Heliyon 8 (2022) e10729

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Methods for assessing the impact of $PM_{2.5}$ concentration on mortality while controlling for socio-economic factors

Dejan Lončar [a,*], Nicholas Brown Tyack [b], Vesna Krstić [c], Jane Paunković [d]

[a] Institute of Global Health - University of Geneva, 9 Chemin des Mines, 1202 Geneva, Switzerland
[b] University of Saskatchewan, 51 Campus Dr, Saskatoon, SK S7N 5A8, Canada
[c] MMI Bor, Zeleni Bulevar 35 & University of Belgrade, TF Bor, VJ 12, 19210 Bor, Serbia
[d] University Business Academy in Novi Sad, Faculty of Social Sciences, Belgrade, Serbia

ARTICLE INFO

Keywords:
$PM_{2.5}$ air pollution and socio economic factors
Health and air pollution
Method research
$PM_{2.5}$
Regression analysis

ABSTRACT

Even though industrial development has brought vast improvements to our daily lives, it carries with it negative effects such as adverse health outcomes caused by $PM_{2.5}$ and other pollutants. The negative externalities and external costs might occur when property rights are not properly defined, which means that if no one holds a property right on the atmosphere and the quality of air, there is no appropriate mechanism to prevent a further expansion of negative effects. An economic burden of pollution related to premature morbidity and mortality in individual countries can account for 5–14% of GDP (World Bank, 2021). In 2019, the worldwide health cost of mortality and morbidity caused by exposure to $PM_{2.5}$ concentration was $8.1 trillion, which is equivalent to 6.1 percent of the global gross domestic product (GDP) (World Bank estimate). Policymakers require evidence-based results that clearly show the impact that air pollution has on the economy and society, in order to be able to establish the proper regulations and ensure their successful implementation. The purpose of this long term study is to provide methods for assessing the negative effects of $PM_{2.5}$ concentration on $PM_{2.5}$-related mortality using panel data structure and demonstrate how socio-economic factors affect this relation. This study employed advanced econometric techniques to analyse the long-term impact of $PM_{2.5}$ on human health, while controlling for socio economic indicators. This study has demonstrated significant effects of socio-economic, health risk and system and governance variables on the relation between $PM_{2.5}$ concentration and $PM_{2.5}$-related mortality.

1. Introduction

Large-scale industrial development is counted as one of the primary reasons for the appearance of $PM_{2.5}$ in the rapidly developing modern world. Polluted air spreads through the environment and affects people's health, and according to Cohen et al. (2017) $PM_{2.5}$ pollution is a contributing factor to cardiovascular and respiratory diseases, including lung cancer. In numerous countries, air pollution is the most important environmental determinant of health, and the World Health Organization (WHO, 2016) estimates that about 7 million premature deaths are attributed annually to the effects of ambient and household air pollution. In order to improve their living standards, as well as healthcare and life expectancy, developing countries have had to prioritize industrial development. Overall, as emissions of air pollution rise, this leads to considerable detrimental changes in the environment and can thus cause negative health effects in people. (Brunekreef, 1997; Correia et al., 2013; Ebenstein et al., 2017; DeFelice, 2020).

During our extensive literature review, we have found that fine particulate matter ($PM_{2.5}$) carries the most association with adverse health outcomes, and therefore causing significant public health concerns (Ito et al., 2011; Ostro et al., 2007; Peng et al., 2009; Thurston et al., 2005; Zhou et al., 2011). Cohen et al. (2005) estimated that outdoor $PM_{2.5}$ air pollution is responsible for adult cardiopulmonary disease mortality (about 3%), trachea, bronchus, and lung cancer mortality (about 5%), and mortality in children under 5 years from an acute respiratory infection (about 1%) in urban areas worldwide.

A notable number of studies have elaborated the effect of fine particulate air pollution $PM_{2.5}$ on health (Pope and Dockery, 2006; Pope et al., 2009; Caccarelli et al., 2016), that take into account economic growth and other control factors. However, most of these studies are

* Corresponding author.
E-mail address: Dejan.Loncar@unige.ch (D. Lončar).

https://doi.org/10.1016/j.heliyon.2022.e10729
Received 29 January 2021; Received in revised form 22 November 2021; Accepted 16 September 2022

focused on short-term exposure to $PM_{2.5}$ in limited geographical settings. In addition, the methodologies adopted for this kind of analyses have not been able to address the issues of autocorrelation and endogeneity in their econometric models. Several studies have suggested that socio-economic factors play a significant role in the epidemiology of diseases and mortality associated with exposure to air pollution, as exposure often varies according to the socioeconomic status of the population (Dockery et al., 1993; Zanobetti and Schwartz, 2000; Bobak and Leon, 1999; Gwynn and Thurston, 2001). The objective of this study is to enhance the comprehension of the association between socio-economic indicators, human health and mortality which is caused by $PM_{2.5}$, while also providing insight related to which econometric models can be utilized to model this association.

In order to be able to incentivize international and national stakeholders to develop policies and instruments that will decrease the concentration of $PM_{2.5}$, it is crucial to analyze and document the negative effects of $PM_{2.5}$ air pollution on health and the economy. So far, significant financial resources, through various institutions and sectors, have been channeled to cope with the direct and indirect consequences of air pollution on health. Kunzli et al. (2001) explain that national and international efforts to mitigate the risks from air pollution should have a multi-sector approach, considering all socio-economic aspects of this significant issue through the prism of government, businesses, population and the wellbeing of the planet.

2. Methodology

In this study, we investigated the effects of $PM_{2.5}$ concentration on the population over 35 years of age on associated health outcomes while controlling for factors such as GDP, education, smoke intensity, inequality of the distribution of wealth, indoor pollution produced by the use of fossil fuels, population density in urban areas and national and health system regulatory capacities. The selection of variables for this study is based on theoretical concepts, an extensive literature review and the World Health Organization's air quality database that provides data disaggregated by ICD codes. Our primary covariate of importance is annual mean exposure to PM with a diameter size of 2.5 microns or less ($PM_{2.5}$). Additionally, $PM_{2.5}$ data is deemed merely as a proxy indication of air quality to inform cross-country comparisons of the health risks due to PM (World Bank Group). Our independent variables included in the regression model are: GDP PPP, HDI, Social Health Determinants and Health Risk Indicators, Smoking Prevalence, DTP coverage, GINI Index, WGI variables, Fossil Fuel consumption as percentage of total energy consumption, Urban/Rural percentage of population, Polity IV project indicators. The mortality data are mapped with the following ICD codes: B27, B29, B101, B323, B325 (ICD9) and I20–I25, I60–I69, C33, C34, J40–J44 and J47 (ICD10).

Mortality data contain a number of limitations which need to be addressed. Our unbalanced panel of mortality data consists of 91 countries (1397 data points) mostly from high and upper middle-income countries and limited number of data for 2015 (15 data points) from low-income countries. We have deliberated applying the projected mortality data from the Institute for Health Metrics and Evaluation (IHME) to increase the completeness of our dataset but this attempt was not used due to bias issues that may appear since this IHME and WHO methodology (Mathers and Loncar, 2006) employs identical independent variables to populate missing mortality data as we do in our regression analysis. In this study we proxied health system capacities and prioritization of health with the coverage of immunization Diphtheria-tetanus-pertussis (DTP), expressed as the percentage of immunized children ages 12–23 months from the World Bank Group, (Bos and Batson, 2000).

We addressed a few methodological challenges when approximating a relationship between $PM_{2.5}$ concentration and $PM_{2.5}$ related mortality and other socio-economic independent variables using cross-country data over time (Lončar et al., 2022). Independent variables could be potentially endogenous which means that we need to adopt appropriate instrumental variables estimation techniques to correct for endogeneity bias. Since there is a data set running across several countries over time, there may be unobserved heterogeneity in the data, and they have to be properly accounted for. We considered dynamic effects and tests for potential autocorrelation in the error term due to a presence of time series in our model.

Our model can be expressed in a general notation:

$$y_{it} = x_{it}'\beta + \varepsilon_{it} \quad i = 1 \ldots N; \quad t = 1 \ldots T; \tag{1}$$

where y_{it} denotes the $PM_{2.5}$-related mortality rate of country i in year t, x_{it}' is $(1 \times K)$ row vector of the observation (i,t) on all the explanatory variables mentioned above as well as the controls, and ε_{it} is the error term. In our study we used following econometric specifications: (i) A classical least squares regression (OLS) which assumes exogeneity of all x_{it}'s and *i.i.d.* ε_{it}, but it ignores the panel data nature, thus we need to use more advanced econometric estimators, (ii) Fixed and Random Effects models that can address panel data structure but not potential endogeneity in the model and (iii) Instrumental Variables method where we address the both endogeneity of our independent variables and panel data nature.

2.1. Introduction to panel data estimators

Before analysing panel data models it would be beneficial to discuss the nature of the unobserved effects. The challenge arises due to the fact that we have presumed the intercept α to be the same for all countries. Disregarding the individual or time-specific effects that occur among cross-sectional or time-series units can lead to inconsistent or meaningless estimates of the relevant parameters. Consider the following error component model shown in Eq. (2):

$$y_{it} = \alpha_i + x_{it}\beta + v_{it} \tag{2}$$

α_i are unobserved time-invariant individual effects; v_{it} is the error term, also not observed. x_{it} is a $(1 \times k)$ vector of regressors, and β is a $(k \times 1)$ vector of parameters to be estimated. We assume the following:

Assumption Panel #1. (Linear model) The model can be written as Eq. (2) where α_i is unobserved.
Assumption Panel #2. (Random sample) y_{it} and x_{it} are identically and independently distributed (i.i.d.)
Assumption Panel #3. (Identification) There is no perfect linear relationship among the explanatory variable (matrix X is non-singular).
Assumption Panel #4. (Spherical disturbances) (i) $V(\alpha_i|X) = \sigma^2$; (ii) $V(v_{it}|X) = \sigma^2 v$; (iii) $E(\alpha_i v_{it}|X) = E(v_{it}v_{is}|X) = 0$.

2.2. Instrumental variables (IV)

The Zero Conditional Mean assumption or exogeneity assumption is at the core of OLS unbiasedness and consistency. Whenever this assumption fails i.e., OLS estimator is biased and inconsistent as some of the regressors are correlated with the error term. This can arise for several reasons: sample selection, reverse causality, omitted variables or measurement error. To address endogeneity issues we use the instrument variables instead of endogenous variables in the model. Instruments are variables, which impact y but only indirectly, through their effect on endogenous variable x_k. Two conditions have to be satisfied by a valid instrument z, the instrument has to be correlated with the endogenous explanatory variable, and has to be uncorrelated with the error term.

The main notion of the IV estimation procedure is to take variation in the explanatory variable that matches up with variation in the instrument, so is uncorrelated with the error and uses only this variation to compute the slope estimate (Anderson and Hsiao, 1981).

IV Estimator: Provided that we have a random sample of observations on y, X, Z and Rank E (Z′X) = k, where IV conditions are satisfied, the instrumental variable estimator is defined as in Eq. (3):

$$\widehat{\beta_{IV}} = (Z'X)^{-1}Z'y \quad (3)$$

We can obtain the IV estimator using a two-step procedure:

1. Regress by OLS endogenous variable X on instrument Z: $X = \delta Z + \mu$
2. Use the predicted $\widehat{X}$ as the explanatory variable in the structural equation: $y = \widehat{X}\beta + u$

Using simple matrix algebra, we can show Eq. (4):

$$\widehat{\beta_{IV}} = (Z'X)^{-1}Z'y = (\widehat{X}'\widehat{X})^{-1}\widehat{X}'y \quad (4)$$

2.2.1. Limitations of IV estimation

The assumption for the validity of the instrumental variable z, whereby Cov (z,u) = 0, is partly subjective since its exogeneity cannot be fully tested statistically. We can, however, perform a number of tests to support the use of our instruments.

The following are some of the aspects and limitations which users of this methodology should be aware of. A weak instrument is a variable, which is weakly correlated with the endogenous regressor although the instrument might be still "relevant". The first issue is that weak instruments lead to more imprecise estimates. A second issue is that weak instruments exacerbate the bias. To confirm two conditions that have to be satisfied by a valid instrument z, we test the instrument for its validity using Sargan-Hansen's test and its strength using Anderson-Rubin under-id tests.

In order to test the relevance of the instruments to meet the rank condition we use a t-test or F-test on the instruments' coefficients in the first stage estimation. With more endogenous regressors (Lončar et al., 2019), we have to rely on the Cragg-Donald F-statistic, the F-statistic of the relevance of the instruments. A value of 10 or more for this F-statistic is a common rule-of-thumb, but it can be contingent on the size of the sample and the number of instruments (reported by IVREG2 command in Stata).

3. Results

3.1. Descriptive statistics

The descriptive statistics of our variables of interest are offered in Table 1.

The summary statistics of $PM_{2.5}$ related death rate, $PM_{2.5}$ concentration and gross domestic product (GDP) show that the skewness values are 0.7, 3.2 and 1.7 while kurtosis values are 3.1, 15.3 and 7.9 respectively, which shows a lack of symmetry and the presence of outliers in $PM_{2.5}$ concentration and GDP data. Regional variations of variables of interest are observed in our dataset. The highest mean of $PM_{2.5}$ related death rate is registered in Europe and Central Asia and the lowest in South Asia, 787 and 163 cases per 100.000 populations, respectively. As anticipated, the Middle East and North Africa (MENA) and South Asia have the highest average values of concentration of $PM_{2.5}$, 51 and 30 μg/m^3, respectively. The lowest level of $PM_{2.5}$ is registered in North America, with an average concentration of 9.4 μg/m^3. North America has the highest GDP based on purchasing power parity (PPP) of USD 39,341, while the lowest level of GDP, as anticipated is recorded in Sub-Saharan Africa of USD 11,994 and South Asia of USD 8,476. The results of this study ought to be analysed from the perspective of countries which were used in this report, as out dataset includes mostly high and upper-middle-income countries.

We postulate that certain income level variations exist in our data, therefore we summarized the variables by income level in Table 2. The highest mean of death rates attributed to $PM_{2.5}$ pollution is found in lower middle income countries, with 964 cases per 100,000 population, and the lowest level is in high income countries, with 622 cases per 100,000 population. This was anticipated when observing the data.

In the bivariate analysis we observed an unexpected negative trend between $PM_{2.5}$ concentration and the $PM_{2.5}$ related death rate. We observed that where $PM_{2.5}$ is less than 50 μg/m^3, there is no clear relation between $PM_{2.5}$ concentration and $PM_{2.5}$ related death rate. The second segment, where $PM_{2.5}$ is higher than 50 μg/m^3, demonstrates that this relationship has a negative trend. The countries that drive the negative trend in the second segment of the graph are: Tajikistan, Kuwait, Egypt, Bahrain, and Qatar.

Figure 1 shows time trends of $PM_{2.5}$ related death rates, $PM_{2.5}$ concentration and GDP in the dataset. The death rate attributed to $PM_{2.5}$ shows a slightly decreasing trend which is not something unforeseen, due to the fact that the majority of our data comes from high and upper-middle income countries where generally the mortality rate caused by $PM_{2.5}$ associated diseases are decreasing. Interestingly, the $PM_{2.5}$ mean shows quite a concave shape, showing initially a decreasing and after an increasing trend. As expected, the GDP means shows constantly an increasing trend.

3.2. Using available models

A Fisher test for unbalanced panel data was employed to test the stationarity of the time series of our variables. Under the null hypothesis, the Fisher's test assumes that all-time series are non-stationary against the alternative that at least one-time series within the panel is stationary. Fisher's test confirmed a sufficient level of stationarity and we rejected the null.

A simple OLS lin-lin model was employed to analyse the determinants of the dependent variable which is death rate attributable to $PM_{2.5}$ concentrations (per 100,000 individuals). Our primary independent variable of interest is the annual mean of $PM_{2.5}$ concentrations, expressed in μg/m^3 and another health risk, socio-economic and governance control variables such as: GDP PPP, smoking intensity and human capital index. The OLS reported negative and non significant association between $PM_{2.5}$ concentration and $PM_{2.5}$ related death rates, however smoking prevalence showed significant and positive association with $PM_{2.5}$ related death rates. Since the OLS model does not address panel

Table 1. Summary statistics.

Variable	N	Mean	SD	Midian	Min	Max
Death rate	1387	678	342	619	2	1753
$PM_{2.5}$	1387	21	16	17	5	122
GDP per capita PPP	1350	26392	19872	24008	1040	132514
Human capital	1387	10	2	10	3	13
Smoking prevalence	1384	0	0	0	0	0
Fossil use (%)	1322	77	19	82	10	100
Gini	1015	0	0	0	0	1
Urban population	1387	69	17	70	9	100
DTP coverage	1337	93	7	95	45	99
Polity	1274	18	5	20	1	21

Table 2. Tabular statistics for death rate by income level (in hundreds).

Income	Summary of Death rate		
	Mean	Std.Dev.	Freq.
High	6.2213176	2.9450498	935
Lower middle	9.6365257	4.3552476	138
Upper middle	7.1978764	3.5926168	314
Total	6.7821959	3.4200922	1,387

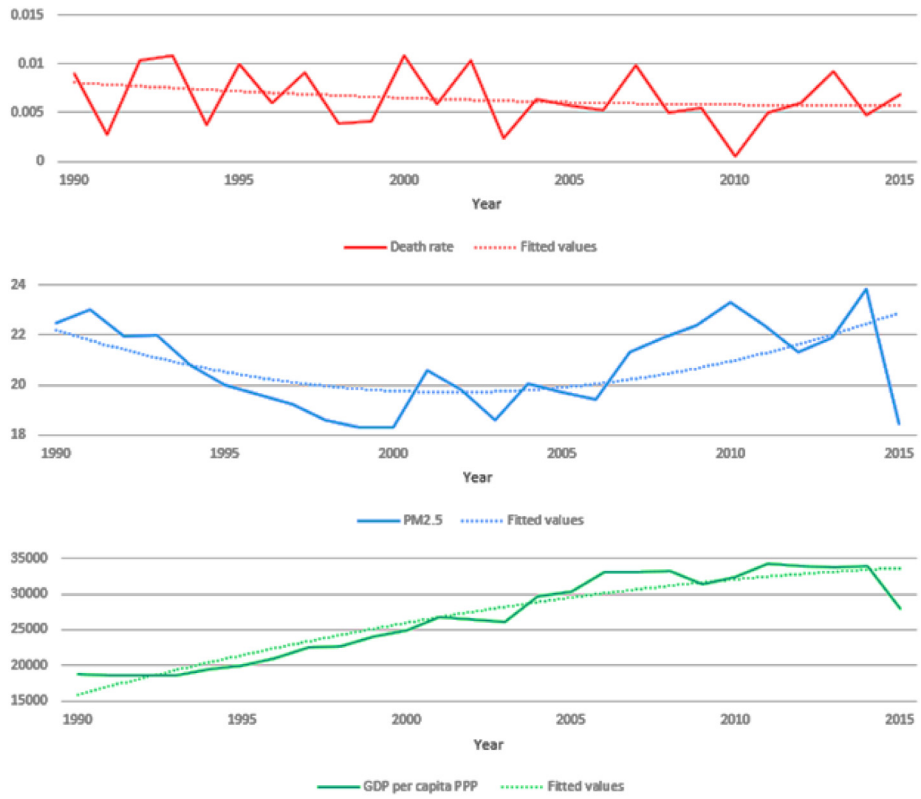

Figure 1. The trend of death rate, $PM_{2.5}$, GDP (World Bank Open Data).

data nature and endogeneity issues, we utilized more advanced models in order to widen our analysis.

Before looking at the Instrumental Variable (IV) models, we briefly report the results of the panel specifications, namely Fixed and Random Effects models, that keep GDP exogenous and enable us to gain more insights into the panel effects over the time. We used a Hausman test to compare the within estimation with the generalized least squares (GLS) approach. We rejected the absence of correlation of the specific effects with the other exogenous variables, and thus employed Fixed Effects (FE) models for the interpretation of the results. To improve normality and heteroscedasticity in the model we corrected the standard errors by using the "robust cluster" option in the regression models.

In order to regulate for the effects of indoor $PM_{2.5}$ pollution and government regulatory capacity on $PM_{2.5}$ related mortality, we included "percentage of use of fossil fuel" and "regulatory quality" variables. Both variables showed a significant association ($p < 0.001$) with $PM_{2.5}$-related mortality, with the estimated regression coefficient for the "percentage of the use of fossil fuel" variable ranging between 2.8 and 3.9 and the estimated coefficient for the regulatory quality variable varying between -68 and -82. The polity variable to proxy democratic and open governance systems is consistently found to be insignificant in the model.

3.3. Instrumental variable (IV) estimation

Finally, the instrumental variable (IV) models to address the issue of the possible endogeneity of GDP was employed. Our strategy was to use "Polity" variable to instrument GDP in the model as it is reasonable to belive that "Polity" variable is positively correlated with GDP. It has been shown in a significant number of studies that higher economic growth seems to be a trait of more democratic governments (Helliwell, 1994). A further motive for employing the "Polity" variable as an instrument, is that we have never found the polity variable to be statistically significant in any of our model estimations. We tested validity of the instrument and performed an over-identification test.

Table 3 reports the results of the IV regression model (IVREG1) estimations with the "polity" and lag of GDP variables used to instrument for GDP and address endogeneity issue. We find that the "Polity" and lag of GDP variables are valid, relevant and not a weak instrument. As the residuals in the model are not autocorrelated, we introduced a dynamic component in the model which is lag of $PM_{2.5}$ related death rates (DIVREG1). The IV regression models show that instrumented GDP is statistically significant ($p < 0.001$) and negatively associated with death rates attributed to $PM_{2.5}$ with coefficient estimates of -0.011 and -0.003, for the IVREG1 and DIVREG1 models, respectively. To be able to proxy the time of exposure to $PM_{2.5}$ concentration we used interaction term of Gini index and $PM_{2.5}$. Our hypothesis is that certain populations with varying income levels are subject to differing levels of $PM_{2.5}$. $PM_{2.5}$ concentration, the lag of $PM_{2.5}$ concentration, and interaction terms $PM_{2.5}$*Gini are not shown to be statistically significant. However, as anticipated, smoking prevalence and % use of fossil fuel demonstrate a significant (p

Table 3. Instrumental Variable (I–V) regression estimator results (lin-lin).

Variable	IVREG1		DIVREG1	
lag Death Rate	-	-	0.806***	(0.026)
$PM_{2.5}$	-7.14	(6.92)	-9.71*	(4.23)
lag $PM_{2.5}$	-0.641	(5.16)	8.20**	(3.16)
GDP PPP per capita	-0.011***	(0.001)	-0.003**	(0.001)
I.T. $PM_{2.5}$* Gini	-4.18	(4.34)	-2.50	(2.65)
Smoking Prevalence	871***	(194)	48.09	(122)
Human Capital	-26.1***	(6.04)	-1.05	(3.75)
Fossil Use	4.33***	(0.864)	0.780	(0.539)
DTP Coverage	-2.03*	(1.01)	-	-
Regulatory Quality	-90.5***	(17.0)	-14.7	(10.6)
R-squared	0.525		0.823	
N	622		622	

Note: *p<; 0.05, **$p < 0.01$, and ***$p < 0.001$.

< 0.001) and positive association with the mortality rate, with estimated regression coefficients of 872 and 4.33, respectively. Human capital ($p < 0.001$), regulatory capacity ($p < 0.001$) and diphtheria-tetanus-pertussis (DTP) coverage ($p < 0.05$) also revealed to have a statistically significant and negative association with the mortality rate, with coefficient estimates of −26, −91, and −2, respectively. The DIVREG1 model reveals *that the* lag of the death rate attributable to $PM_{2.5}$ pollution has a significant and positive association ($p < 0.001$) with mortality with a coefficient estimate of 0.81. In the DIVREG model, it can be seen for the first time that the lag of $PM_{2.5}$ concentration has a positive association with mortality with an estimated coefficient of 8.2 at a statistically significant level ($p < 0.01$). The log-transformed IVREG1 and DIVREG2 models do not produce results that vary significantly from those observed so far in the other regression models (Tables 3 and 4).

We use the same I–V models but with log transformed variables of interest for the instrumental variable regression models with and without the dynamic component (Table 4). To summarize the data, the lag of death rate is found to be significant at a level of $p < 0.001$ with a positive regression coefficient of 0.7. The results for the marginal effects of PM2.5 concentration on mortality follow the findings that we have observed in previous models. Specifically, we cannot observe the significance of PM2.5 concentration either in IVREG2 nor DIVREG2, although the signs of the regression coefficients are negative.

Furthermore, the lag of PM2.5 concentration exhibits a positive but insignificant association with mortality (DIVREG2). The results for GDP demonstrate a negative and statistically significant association with mortality ($p < 0.001$) for both models, with coefficient estimates of −0.28 and −0.09, respectively. The "smoking prevalence" and socio-economic indicators are found to be significant at the $p < 0.001$ level with the expected association with mortality attributed to PM2.5 (IVREG2): smoking prevalence (0.72), human capital (−0.5), % use of fossil fuels (0.36), DTP coverage (−0.5) and regulatory quality (−0.18) at a significance level of $p < 0.05$. The socio-economic variables in the DIVREG2 model are found to have a similar association with mortality.

Consequently, we use the Arellano-Bond (ABOND) method for the endogenous covariate in the model as a dynamic panel model specification. The models ABOND1 consider GDP to be an exogenous variable, while the ABOND2 models' instrument GDP with the polity variable and the lag of the GDP (Table 5). As anticipated, for all variants of the dynamic panel models that we inspected, the lag of death rate is found to be statistically significant at a high level ($p < 0.001$), with a positive regression coefficient varying between 0.38 and 0.46. $PM_{2.5}$ concentration and the lag of $PM_{2.5}$ concentration in most of the models are not found to be statistically significant, showing a mixed association with mortality. The variables such as GDP, human capital, smoke prevalence and fossil use give consistently statistically significant results with effects on $PM_{2.5}$ related mortality as we have anticipated.

Table 4. Instrumental Variable (I–V) regression estimator results (log-log).

Variable	IVREG2		DIVREG2	
ln (lag Death Rate)	-	-	0.704***	(0.025)
ln ($PM_{2.5}$)	-0.164	(0.176)	-0.168	(0.116)
ln (lag $PM_{2.5}$)	-0.374*	(0.155)	0.056	(0.104)
I.T. ln ($PM_{2.5}$* Gini)	0.192*	(0.083)	0.006	(0.055)
ln (GDP PPP per capita)	-0.280***	(0.034)	-0.094***	(0.024)
ln (Human Capital)	-0.500***	(0.083)	-0.154**	(0.056)
ln (Smoke Prevalence)	0.718***	(0.051)	0.187***	(0.039)
ln (Fossil Use)	0.360***	(0.062)	0.112**	(0.042)
ln (DTP Coverage)	-0.500***	(0.118)	-	-
ln (Regulatory Quality)	-0.178	(0.077)	-0.058	(0.051)
R-squared	0.630		0.838	
N	622		622	

Note: *p < 0.05, **p < 0.01, ***p < 0.001.

Table 5. Arellano-Bond regression results (log-log).

Variable	ABOND-1		ABOND-2	
log ln (Death Rate)	0.377***	(0.044)	0.459***	(0.012)
ln ($PM_{2.5}$)	-0.200	(0.116)	-0.036	(0.029)
ln (lag $PM_{2.5}$)	0.100	(0.096)	-0.030	(0.045)
I.T. ln ($PM_{2.5}$* Gini)	-0.003	(0.077)	-	-
ln (GDP PPP per capita)	0.352	(0.438)	-0.108***	(0.013)
ln (SQ GDP PPP per capita)	-0.024	(0.023)	-	-
ln (Human Capital)	-0.124*	(0.063)	-0.081***	(0.015)
ln (Smoke Prevalence)	0.449***	(0.075)	0.423***	(0.054)
ln (Fossil Use)	0.215***	(0.045)	0.269***	(0.025)
ln (Regulatory Quality)	-0.051	(0.080)	-0.018	(0.019)
ln (Gini)	-	-	0.009	(0.007)
Constant	3.29	(2.156)	4.43***	(0.214)
N	429		408	

Note: *p < 0.05, **p < 0.01, ***p < 0.001.

4. Discussion

In this study we investigated our assumptions concerning the association of $PM_{2.5}$ concentration, socioeconomic, health risk and system and governance indicators on $PM_{2.5}$ related mortality, controlling for time of $PM_{2.5}$ exposure where these variables were operationalized by $PM_{2.5}$ emission, GDP, Gini, education, smoking intensity, immunization coverage, ratio of urban population, and government regulatory capacity. The income level of population was controlled by GDP, and the fossil use variable was included to control effects of industrial development and indoor $PM_{2.5}$ pollution. The effect of human capital on mortality has been very well documented in literature, so we utilized the education variable to control this association. Numerous studies revealed a strong association between risk factors such as smoking prevalence and mortality (Mathers and Loncar, 2006) thus we employed this variable in the model. As argued, the population with lower income levels is exposed to higher levels of $PM_{2.5}$ emissions for longer periods of time and consequently, will have more negative health consequences of $PM_{2.5}$ emission. Population density and industrialization were the primary reasons for influencing heightened PM 2.5 pollution (Xiangxue et al., 2021). A study in China demonstrated that population accumulation, industrialization, foreign investment, transportation, and pollution emissions added to the escalation of $PM_{2.5}$ concentration. (Yazhu et al., 2019). Analysing the results of a study in California, authors concluded that future mortality studies should reflect on adjusting for differences with rural–urban variables (Garcia et al., 2015). In this study we postulate that a country with a proportionally larger urban population, lower income and a more significant inequality in wealth distribution is subject longer to higher levels of $PM_{2.5}$ emissions. Thus, we consider to proxy the time of exposure to $PM_{2.5}$ concentration with the Gini index to address inequity of wealth distribution and share of urban population to control the ratio of urban rural population in the country. It is reasonable to postulate that the population with lower levels of wealth distribution is exposed to higher levels of $PM_{2.5}$ concentration, thus we use the iteration term as a product of Gini and $PM_{2.5}$ in our model. Lastly, our literature review provided strong evidence that health system capacities proxied by immunization coverage and regulatory capacity carries a significant association with mortality rates.

The study has demonstrated significant effects of socioeconomic, health risk and system and governance variables on the relation between $PM_{2.5}$ concentration and $PM_{2.5}$ related mortality with less obvious and mixed evidence of the direct effect of $PM_{2.5}$ concentration on $PM_{2.5}$ related mortality.

Initial models, OLS and FE, unexpectedly exposed that the relation between $PM_{2.5}$ concentration expressed in μg/m and mortality rates

attributed to $PM_{2.5}$, has a negative coefficient at a statistically significant level, while more robust and sophisticated estimators that control potential heteroscedasticity, autocorrelation and endogeneity show that this relation is not statistically significant with mixed results. Similar results were reported in several empirical studies, i.e., Venners et al. (2003). However, this result was not consistent during the entire study as it can be observed that $PM_{2.5}$ concentration has a various effect on mortality on differing levels of GDP and $PM_{2.5}$ concentration.

In the more advanced dynamic instrumental regression model, an important finding we observed is that the lag of $PM_{2.5}$ concentration has a positive association with mortality with an estimated coefficient of 8.2 at a statistically significant level ($p < 0.01$). This finding highlights that an increased level of $PM_{2.5}$ concentration on average increases the $PM_{2.5}$-related mortality estimate. The results show that GDP per capita PPP shows a consistently strong negative effect on mortality rate attributed to $PM_{2.5}$ at high levels of statistical significance ($p < 0.001$). We can approximate from more advanced models, while controlling the endogeneity and autocorrelation that for each increase of US$ 10,000 in annual GDP per capita, between 20 and 110 deaths in 100,000 in population will be decreased on average. Smoking prevalence was documented as one of the main risk factors for cardiovascular, respiratory and lung cancer causes of death (Mathers and Loncar, 2006). This study confirms that smoking prevalence is consistently, strongly and positively associated with the mortality rate at high significance level of $p < 0.001$ across all regression models. As our long term study has exhibited a mixed result of association between $PM_{2.5}$ concentration and $PM_{2.5}$ related mortality we postulate that a strong effect of smoking intensity in the models diminished the effect of $PM_{2.5}$. We estimate that an increase of smoking prevalence of 1 % in total population will increase on average between 668 and 927 deaths per 100,000 of a population. The average year of schooling, as foreseen, shows a negative association at a statistically significant level on mortality in most of the models. Average education shows a negative association with mortality and based on the coefficient estimates, we can anticipate that every 1-year increase in a country's average years of schooling reduces the $PM_{2.5}$-related mortality rate by on average between 3 and 26 deaths per 100,000 individuals. Moreover, as previously discussed, the $PM_{2.5}$ related health outcomes are caused by both indoor and outdoor $PM_{2.5}$ pollution. Since our $PM_{2.5}$ covariate measures the average value of outdoor pollution we proxy the effects of indoor pollution by percentage of use of fossil fuel. Our hypothesis is that a higher percentage of use of fossil fuels increases the population exposure to indoor $PM_{2.5}$ pollution. The log-log regression models also demonstrate a positive and statistically significant relationship between indoor $PM_{2.5}$ pollution proxied by use of fossil fuels and $PM_{2.5}$-related mortality rate with a regression coefficient between 0.1 and 0.36. The government regulatory capacity to develop and implement policy and institutional reforms has emerged as an important area that can drive sectorial improvements. Thus, the abatement of health outcomes through either the reduction of $PM_{2.5}$ concentration or mitigation of any other associated health risk factors might be a direct consequence of good regulations. Our analysis shows that a strong government regulatory capacity affects the decrease in the $PM_{2.5}$-related mortality rate. A strong health system and government commitment to health through investment in new technologies, human capital for health, research and development, medicines and prevention activities, reduce negative health outcomes and extend human life expectancy.

5. Limitations

Our study design encompasses certain limitations, which should be addressed. Firstly, one of the limitations of this study is that we do not take into account the amount of time a person has been exposed to $PM_{2.5}$. Studies have shown that there are also significant differences in exposure of populations in different areas, namely urban and rural. Additionally, we haven't been able to consider the individual characteristics of the population (previous illnesses, gender, etc.). In this study, our variable is the average, which can be a concern. Our main strategy in this document was to use proxy variables which control for the unobserved factors.

6. Conclusions and future perspective

Evidence was found that our socio-economic factors and risk health factors have a larger impact on $PM_{2.5}$ related mortality. It is observed that the relation between $PM_{2.5}$ emission and health outcomes cannot be modelled without socio-economic and health risk factors. The investments in mitigation of negative consequences of $PM_{2.5}$ concentration need to be integrated with policy instruments that will address risks associated with socio-economic factors. This reiterates the importance of interconnection in a variety of sectors, which then provide an opening to enhance new ways to reach the SDGs.

Our study demonstrates that more advanced econometric models can better model the connection between annual mean $PM_{2.5}$ concentrations and mortality rate attributable to $PM_{2.5}$. Initial models show that this relationship is negative at a statistically significant level, while more sophisticated estimators that can control for heteroscedasticity, autocorrelation and endogeneity in the model show that this relationship is not statistically significant with mixed results. However, it is vital to report that lag of $PM_{2.5}$ concentrations increase mortality rate attributable to $PM_{2.5}$.

Current research studies concentrate mostly on the effect of air quality on mortality, and much less on morbidity. Further studies would be required to investigate the extent to which morbidity in specific groups may be associated with changes in $PM_{2.5}$ related mortality controlling for socioeconomic, health and governance indicators.

This article gives a useful analytical structure for analyzing the health effects of $PM_{2.5}$ exposure with advanced econometric estimators that can accommodate autocorrelation and endogeneity issues in the panel data and confirms significant association of lagged $PM_{2.5}$ concentration, socio-economic, health and governance indicators on $PM_{2.5}$ related mortality.

Declarations

Author contribution statement

Dejan Lončar: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Vesna Krstić: Performed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Nicholas Brown Tyack: Performed the experiments; Contributed reagents, materials, analysis tools or data.

Jane Paunković: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Not applicable as all data was extracted from publicly available databases.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

The authors are thankful to University LUM, Bari, Italy and WHO Geneva, Switzerland, for the helpful information during the preparation of this research.

References

Anderson, T., Hsiao, C., 1981. Estimation of dynamic models with error components. J. Am. Stat. Assoc. 76 (375), 598–606.

Bobak, M., Leon, D., 1999. The effect of air pollution on infant mortality appears specific for respiratory causes in the postneonatal period. Epidemiology 10, 666–670.

Bos, E., Batson, A., 2000. Using immunization coverage rates for monitoring health sector performance, measurement and interpretation issues. In: HNP Discussion Paper.

Brunekreef, B., 1997. Air pollution and life expectancy: is there a relation? Occup. Environ. Med. 54, 781–784.

Caccarelli, A.A., Hales, N., Burnett, R.T., Jerrett, M., Mix, C., Dockery, D.W., Pope, C.A., 2016. Particulate air pollution, exceptional aging, and rates of centenarians: a nationwide analysis of the United States, 1980−2010. Environ. Health Perspect. 124, 1744–1750.

Cohen, A.J., Ross Anderson, H., Ostro, B., Dev Pandey, K., Krzyzanowski, M., Künzli, N., Gutschmidt, K., Pope, A., Romieu, I., Samet, J.M., Smith, K., 2005. The global burden of disease due to outdoor air pollution. J. Toxicol. Environ. Health, Part A 68, 1301–1307.

Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope III, C.A., Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C.J.L., Forouzanfar, M.H., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. Lancet 389, 1907–1918.

Correia, A.W., Pope III, C.A., Dockery, D.W., Wang, Y., Ezzati, M., Dominici, F., 2013. Effect of air pollution control on life expectancy in the United States.An Analysis of 545 US counties for the period 2000 to 2007. Epidemiology 24, 23–31.

DeFelice, T.P., 2020. Relationship between temporal anomalies in $PM_{2.5}$ concentrations and reported influenza/influenza-like illness activity. Heliyon 6, e04726.

Dockery, W.D., Pope III, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris Jr., B.G., Speizer, F.E., 1993. An association between air pollution and mortality in six U.S. Cities. N. Engl. J. Med. 329, 1753–1759.

Ebenstein, A., Fan, M., Greenstone, M., He, G., Zhou, M., 2017. New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. Proc. Natl. Acad. Sci. U.S.A. 114, 10384.

Garcia, Cynthia A., Yap, Poh-Sin, Weller, Hye-Youn Park & Barbara L., 2015. Association of long-term $PM_{2.5}$ exposure with mortality using different air pollution exposure models: impacts in rural and urban California. Int. J. Environ. Health Res. 26 (2), 2016.

Gwynn, R., Thurston, G., 2001. The burden of air pollution: impacts among racial minorities. Environ. Health Perspect. 109 (suppl 4), 501–506.

Helliwell, J.F., 1994. Empirical linkages between democracy and economic growth. Br. J. Polit. Sci. 24 (2), 225–248. https://www.mcc.gov/who-we-fund/indicator/immunization-rates-indicator. https://www.k4health.org/toolkits/immunization-trm/key-indicators.

Ito, K., Mathes, R., Ross, Z., Nadas, A., Thurston, G., Matte, T., 2011. Fine particulate matter constituents associated with cardiovascular hospitalizations and mortality in New York City. Environ. Health Perspect. 119, 467–473.

Künzli, N., Medina, S., Kaiser, R., Quenel, P., Horak Jr., F., Studnicka, M., 2001. Assessment of deaths attributable to air pollution: should we use risk estimates based on time series or on cohort studies? Am. J. Epidemiol. 153, 1050–1055.

Lončar, D., Paunković, J., Jovanović, V., Krstić, V., 2019. Environmental and social responsibility of companies across EU countries - panel data analysis. Sci. Total Environ. 657, 287–296.

Lončar, D., Krstić, V., Brown Tyack, N., Jovanović, M., Paunković, J., 2022. Economic and technological aspects of the impact of $PM_{2.5}$ particles on human health and productivity. Zaštita Materijala 63 (2), 135–145.

Mathers, C., Loncar, D., 2006. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 2011–2030.

Ostro, B., Feng, W., Broadwin, R., Green, S., Lipsett, M., 2007. The effects of components of fine particulate air pollution on mortality in California: results from CALFINE. Environ. Health Perspect. 115 (1), 13–19.

Peng, R., Bell, M., Geyh, A., McDermott, A., Zeger, S., Samet, J., Dominici, F., 2009. Emergency admissions for cardiovascular and respiratory diseases and the chemical composition of fine particle air pollution. Environ. Health Perspect. 117, 957–963.

Pope, C.A., Dockery, D.W., 2006. Health effects of fine particulate air pollution: lines that connect. J. Air Waste Manage. Assoc. 56, 709–742.

Pope, C.A., Burnett, R.T., Krewski, D., Jerrett, M., Shi, Y., Calle, E.E., Thun, M.J., 2009. Cardiovascular mortality and exposure to airborne fine particulate matter and cigarette smoke: shape of the exposure-response relationship. Circulation 120, 941–948.

Thurston, G., Ito, K., Mar, T., Christensen, W., Eatough, D., Henry, R., Kim, E., Laden, F., Lall, R., Larson, T., Liu, H., Neas, L., Pinto, J., Stölzel, M., Suh, H., Hopke, P., 2005. Workgroup report: workshop on source apportionment of particulate matter health effects—intercomparison of results and implications. Environ. Health Perspect. 113, 1768–1774.

Venners, S., Wang, B., Peng, Z., Xu, Y., Wang, L., Xu, X., 2003. Particulate matter, sulfur dioxide, and daily mortality in Chongqing, China. Environ. Health Perspect. 111, 562–567.

WHO, 2016. Ambient Air Pollution: A Global Assessment of Exposure and burden of Disease. In: World Bank Open Data, Free and Open Access to Global Development Data. World Health Organization, Geneva, Switzerland. https://data.worldbank.org/.

Xiangxue, Z., Yue, L., Changxiu, C., Junming, L., 2021. Determinant powers of socioeconomic factors and their interactive impacts on particulate matter pollution in North China. J. Environ. Res. Public Health 2021 (18), 6261.

Yazhu, W., Xuejun, D., Lei, W., 2019. Spatial-Temporal evolution of $PM_{2.5}$ concentration and its socioeconomic influence factors in Chinese cities in 2014–2017. Int. J. Environ. Res. Publ. Health 16, 985.a.

Zanobetti, A., Schwartz, J., 2000. Race, gender, and social status as modifiers of the effects of PM10 on mortality. J. Occup. Environ. Med. 42 (5), 469–474.

Zhou, J., Ito, K., Lall, R., Lippmann, M., Thurston, G., 2011. Time-series analysis of mortality effects of fine particulate matter components in Detroit and Seattle. Environ. Health Perspect. 119, 461–466.