthermore, the table shows that $TrustMedics$ is negatively related to the number of cases of COVID-19 per million inhabitants. This may point out that societies that trust the recommendations from the health authorities are better equipped to slow the spread of the virus. In this regard, constant honest communication from the health authorities may help in controlling the pandemic.

The table presents results from estimating Equation (1) using forward stepwise selection to alleviate the concern on the degrees of freedom. The idea behind forward stepwise selection is to start with the simplest model containing only the constant term and add variables one at a time following some fitness criteria. A typical rule is to add a regressor only if it increases the adjusted R-squared of the regression, and we stop adding regressors when we can no longer increase the adjusted R-squared. The estimation is thus simple but potentially computationally burdensome. The advantage of forward stepwise selection is that the selected model is usually much smaller than the complete model, which results in more degrees of freedom and thus more efficient estimates.

The increase in efficiency gained from forward stepwise selection allows us to isolate some other factors that may explain the different number of cases of COVID-19 per million inhabitants between countries.

First, in agreement with ordinary least squares, forward stepwise selection identifies $Drinking$ and $TrustMedics$ as significant variables capable of explaining the number of cases.

Furthermore, forward stepwise selection finds that $HumanRights$ is negatively associated with the number of cases. This may suggest that protecting and respecting human rights may allow the population of a country to decide the best way to avoid contagious on a macro-level.

Then, both $GiniIndex$ and $GDPpcp$ are positively associated with the number of cases. In conjunction, these factors suggest that more developed countries with high levels of economic inequality are associated with a higher number of cases. This may point to the fact that the population in this type of countries are typically more individualistic, and thus less prone to act in a way that benefits society in general at a small personal cost. The current debate around the wearing of masks in the United States, Mexico, and Brazil is an example of an egoistic behaviour that makes it harder to stop the spread of the virus. Thus, the results suggest that societies that show more solidarity are better equipped to control the spread of the virus.

A further alternative to increase the degrees of freedom in the estimation is regularised regression. In this paper, we focus on the lasso estimator. Lasso adds a penalty term as a function of the size of the estimates to the residual sum of squares defined by ordinary least squares. That is, lasso solves the problem given by

$$\min_{\beta} = (TotCases - \alpha - x\beta)'(TotCases - \alpha - x\beta) + \lambda|\beta|_{\ell_1},$$

where λ is the weight associated to the penalty term, and $|\beta|_{\ell_1}$ is the ℓ_1-norm given by $|\beta|_{\ell_1} = \sum_i^n |\beta_i|$ where $\beta' = (\beta_1, \cdots, \beta_n)$. The use of ℓ_1-norm allows the lasso estimator to make variable selection. That is, the lasso is capable of selecting the variables that increase the fit of the model and removing the rest. As such, lasso estimation increases the degrees of freedom by constructing a smaller model. The weight λ is typically selected by cross-validation so that the model is fit in a subset of the data and evaluated in the remaining observations. We select the λ with the best fit to the data.

One drawback of lasso estimation is that standard errors and p-values for inference are no longer a byproduct of the estimation. Nevertheless, lasso estimates can give us an indication regarding which variables better seem to explain the number of cases of COVID-19 per million inhabitants.

Results from the lasso estimation in Table 2 are much in line with forward stepwise selection except for the swap of $Cardiovascular$ for $TrustShare$. As such, the main outcomes are maintained.

A final alternative to deal with the reduced degrees of freedom is presented in Table 3. The table shows the results from principal component regression for $TotCases$. Principal component regression is a dimensionality reduction method to estimate a linear model with a large number of regressors, see Jolliffe (1982); Park (1981). The idea is to project the regressors into a lower-dimensional space spanned by the principal components, and estimate the specification in this lower-dimensional space. The method thus gains degrees of freedom and efficiency.

	Principal comp. reg.			Principal comp. reg.		
Variable	Est.	SE	pVal.	Est.	SE	pVal.
$(Intercept)$	0.000	0.091	1.000	0.000	0.091	1.000
$PollutionVars$	-0.091	0.213	0.669	-0.137	0.224	0.543
$HealthVars$	-0.235	0.114	0.043**	-0.233	0.115	0.045**
$PoliticalVars$	0.172	0.182	0.347	0.158	0.184	0.393
$EconomicVars$	0.289	0.160	0.075*	0.346	0.182	0.061*
$TrustVars$	-0.317	0.128	0.015**	-0.365	0.147	0.015**
$TrustVars\ 2PC$	–	–	–	0.091	0.136	0.504
$RMSE$	0.881			0.883		
DoF	88			87		
$Adj.\ R^2$	0.224			0.22		

Table 3. Results estimation by principal components regression for number of cases of COVID-19 per million inhabitants. Est., SE, pVal, $RMSE$, DoF, $Adj.\ R^2$ stand for estimates, standard errors, p-values, root-mean-squared error, degrees of freedom, and adjusted R-squared, respectively.

Principal component regression uses ordinary least squares in the equation given by

$$TotCases = \alpha + \beta_1 PollutionVars + \beta_2 HealthVars + \beta_3 PoliticalVars + \beta_4 EconomicVars + \beta_5 TrustVars + \varepsilon, \quad (2)$$

where the variables are obtained by principal component analysis as described in 2.2. Furthermore, as previously discussed, the first principal component already explains more than 80% of the variation for all variables besides $TrustVars$ where two first principal components may be needed in the analysis.

Notice that by construction, Equation (2) has fewer regressors than Equation (1), and thus more degrees of freedom. Moreover, from Table 3 notice the decrease in the adjusted R-squared and increase in root-mean-squared error from the specification with only the first principal component against the one including two first principal components for $TrustVars$. These statistics suggest that including two first principal components result in a worse fit. Thus, we continue the analysis considering only the specification with just one first principal component for all variables.

The results from principal component regression in Table 3 show that $HealthVars$, $EconomicVars$, and $TrusVars$ are significantly related to the number of cases per million inhabitants. Recalling that $Drinking$ is contained in the $HealthVars$ category (with negative loading as shown in Figure 2 and Appendix A.3), $TrustMedics$ is contained in the $TrustVars$ category (with positive loading), and that $GiniIndex$ and $GDPpcp$ are contained in the $EconomicVars$ category (both with positive loadings), the results from principal component regression are much in line with the results from the forward stepwise selection and lasso estimation. As such, the results are quite robust to the estimation method.

3.2. Deaths by COVID-19 per million inhabitants

In this section, we model the total number of deaths by COVID-19 per million of inhabitants using a linear specification on the health, political, and economic-related variables. The model is thus given by

$$TotDeaths = \alpha + X\beta + \epsilon, \tag{3}$$

where X is a matrix containing the 23 regressors presented in Table 1.

Analogously to the analysis for the number of cases, we estimate Equation (3) using three distinct statistical methods. The results from the estimations using ordinary least squares, forward stepwise selection, and lasso are presented in Table 4.

	Ordinary least squares			Forward stepwise selec.			Lasso
Variable	Est.	SE	pVal.	Est.	SE	pVal.	Est.
$(Intercept)$	0.000	0.089	1.000	0.000	0.084	1.000	–
$Pollution$	-0.063	0.158	0.694	–	–	–	–
$DeathsPollution$	0.088	0.346	0.800	0.028	0.217	0.897	–
$SmokeDaily$	0.061	0.168	0.717	–	–	–	–
$Drinking$	0.251	0.205	0.224	0.361	0.147	0.016**	0.109
$UnsafeWater$	-0.080	0.569	0.889	–	–	–	–
$Sanitation$	0.189	0.805	0.815	–	–	–	–
$Overweight$	0.123	0.278	0.660	–	–	–	0.217
$Cardiovascular$	-0.212	0.196	0.283	-0.137	0.131	0.298	-0.221
$Diabetes$	-0.182	0.126	0.152	–	–	–	–
$Aged65$	0.002	0.257	0.993	–	–	–	–
$HospBeds$	-0.352	0.172	0.044**	-0.286	0.123	0.022**	-0.005
$Corruption$	-0.054	0.251	0.831	–	–	–	–
$TrustShare$	-0.174	0.172	0.315	-0.180	0.104	0.087*	-0.026
$TrustMedics$	-0.100	0.169	0.557	–	–	–	–
$Literacy$	0.094	0.244	0.701	–	–	–	–
$HumanRights$	-0.348	0.182	0.060*	-0.306	0.136	0.026**	–
$PoliticalRegime$	0.048	0.130	0.711	–	–	–	–
$GiniIndex$	-0.049	0.138	0.724	–	–	–	–
$EconomicFreedom$	0.097	0.144	0.502	–	–	–	–
$HealthShare$	0.230	0.180	0.206	0.253	0.142	0.080*	0.070
$PopDensity$	0.122	0.119	0.308	–	–	–	–
$GDPpcp$	0.216	0.209	0.305	0.209	0.164	0.207	–
$Poverty$	-0.271	0.324	0.406	-0.188	0.149	0.210	–
$RMSE$	0.867			0.813			0.8155
DoF	70			84			88.000
$Adj.\ R^2$	0.249			0.339			

Table 4. Results from parametric estimators for number of deaths by COVID-19 per million inhabitants. Est., pVal, $RMSE$, DoF, $Adj.\ R^2$ stand for estimates, p-values, root-mean squared error, degrees of freedom, and adjusted R-squared, respectively.

The results from ordinary least squares estimation using all of the variables show that only *HospBeds* and *HumanRights* seem to be significantly associated with the number of deaths, both with negative signs. These results suggest that hospital capacity measured in the number of beds available per thousand inhabitants help in controlling the number of deaths due to COVID-19. This result supports the high priority that several countries set around hospital capacity to deal with the effects of COVID-19, particularly after noticing the impact of the pandemic in Italy and Spain. In particular, having more available hospital beds to treat patients worst hit by COVID-19 is shown to reduce the number of deaths.

Moreover, but perhaps less intuitively, countries that better respect human rights seem to be associated with fewer number of deaths due to COVID-19. As previously argued, this result may relate to a population that is more free to decide how to avoid getting infected, or once infected how to look for proper care.

The high number of regressors, and thus the lower degrees of freedom, could be one of the reasons behind the few significant estimates computed by ordinary least squares. As previously discussed, one way to alleviate these concerns is to estimate the model using forward stepwise selection to increase the degrees of freedom, and thus the efficiency of the estimates. Results from Table 4 using forward stepwise selection show that in addition of *HospBeds* and *HumanRights*, *Drinking*, *TrustShare*, and *HealtShare* are also found to be significantly associated to the number of deaths due to COVID-19 per million inhabitants.

The results from *Drinking* are qualitatively similar to the analysis for the number of cases. That is, societies with a higher share of the population that drink alcohol are associated to a higher number of deaths due to COVID-19. This result may be related to the argument for the number of cases. Societies with a higher proclivity to alcohol drinking may be more prone to infection. Furthermore, once infected, alcohol-related health conditions may perversely affect recovery. As argued in World Health Organization (2020), 'Alcohol use, especially heavy use, weakens the immune system and thus reduces the ability to cope with infectious diseases.' In this regard, our results support the World Health Organization's recommendation of reducing alcohol drinking during the pandemic.

Furthermore, the positive coefficient for *TrustShare* advocates that societies that trust their governments are better equipped to control the number of deaths due to COVID-19. This may relate to more people believing, and thus following, the government's guidelines regarding COVID-19 mitigation policies. Thus, our results indicate that governments should gain the trust of their citizens as a way to reduce the number of deaths during the pandemic. In this regard, a clear and honest set of guidelines should be implemented.

The result regarding *HealthShare* seems counterintuitive at first. The sign of the coefficient appears to suggest that higher public spending on health care is associated with a more significant number of deaths. Nonetheless, note that spending on health care services includes treatments for cancer, diabetes, and a broad range of cardiovascular diseases. As found by Dieleman et al. (2017), 'Increases in US health care spending from 1996 through 2013 were largely related to increases in health care service price and intensity but were also positively associated with population growth and aging and negatively associated with disease prevalence or incidence.' Thus, it may be the case that countries that spend more in health care are those with a higher proportion of older individuals, with more comorbidities to COVID-19. In this regard, the result from *HospBeds* is particularly revealing. Our results suggest that countries should consciously assign a proportion of their health care spending on policies directly aimed to care for COVID-19 patients to reduce the number of deaths.

The results from lasso estimation broadly agree with the ones from forward stepwise selection except for $HumanRights$. In this regard, the main outcomes are maintained.

Finally, Table 5 presents the results from principal components regression for $TotDeaths$.

	Principal comp. reg.		
Variable	Est.	SE	pVal.
$(Intercept)$	0.000	0.089	1.000
$PollutionVars$	-0.132	0.209	0.530
$HealthVars$	-0.320	0.112	0.005***
$PoliticalVars$	0.089	0.178	0.619
$EconomicVars$	0.263	0.167	0.098*
$TrustVars$	-0.243	0.125	0.055*
$RMSE$	0.862		
DoF	88		
$Adj.\ R^2$	0.257		

Table 5. Results estimation by principal components regression for number of cases of COVID-19 per million inhabitants. Est., SE, pVal, $RMSE$, DoF, $Adj.\ R^2$ stand for estimates, standard errors, p-values, root-mean-squared error, degrees of freedom, and adjusted R-squared, respectively.

To model $TotDeahts$, principal component regression uses ordinary least squares in the equation given by

$$TotDeaths = \alpha + \beta_1 PollutionVars + \beta_2 HealthVars + \beta_3 PoliticalVars + \beta_4 EconomicVars + \beta_5 TrustVars + \varepsilon, \quad (4)$$

where the regressors are defined as in Equation (2). Following the discussion in the section for the number of cases, we consider only one first principal component for all variables.

Notice that the results from principal component regression are much in line with the results on the high-dimensional specification in Equation (3). In particular, $HealthVars$, $EconomicVars$, and $TrustVars$ seem to be significantly associated with the number of deaths due to COVID-19 per million inhabitants.

Recalling that $Drinking$ and $HospBeds$ are contained in the $HealthVars$ category (with negative and positive loading, respectively), $HealthShare$ is contained in the $EconomicVars$ (with positive loading), and $TrustShare$ is contained in the $TrustVars$ category (with positive loading), the results from principal component regression are much in line with the results from the forward stepwise selection and lasso estimation. As such, the results from the analysis of the number of deaths due to COVID-19 are robust to the estimation method.

4. Conclusions

We have analysed a broad range of health, political, and economic variables at a macro-level to assess the effects that they may have on the number of cases and deaths due to COVID-19. The data contains information on 23 variables for 94 countries. Given the large amount of regressors, we use a broad range of advanced statistical tools for the analysis. We use regularised regression, forward stepwise selection, and principal component regression to increase the degrees of freedom and estimation efficiency.

Our results suggest that alcohol drinking has perverse consequences on the effects of the pandemic. We find that, at a macro-level, alcohol drinking is positively associated with both the number of cases and deaths due to COVID-19. In this regard, our results support the World Health Organization's advice that people should minimise their alcohol consumption during the pandemic.

Furthermore, we find that higher GDP per capita and income inequality are associated with an increase in the number of cases. The combination of these two factors may point to the fact that an individualistic society is less prone to adopt the necessary altruistic behaviour that may be needed to control the pandemic. As previously argued, this result may relate to the neglect of the use of masks in some of the countries worst hit by COVID-19.

Moreover, our analysis does not seem to support the notion that pollution levels increase the morbidity of COVID-19, at least at a macro-level. Pollution was not found to be a significant factor for either the number of cases or deaths. In this regard, it may be the case that the effect of pollution is a much more local affair that is not maintained in the aggregate at the country level.

Finally, our results highlight the importance that trust has on controlling the pandemic. We find that higher levels of trust in medical personnel are associated with a lower number of COVID-19 cases per million inhabitants. This, of course, reflects the fact that the more people trust the medical staff, the easier it will be for people to follow their recommendations to control the spread of the virus. Moreover, trust in government is found to be significantly associated with the number of deaths due to the pandemic. A higher level of trust in government translates into a lower level of deaths due to COVID-19. Thus, our results highlight that governments with clear and honest communication are better at controlling the perverse effects of the pandemic.

Overall, our results show that collective action is needed to control the pandemic. Countries that are better prepared to cope with COVID-19 are those that can react to the pandemic in a timely, coordinated fashion.

Disclosure statement

The authors declare no conflict of interest.

Data availability statement

The data that support the findings of this study are available at `https://figshare.com/s/15c09557f9aa4c33da46`. These data were derived from resources freely available in the public domain in 'Our World in Data' at `https://ourworldindata.org/`.

References

Alqahtani, J. S., Oyelade, T., Aldhahir, A. M., Alghamdi, S. M., Almehmadi, M., Alqahtani, A. S., ... Hurst, J. R. (2020). Prevalence, Severity and Mortality associated with COPD and Smoking in patients with COVID-19: A Rapid Systematic Review and Meta-Analysis. *PloS one*, *15*(5), e0233147.

Atkins, J. L., Masoli, J. A. H., Delgado, J., Pilling, L. C., Kuo, C.-L., Kuchel, G. A., & Melzer, D. (2020). Preexisting Comorbidities Predicting COVID-19 and Mortality in the

UK Biobank Community Cohort. *The Journals of Gerontology: Series A*. Retrieved from https://doi.org/10.1093/gerona/glaa183
Chudasama, Y. V., Gillies, C. L., Appiah, K., Zaccardi, F., Razieh, C., Davies, M. J., ... Khunti, K. (2020, 6). Multimorbidity and SARS-CoV-2 infection in UK Biobank. *Diabetes & metabolic syndrome*, *14*(5), 775–776.
Dieleman, J. L., Squires, E., Bui, A. L., Campbell, M., Chapin, A., Hamavid, H., ... Murray, C. J. L. (2017, 11). Factors Associated With Increases in US Health Care Spending, 1996-2013. *JAMA*, *318*(17), 1668–1678. Retrieved from https://doi.org/10.1001/jama.2017.15927
Gupta, A., Bherwani, H., Gautam, S., Anjum, S., Musugu, K., Kumar, N., ... Kumar, R. (2020). Air pollution aggravating COVID-19 lethality? Exploration in Asian cities using statistical models. *Environment, Development and Sustainability*. Retrieved from https://doi.org/10.1007/s10668-020-00878-9
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning.* Retrieved from https://link.springer.com/content/pdf/10.1007%2F978-0-387-84858-7.pdf%0Ahttp://link.springer.com/10.1007/978-0-387-84858-7
Hastie, T., Tibshirani, R., & Wainwright, M. (2016). *Statistical Learning with Sparsity: The Lasso and Generalizations.* Chapman & Hall. Retrieved from http://doi.wiley.com/10.1111/insr.12167
Hendryx, M., & Luo, J. (2020). COVID-19 prevalence and fatality rates in association with air pollution emission concentrations and emission sources. *Environmental Pollution*, *265*, 115126. Retrieved from http://www.sciencedirect.com/science/article/pii/S0269749120331419
Jolliffe, I. T. (1982, 8). A Note on the Use of Principal Components in Regression. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *31*(3), 300–303. Retrieved from http://www.jstor.org/stable/2348005
Park, S. H. (1981, 8). Collinearity and Optimal Restrictions on Regression Parameters for Estimating Responses. *Technometrics*, *23*(3), 289–295. Retrieved from http://www.jstor.org/stable/1267793
Tibshirani, R. (1996). *Regression Selection and Shrinkage via the Lasso* (Vol. 58) (No. 1). Retrieved from http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.35.7574
World Health Organization. (2020). Alcohol and COVID-19 : what you need to know. *World Health Organization*. Retrieved from https://www.euro.who.int/__data/assets/pdf_file/0010/437608/Alcohol-and-COVID-19-what-you-need-to-know.pdf
Yang, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., ... Zhou, Y. (2020). Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *International Journal of Infectious Diseases*, *94*, 91–95. Retrieved from http://www.sciencedirect.com/science/article/pii/S1201971220301363
Zoran, M. A., Savastru, R. S., Savastru, D. M., & Tautan, M. N. (2020). Assessing the relationship between surface levels of PM2.5 and PM10 particulate matter impact on COVID-19 in Milan, Italy. *Science of The Total Environment*, *738*, 139825. Retrieved from http://www.sciencedirect.com/science/article/pii/S0048969720333453

Appendix A. Additional information

A.1. List of countries

List of countries included in the analysis: Albania, Algeria, Argentina, Armenia, Australia, Austria, Bangladesh, Belgium, Benin, Bolivia, Brazil, Bulgaria, Burkina Faso, Burundi, Cameroon, Canada, Chile, Colombia, Costa Rica, Croatia, Denmark, Dominican Republic, Ecuador, El Salvador, Estonia, Ethiopia, Gabon, Gambia, Georgia, Ghana, Greece, Guatemala, Guinea, Haiti, Honduras, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Jordan, Kazakhstan, Kenya, Kyrgyzstan, Latvia, Liberia, Lithuania, Luxembourg, Macedonia, Madagascar, Malawi, Malaysia, Mauritius, Mexico, Moldova, Mongolia, Montenegro, Morocco, Mozambique, Myanmar, Nepal, Nicaragua, Niger, Norway, Pakistan, Panama, Paraguay, Peru, Portugal, Romania, Russia, Slovakia, South Africa, South Korea, Spain, Sri Lanka, Sweden, Tanzania, Thailand, Togo, Tunisia, Turkey, Uganda, Ukraine, United Kingdom, United States, Uruguay, Yemen, Zambia, Zimbabwe.

A.2. Summary statistics

Variable	Min	Median	Mean	Max
$Pollution$	5.20	22.00	32.79	203.74
$DeathsPollution$	9.40	46.46	59.53	160.50
$SmokeDaily$	3.70	16.40	17.47	36.90
$Drinking$	1.90	47.85	45.36	93.90
$UnsafeWater$	0.00	0.21	1.62	10.73
$Sanitation$	0.00	0.27	11.66	76.93
$Overweight$	18.10	56.85	48.65	70.20
$Cardiovascular$	85.76	235.17	246.21	539.85
$Diabetes$	0.99	6.77	6.83	22.02
$Aged65$	2.17	7.40	9.65	23.02
$HospBeds$	0.20	2.16	2.80	12.27
$Corruption$	14.00	37.00	43.30	88
$TrustShare$	10.95	50.45	51.55	92.34
$TrustMedics$	49.77	82.23	80.60	98.20
$Literacy$	19.10	94.84	86.45	100
$HumanRights$	-2.47	0.34	0.47	5.13
$PoliticalRegime^*$	-7.00	8.00	6.18	10
$GiniIndex$	25.50	37.55	38.17	63.40
$EconomicFreedom$	4.84	7.00	6.87	8.07
$HealthShare$	0.79	3.84	4.07	10.02
$PopDensity$	1.98	78.74	126.41	1265.04
$GDPpcp$	702.23	12703.05	17788.49	94277.97
$Poverty$	0.10	1.75	11.31	77.60
$TotCases$	6.60	1492.39	3265.22	19181.16
$TotDeaths$	0.00	29.94	122.65	850.85

Table A1. Summary statistics. *We have updated Iceland's *PoliticalRegime* from a 1816 designation as a colony to a current democracy.

A.3. Principal component loadings

HealthVars		***EconomicVars***		***PoliticalVars***	
SmokeDaily	0.011	*GiniIndex*	-0.000	*Literacy*	0.995
Drinking	-0.074	*PopDensity*	-0.001	*HumanRights*	0.030
Overweight	-0.032	*GDPpcp*	0.999	*PoliticalRegime*	0.094
Aged65	-0.013	*Poverty*	-0.001	***TrustVars***	
Diabetes	-0.002	*EconomicFreedom*	0.000	*Corruption*	0.661
UnsafeWater	0.004	*HealthShare*	0.001	*TrustShare*	0.610
Sanitation	0.301	***PollutionVars***		*TrustMedics*	0.438
Cardiovascular	0.996	*Pollution*	0.553		
HospBeds	0.003	*DeathsPollution*	0.833		

Table A2. Loadings from principal components analysis.