



HAL
open science

Does particulate matter pollution affect school attendance in South Africa?

Macoura Doumbia

► **To cite this version:**

Macoura Doumbia. Does particulate matter pollution affect school attendance in South Africa?. 2024. hal-04446532

HAL Id: hal-04446532

<https://univ-orleans.hal.science/hal-04446532v1>

Preprint submitted on 27 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Document de Recherche du Laboratoire d'Économie d'Orléans
*Working Paper Series, Economic Research Department of the University of Orléans, University Of Tours
and University Clermont-Auvergne, France*

DR LEO 2024-02



Does particulate matter pollution affect school attendance in South Africa?

Macoura DOUMBIA

MISE EN LIGNE / ONLINE | 08 /02/2024

Laboratoire d'Économie d'Orléans
Collegium DEG
Rue de Blois - BP 26739
45067 ORLÉANS Cedex 2
TÉL | (33) (0)2 38 41 70 37
MAIL | leo@univ-orleans.fr
www.leo-univ-orleans.fr

Does particulate matter pollution affect school attendance in South Africa?

Macoura DOUMBIA

Laboratoire d'Economie d'Orléans-Université Clermont Auvergne,
26 Av. Léon Blum, 63000 Clermont-Ferrand, FRANCE.

January 2024

Abstract

Rising air pollution is a growing problem in South Africa. This study aims to find out how pollution affects school attendance in South Africa. By matching survey and satellite pollution data and using an instrumental variable approach over the period 2015-2016, this paper finds evidence that the risk of absenteeism is larger for pupils exposed to fine particulate matter (PM_{2.5}) and especially for those aged between 5 and 15 years old. This finding is consistent across provinces but varies with age of individuals and household's characteristics such as wealth level, household size and structure, and household head subgroup. The study highlights the causal relationship between air pollution and school absenteeism at least in South Africa which has up to here remain mixed in the literature.

Keywords : *PM 2.5, ambient air pollution, school attendance, South Africa*

JEL Classification : I15, I18, Q53, Q56

1 Introduction

According to World Health Organization, more than 90 per cent of people live in a place that does not meet air quality standards, and the part of the world exposed to particulate pollution is mostly made up of developing countries (WHO, 2016). The greatest environmental risk on the African continent is air pollution (THE BORGES PROJECT, 2020). In particular, South Africa is one of the most polluted countries in the world in 2019 (IQAIR, 2020). Two types of pollution undermine the South African environment : air pollution and marine pollution. The first source of air pollution is the industrial sector. This is because South Africa is a newly industrialised country according to the World Bank and has a relatively large contribution of its industrial sector to the national GDP. It ranks 33rd in the world in terms of GDP, has a diversified industrial sector. And then, its manufacturing sector is dominated by agriprocessing industry, automotive industry, chemicals industry, Information and Communications Technology and electronics industries, metals industry, textiles and clothing and footwear industry. The development of the sector explains the high level of pollution in the country. The energy mix could also explain the country's air pollution levels. Indeed, most of the energy consumed is generated by coal and oil (Ritchie et al., 2022). In 2022, coal will account for 70% of primary energy (Ritchie et al., 2022).

Regarding this situation South African government put in place 2004 Air quality act which is a law for air quality management including control at the source until its impact on people. The Act has established a list of pollutants and polluting activities, which has resulted in the establishment of certain standards. In 2009, emission limits were set for sulphur dioxide (SO₂), nitrogen dioxide (NO₂), particles with an aerodynamic diameter of less than 10 microns metres (PM₁₀), ozone (O₃), benzene (C₆H₆) and lead (Pb). Then, in view of the adverse public health effects of fine particulate matter (PM_{2.5}), these are added to the list in 2012. In 2013, the list of highly polluting activities includes 10 categories of activities which are combustion installations, petroleum industries, carbonization and gasification, metallurgical industry, mineral processing, storage and handling, organic chemicals industries, inorganic chemicals industries, thermal treatment of hazardous and general waste, pulp and paper manufacturing industries, and animal matter processing with emission limits defined for each.

The environmental efficiency of South Africa's regulation on air quality depends on its impact on health and more specifically on human capital. The objective of this study is to study the impact of air pollution on one of the most vulnerable individuals : children. Indeed, children are more exposed to pollution because of their height since pollutants are more concentrated at one (01) meter than at two (02) meters from the ground (Umwelthilfe, 2019) and early exposure air to pollution has more damaging effects (Pope III, 2002).

PM 2.5 from anthropogenic or natural sources is very harmful to human health and the cause of mortality from respiratory and cardiovascular diseases (Kampa and Castanas, 2008). By affecting human health, it can affect economic performance of countries through its impact on human capital. In this study, I am particularly interested in the relationship between air pollution and school absenteeism. Previous studies test the link between particulate pollution and school absenteeism and find a mixed effect. The impact may be positive (Ransom and Pope III, 1992; Makino, 2000; Park

et al., 2002; Ransom and Pope, 2013; MacNaughton et al., 2017), negative (Chen et al., 2000) or null (Gilliland et al., 2001; Rondeau et al., 2005).

This study differs from the previous ones in two ways. First difference is the sample. Contrary to the previous studies, this study focuses on an African developing country which is one of the most polluted by PM2.5 in the world (WHO, 2016) and with less regulation. The absence of African countries as a sample in this literature could be explained by the absence or poor quality of the data (Kalisa et al., 2023). The second distinction concerns the methodology used. The choice of residence supposedly leads to a non-random distribution of exposure to air pollution. As a result, PM2.5 concentration is potentially subject to an endogeneity bias. Another originality of this study is to take into account this endogeneity bias using control function method of Wooldridge (2015) which is similar to two stage least squares (2SLS) approach but with some interesting advantages over 2SLS. The question this study is trying to answer is what is the effect of PM exposure on school absenteeism. The interest of this study is to help us to better understand the impact of air pollution on school attendance in an African developing country.

In order to answer our research question, we use data from the General Household Survey of South Africa over the period 2015-2016. The variable of interest is constructed using data from van Donkelaar et al. (2018). This study is the first to use these data to study the relationship between particulate pollution and school absenteeism. Using a random effects logit model and an instrumental variable method, results show an increase in PM2.5 concentration by $10 \mu\text{g}/\text{m}^3$ results in a 1.33 to 2.46 percent risk of absenteeism. However, the impact depends on individual and household characteristics.

The next section reviews the existing literature. Data, stylized facts and hypotheses are presented in the third section. The fourth section describes the econometric analysis. Then, the fifth and sixth sections deal respectively with the transmission channels and robustness checks. Section 7 shows heterogeneity of the impact. The last part discusses conclusion and policy implications.

2 Background

This section provides a summary of the literature on the characteristics of the South African education system, the determinants of school absenteeism in South Africa, the relationship between air pollution and health and the impact of air pollution on school absenteeism.

2.1 South African Education System

"Apartheid" introduced in 1948 and abolished on June 30, 1991 did not spare the South African education system. Indeed, this policy of "separate development" led to a division and unequal distribution of the educational system by skin color (the "blacks" on one side and the "whites" on the other) (Vambe, 2005). This unequal educational system trained children differently because training programs differed according to the positions that child is expected to have in social, economic and political life in the future. As a result, best training programs were reserved for "whites" (Education Department, 2002). Immediately after apartheid, South African government introduced a unique educational program to

reduce educational inequalities created during racial segregation ([Education Department, 2002](#)).

After the election of Jacob Zuma as president in 2009, the Ministry of Education was split in two. South African education sector is now shared between the Ministry of Basic Education and the Ministry of Higher Education and Training. The education system is divided into three successive levels of education: primary education, secondary education and tertiary education. Primary students are aged between 7 and 15 years old. In 2001, the average age of admission to grade 1 was set at 7 years by the Education Department ([Education Department, 2001](#)). As in most countries, in South Africa the private sector also provides educational services.

In addition, there are major geographical inequalities in the education system. In 2007, 41% of rural students in grade 6 which is the exit-level of the Intermediate Phase, performed poorer compared to 13% of students of the same level who live in cities. In addition to the inequalities within the country, grade 6 students have a lower level of education than those in Kenya or Tanzania for example ([Spaull, 2015](#)). Furthermore, school dropout is not negligible. For example, in 2014, about 50% (532,860) of the students in the cohort took exams (Matric (grade 12)) while there were 1,085,570 students enrolled in the 12-year old cohort, i.e. in grade 1 ([Spaull, 2015](#)). This school dropout is mainly related to financial constraints and pregnancy, which explains 50% of the dropout rate among girls ([Gustafsson and al., 2011](#)).

The Ministry of "Basic Education" has put in place various policies that would allow the proper functioning and efficiency of the education system. First, since the health status of students is paramount for their academic success, policies have been put in place to reduce the negative effects of diseases such as HIV, tuberculosis and sexually-transmitted infections on learners (students), educators and all those working in schools ([Basic Education Department, 2017a](#)). These policies aim to improve the academic achievement of students in a health context weakened by HIV, sexually-transmitted infections and tuberculosis. Also, punishment in the context of school learning, the Ministry of Basic Education is implementing a protocol regarding corporal punishment. This protocol prohibits corporal punishment and provides guidance to schools, provinces, and districts on how to deal with issues related to corporal punishment ([Basic Education Department, 2017b](#)). Then, policies implemented since 2001 include the Integrated Quality Management System (IQMS), which aims to contribute to the professional development of teachers. The goal is to improve learning in schools ([Mestry et al., 2009](#)). Furthermore, in 1995, in order to maintain a healthy working environment for teachers, the Department of Labour set up an Education and Labour Relations Board. The primary goal is to maintain and promote labour peace in education ([Labour Department, 1995](#)). According to a joint study by International Labour Organization and UNESCO, the Education and Labor Relations Board is a framework for social dialogue. Only African countries such as Namibia and South Africa have been able to set up this kind of institution to promote social dialogue ([Hilsdon and Randell, 2012](#)). In addition, in order to improve academic and social conditions of learners, the Ministry of Basic Education is putting in place indicators to determine whether a learner is a victim of sexual, drug or alcohol abuse ([The Education Labour Relations Council, 2003](#)).

According to literature, South African education system is structured with policies in place to

provide a healthy working environment for teachers and learners. Unfortunately, these policies are insufficient in terms of health protection against environmental degradation and more specifically air pollution. One of the main objectives of our study would be to draw the attention of South African policy-makers to the importance of air quality that students breathe and its adverse health effects that affect their school performance.

There are several factors that could explain absenteeism and the following section gives an idea of the determinants of truancy in South Africa.

2.2 Determinants of school absenteeism and performance in South Africa

School absenteeism has several drivers at both the social and economic levels. First, poverty is an important determinant. Authors such as [Adato et al. \(2016\)](#) study the effect of cash transfers on school participation in South Africa. They use the Child Support Grant program as a proxy for cash transfers. The goal of this subsidy program for poor households with young children is to provide social protection and reduce poverty and inequality between rich and poor children in terms of nutrition and to reduce early pregnancy and risky sexual behavior. Their results show that having a scholarship has a positive but partial effect on students' school participation. The effect of poverty on school performance is also proven by [Hunter and May \(2003\)](#). The authors add that this is a controversial determinant because a family's poverty level can push it to keep the child in school because the child represents family's future. That is the only asset that the family holds. [Timæus et al. \(2013\)](#) based on data from "the National Income Dynamics Study" collected at the community level in 2008 also show that the household poverty level also determines children's school participation. In this study we don't use income level of households as control because of potential endogeneity bias the child's absence could be due to the fact that he or she is working and thus generating income for the household.

Furthermore, a second determinant of school performance in South Africa is the presence of biological parents in the household. The death of one of the biological parents, especially the mother, has a negative impact on children's school performance. In a study conducted in the province of KwaZulu Natal, [Case and Ardington \(2006\)](#) show that the death of a father or mother has a different impact on school performance. Maternal orphans appear to perform poorly academically compared to other children their age whose mothers are alive. They are also less likely to be enrolled in school and to complete several years of schooling. Paternal death has no significant effect on the child's school performance. According to [Operario et al. \(2008\)](#), the effect of parental death differs depending on whether the child is a girl or a boy. Their study shows that the absence of the mother or father has a greater effect on the schooling of girls than boys. The study finds that orphan girls are less likely to complete school than orphans after controlling for household poverty level. [Timæus and Boler \(2007\)](#) find maternal absence does not significantly affect a child's school performance. On the other hand, paternal absence has an effect. This can be explained by the fact that paternal absence could affect the household's poverty level.

In addition, mother's level of education has a positive effect on the child's academic success in South Africa ([Timæus and Boler, 2007](#); [Timæus et al., 2013](#); [Visser et al., 2015](#)). A well-educated

mother is more likely to get her children into school. Indeed, parents with a high level of education value education more and consider education as a good way to succeed in life. Therefore, they monitor child's school education. In other words, this kind of parent is more involved in the child's academic success.

Also, school resources made available to students determine their success at school ([Case and Deaton, 1999](#); [Visser et al., 2015](#)). Socio-economic factors such as the pupil-teacher ratio, which is the ratio of the number of pupils to the number of teachers, and the size of classes have a significant impact on pupils' schooling and academic results.

Health and nutrition profile of students is an important determinant of school performance in Nelson Mandela's country ([Jinabhai et al., 2001](#)). Findings from the study reveal micronutrient deficiencies, parasitic infections, and stunting remain significant problems among school-aged children in rural South Africa. Addressing these problems would promote children's health and education. In addition, school tardiness is a major determinant of school dropout ([Branson et al., 2014](#)). Indeed, repeating or resuming classes pushes some students to drop out of school for good.

To better understand the channel through which air pollution affects school absenteeism, it is very important to focus on the effects of air pollution on human health. This is the objective of the following section.

2.3 Local air pollution and health outcomes

Air pollution is a serious but neglected problem in Africa ([Tanmowo, 2000](#)) and is among the main causes for death in Africa ([Bauer et al., 2019](#)). The main sources of pollution include road traffic, industrial activities, biomass burning ([Petkova et al., 2013](#)). In recent decades, researchers are beginning to focus on its effects on the African population living in the most vulnerable countries due to the lack of good pollution regulation policies. The impacts on health are not negligible. The majority of researchers are interested in the health impacts of this atmospheric deterioration in Africa. This degradation of air quality affects life expectancy, causes cardiovascular diseases and mortality ([Lelieveld et al., 2018](#); [Noubiap et al., 2015](#); [Mustapha et al., 2011](#)). [Bauer et al. \(2019\)](#) looks at the effects of air pollution on mortality in Africa according to the sources of pollution (natural, industrial or domestic and biomass combustion). The results show that about 43,000 premature deaths in Africa are related to biomass burning due to agriculture in 2016. In South Africa and Nigeria, the high mortality rate due to pollution is explained by the industrial sector ([Bauer et al., 2019](#)).

Furthermore, [Roy \(2016\)](#) looks at the significant economic cost to Africa of ambient particulate pollution. In particular, the economic cost related to premature deaths. Indeed, the number of premature deaths as a result of ambient pollution increased by 36% between 1990 and 2013, and those related to household pollution increased by 18% over the same period. The estimated economic cost of premature deaths due to ambient particulate air pollution is around 215 billion US dollars and the cost due to household air pollution is around 232 billion US dollars.

Air pollution has important health effects on the South African population. It causes respiratory diseases, cerebrovascular (Wichmann and Voyi, 2012) and acute respiratory diseases in children (Barnes et al., 2009). Most of the studies done on the subject in South Africa link (indoor) air pollution to the source of energy used by the households. These studies show that the use of solid fuels (Barnes et al., 2009; Norman et al., 2007; Röllin et al., 2004; Vanker et al., 2015) or tobacco (Elf et al., 2017) leads to an increase in the level of indoor pollution which results in respiratory diseases. In 2007, Norman et al. (2007) investigate the relationship between indoor pollution from solid fuel use and respiratory disease in South Africa. The study finds that about 20% of South African households are exposed to indoor pollution from heavy fuel use. This pollution is responsible for 0.5% of deaths in the country. It also leads to a 0.4% increase in the loss of healthy life years. Röllin et al. (2004) investigates the relationship between rural electrification and indoor pollution in South African villages. The results show that non-electrified houses had an average daily particulate concentration level of $162\mu\text{g}/\text{m}^3$ compared to an average pollution of $77\mu\text{g}/\text{m}^3$ in electrified houses.

Aware of the level of pollution in the country, South African authorities have put in place policies and interventions to reduce the health costs of air pollution. The results of Leiman et al. (2007) show interventions at the household level are more effective than interventions at the industrial level. The authors conclude that addressing the health costs of air pollution in South Africa should start with addressing the household rather than industrial sources of pollution. Thambiran and Diab (2011) proposes a double benefit policy in terms of reducing pollutants and greenhouse gases. These policies are the reduction of the number of kilometers driven by private motor vehicles and the improvement of the efficiency of road freight transport. Katoto et al. (2019) and Coker and Kizito (2018) through a critical analysis of studies already done on the subject, conclude that the negative health effects of air pollution are under-researched in Sub-Saharan Africa compared to other continents because of a lack of air quality monitoring and epidemiological studies on air pollution, and because of the population's considerable vulnerability.

In view of this abundant literature on Africa, there is less literature on one dimension of human capital, namely education and educational performance. This study aims to fill this gap by examining the relationship between air pollution and educational performance in South Africa. While health is important for good human capital, education also matters.

The following section focuses on the results of previous studies regarding the impact of air pollution on school absenteeism.

2.4 Air pollution and school absenteeism

There is a vast literature on the relationship between suspended particulate pollution and school absenteeism and most of this literature focuses on developed countries. Considering the USA, Ransom and Pope III (1992) investigated the effect of PM10 on school attendance in an elementary school. The results show a positive effect of PM10 on weekly and daily school absenteeism. In a similar vein, Makino (2000) studied the impact of suspended particulate matter on the daily absence of schoolchildren in Japan between 1993 and 1997. The authors find that suspended particulate matter positively

affects school absenteeism. Also, [Park et al. \(2002\)](#) finds a positive effect of PM10 on illness-related school absenteeism in Seoul between March 1996 and December 1999. [Ransom and Pope \(2013\)](#); [MacNaughton et al. \(2017\)](#) also find a positive effect of PM10 and PM2.5 on school attendance in the United States of America. The various authors explain the positive impact of air pollution on school absenteeism through the illnesses that exposure to air pollution causes. As a result, some sick pupils are unable to attend school.

Contrary to our intuition, in the literature, the effect of air pollution on school absenteeism is not always positive. For example, [Chen et al. \(2000\)](#) on a sample of 57 elementary schools in the United States of America between 1996 and 1998, found that PM10 negatively affected school absenteeism. In contrast, another group of authors found that particulate air pollution had no effect on school absenteeism. Using a sample of 2081 students, [Gilliland et al. \(2001\)](#) found a non-significant effect of PM10 exposure on school absenteeism in California in 1996. Similarly, [Rondeau et al. \(2005\)](#) investigated the impact of PM10 on the daily absence of 1932 pupils aged between 9 and 10 in California. The results show that PM10 exposure has no effect on children's daily school absenteeism.

The conclusion of the economic literature on the subject is that the impact of suspended particulates on school absenteeism is mixed. The effect can be positive, negative or nil. This study will contribute to the debate on the subject. The [Table A1](#) in the appendix provides a brief summary of these studies.

3 Data and stylized facts

This study combines household survey data with satellite data of fine particles. In this section are described the databases, the variables used and the stylized facts obtained from the data.

3.1 Data

Panel data over 2 years, 2015 and 2016 are used. The data come from the general household survey ([Statistics South Africa, 2016, 2015](#)), a typical household survey conducted approximately every year throughout South Africa. This survey provides information about socio-economic characteristics, health, education, general functioning and social security of South African households. Data is available at provincial level, so there is no information on the location of households at a finer level.

These survey data offer several advantages. They better describe the relative characteristics of the South African population. First, they enable the collection of data that are close to the exact characteristics of the general population. Second, thanks to the high representativeness of the population, survey data enable us to obtain statistically significant results compared with other methods of data collection. These databases also allow us to have several variables at our disposal for analysis.

To these survey data is added the satellite database of [van Donkelaar et al. \(2018\)](#) which contains information on the average level of PM2.5 concentration in the South African provinces. The two

databases were merged using information on households' province of residence and provide information on local pollution in South Africa. The joint use of these databases is advantageous on several levels. First, it allows a better understanding of the different aspects and impacts of air pollution on the local population. Second, it provides information on whether or not there is heterogeneity of impacts according to provinces and the socio-economic status of households.

In addition, several reasons explain the choice of study period. The first is the availability of data to exploit the panel structure and South African authorities started to survey the same households every year from 2015. Because of temporal dimension of the interest variable (1998-2016) we are obliged to limit the study to a period of 2 years.

The explained variable is a dichotomous variable that takes 1 if the child was absent during the previous school week and 0 otherwise. This variable is the response to the question "Has... been absent from school during the past school calendar week (Monday to Friday) ?", with four (04) possible responses which are "Yes", "No", "Do not know" and "Not applicable - school closed e.g. school holiday". Thus, explained variable is reconstructed from raw data obtained from these answers. Overall, there is a higher level of absenteeism in 2015 compared to 2016. In fact, on graph 1 1202 people were absent in 2015 compared to 1078 people in 2016. There is also an uneven distribution of absences within provinces by year. In 2015, KwaZulu Natal stands out from the other provinces with a high number of people absent.

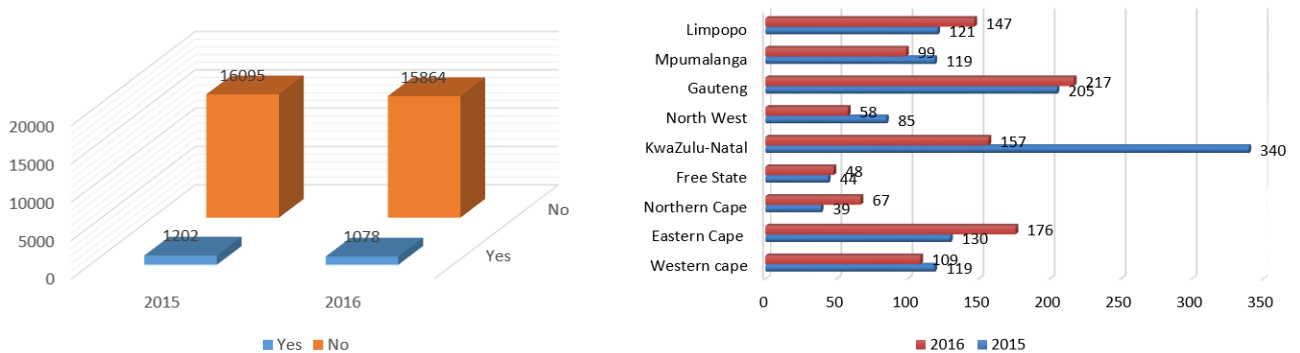


Figure 1: Absenteeism by year (on the left) and province (on the right)

Pollution variable used is a continuous quantitative variable. It measures the annual average concentration of fine particulate matter in $\mu g/m^3$ per province and is taken from the [van Donkelaar et al. \(2018\)](#) database. This is satellite data and extracted using QGIS software. Over the study period and throughout the country, the average pollution is $117.47 \mu g/m^3$ with a dispersion from the average of $67.45 \mu g/m^3$. The minimal pollution is $16.06 \mu g/m^3$ and is registered in the Western Cape province in 2016. On the other hand, the maximum pollution is $221.42 \mu g/m^3$. It is registered in Gauteng province in 2015. These average annual concentrations of PM2.5 far exceed the WHO recommendation of $5 \mu g/m^3$.

Age, gender of the individual, gender of the household head and his or her level of education, household wealth index, household size, presence of biological parents in the household, lack of teacher, travel

time between home and school, rainfall and annual relative humidity are used as control variables.

Individual's age is a very important control variable. Indeed, as the child's age increases, the more likely he or she is to be able to engage in other income-generating activities and the more likely it is to be absent from school. Also, the older the child gets, the less the parents become involved in the child's school affairs. The child could therefore take advantage of this lack of parental supervision to truant.

In addition, the presence of gender in the study matters (Currie et al., 2009). In some African countries, girls are less educated than boys. This discrimination could be a determining factor in school absenteeism and early pregnancy for girls (Spaull, 2015).

The characteristics of the household head help to understand child's behaviours. Education level of the household head could influence his or her interest in school. Household head education level is therefore expected to have a negative effect on the child's school absenteeism. Household head gender is also important because it gives a signal on the financial capacities of the household.

Also, we add information on household's characteristics such as household wealth, household size and biological parents presence in the household. In effect, household size could help us understand the living conditions of child. And then, the presence of the biological parents would indicate whether there is someone caring for the child (Case and Paxson, 2001) or concerned about the child's academic performance.

The level of wealth index permits to approximate socio-economic status of the household. Since information on household total income is not usable. Indeed information on household income is declared data, and may therefore be biased. This is why we use an index based on the characteristics of household assets. The index is constructed from variables that inform on the economic status of the household (Rustein and Johnson, 2004). Information on the type of dwelling, access to electricity, presence of a household helper and the ownership of a vehicle, TV, DVD, air conditioner, computer, dish machine, wash machine and refrigerator are used. Principal component analysis method is used to synthesize all the information and create an index. According to the Kaiser Guttman rule (KG rule), two axes have an eigenvalue greater than 1 (Figure C1 and C2 in appendix C) but we keep the first axis which contains 35% of the total inertia. This means that this axis accounts for the greatest percentage of the total variation in the data used. The wealth index allows us to take into account the living conditions of the household, which is very important in understanding the reasons for the absence of individuals from school (Behrman, 1996; Currie et al., 2009). Wealthy households could require the child to stay at home in case of pollution peaks in order to avoid exposure to pollution. Also, a wealthy household could choose to elect its residence in a non-polluted area. As a result, poor households are the most exposed.

In addition, commuting time to school and lack of teachers are factors that could influence student absenteeism. The further away the home is from school, the less incentive there is to attend. This is also the case with the lack of teachers.

Furthermore, meteorological conditions such as *rainfall and relative humidity* are important in this study. In fact, the weather can have a direct impact on school absenteeism. Bad weather conditions, such as storms, can prevent pupils from attending school. Also weather could impact the likelihood of individuals attending school. Weather impacts through the diseases that bad weather conditions could cause (Currie et al., 2009; Lepeule et al., 2018). And when pupils get sick, some can not attend school either because they stay at home or are hospitalized.

3.2 Stylized Facts

Pollution is an unrecognized silent killer in Africa (Abera et al., 2020). According to the World Bank, poor air quality is responsible for 20,000 premature deaths each year in the country with an economic cost of R300-million (journal, 2016). South Africa is a country with nine provinces whose most polluted are Gauteng, Mpumalanga, North West and Limpopo, Free State and Kwazulu-Natal (Figures 2 and 3) on study period. The high level of air pollution in Gauteng, Mpumalanga and Limpopo can be explained by the presence of coal-fired power stations.

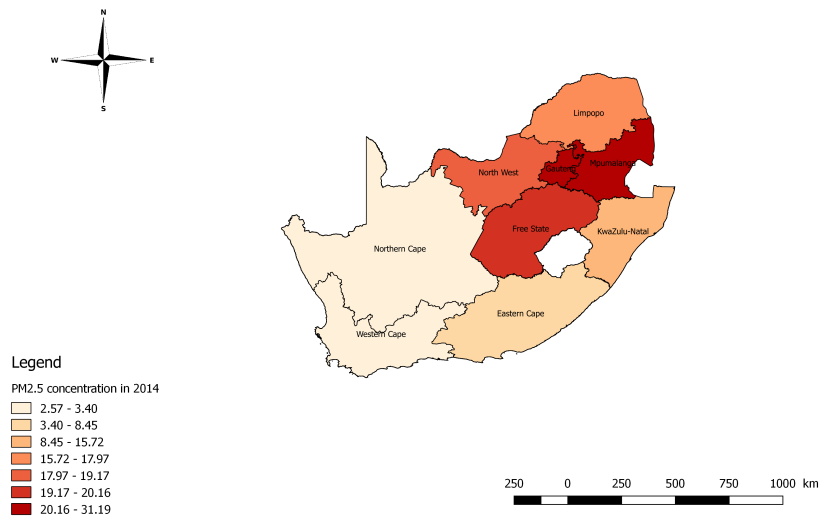


Figure 2: Air pollution in 2014
Source: Author's own elaboration

Also, the high level of particulate pollution would be due to the high level of economic activity in these regions, and more specifically manufacturing activity. Indeed, four of the most polluted provinces are part of the country's top six economic hubs. These are the provinces of Gauteng, Kwazulu-Natal, Mpumalanga and Limpopo. In 2018, these provinces contributed respectively 34.94%, 16.04%, 7.24% and 7.18% to the national GDP (INSIGHTS, 2020). These areas have a very important economic weight for the country. Apart from the fact that the country is the manufacturing hub of Africa, it also has the most industrialized and diversified economy on the continent. Moreover, talking about industries automatically reminds us of one of its side effects, which is pollution. The most economically active provinces are those that emit more pollutants into the atmosphere via road traffic and/or

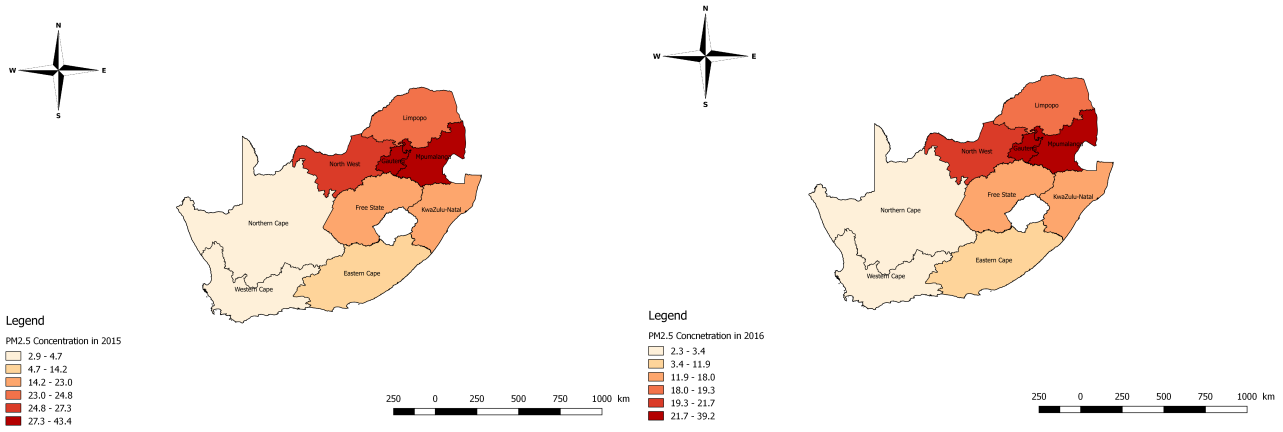


Figure 3: Air pollution in 2015 (left) and air pollution in 2016 (right)
 Source: Author's own elaboration

the industrial sector, which are among the main sources of pollution. The manufacturing sector is the largest source of air pollution in South Africa. In addition, there is the growing urbanisation of certain regions and the neighbourhood effect. The North West Province is one of the most polluted provinces not only because of its proximity to Gauteng but also because of the mining activities in the province. Mining operations are accelerating the rise of pollution in some parts of the country.

According to Greenpeace, NASA satellites data ranks Kriel area in Mpumalanga as second worst sulphur dioxide (SO₂) emission hotspot in the world (Greenpeace, 2019). This high level of sulfur dioxide is due to its coal industry. Indeed, by constituting 92% of energy production and 74.3% of primary energy consumption, coal dominated the energy sector in 2017 and the country was the 3rd largest coal producer in the world in 2018. Sulfur dioxide is very harmful for human health and contributes to the formation of fine particulate matter (PM 2.5). It can lead to respiratory infections, stroke, diabete and lung cancer.

In graph 3 (graph on the right), we can see a relative clean up of the "Free State" province compared to 2015 (graph 3 on the left). This province is very rich in minerals and recognized for its numerous mining activities, especially gold, which represents 20% of world production. Free State also has an industrial sector specialized in high technology. The relative decrease in particulate concentration could probably be explained by a decrease in mining activities or by the use of less polluting technologies.

On the other hand, the "North West" province is less polluted in 2014 (graph 2), and has a higher level of pollution during study period. This could be explained by the increase in economic activities which leads to an intensification of road traffic or an increase in industrial or agricultural emissions. Indeed, mining activity is highly developed in this province. There are gold, uranium, diamond and platinum mines. This activity accounts for more than half of the province's GDP and employs one-quarter of the province's labour force. In the same vain, according to NORTH WEST DEVELOPMENT CORPORATION (2023), economy of North West is dominated by Mining industry. The area contributes 23.9% of national mining and 6.9% to national agriculture, but only 2.3% to manufactur-

ing and 3.8% to construction. Air pollution increase would be explained by the rise in mining activities.

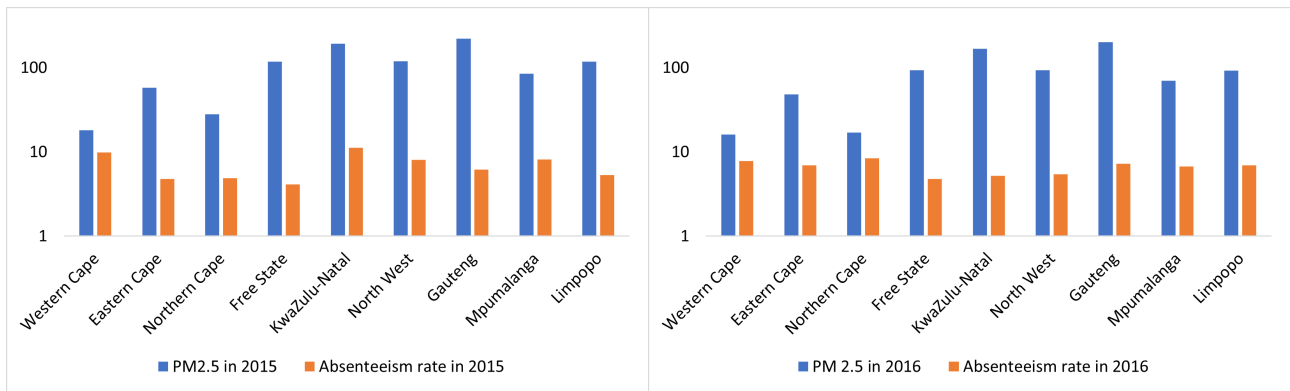


Figure 4: Pollution and absenteeism rate by province and year

3.3 Air pollution and absenteeism

Furthermore, the relationship between absenteeism and pollution is not clear. Indeed, through graph 4, we can see absenteeism rate can be as high in polluted provinces as in less polluted ones. Therefore, with a simple analysis of the data, we can not find the existence of a clear relationship between pollution and absenteeism. The link between the two phenomena could depend on several aspects that are intrinsic to each province. It could depend, for example, on the level of development and the type of economic activity most practised in the province. In the other words, heterogeneity of each area plays a relevant role in explaining relationship between air pollution and child school absenteeism. To take into account this heterogeneity, province and time fixed effects are used and take into account province- and time-specific effects.

Several previous studies show that pollution affects the economy through human capital (Ostro, 1983; Hausman, 1984; Graff Zivin, 2013). Air pollutants affect the health status of individuals and this leads to a decrease in performance in their activities. There are several reasons why pupils were absent the previous school week. Graph 5 shows a small breakdown by reasons for absence. Several reasons explain why pupil did not go to school during previous school week. The most common reason is "illness". Indeed, 555 people were absent because of illness, which is not negligible. In addition, the number of days of absence seems high for some pupils. Indeed, still on Figure 5 we can see 42%, 21%, 10%, 5% and 27% have respectively a duration of absence of 1 day, 2 days, 3 days, 4 days and 5 days.

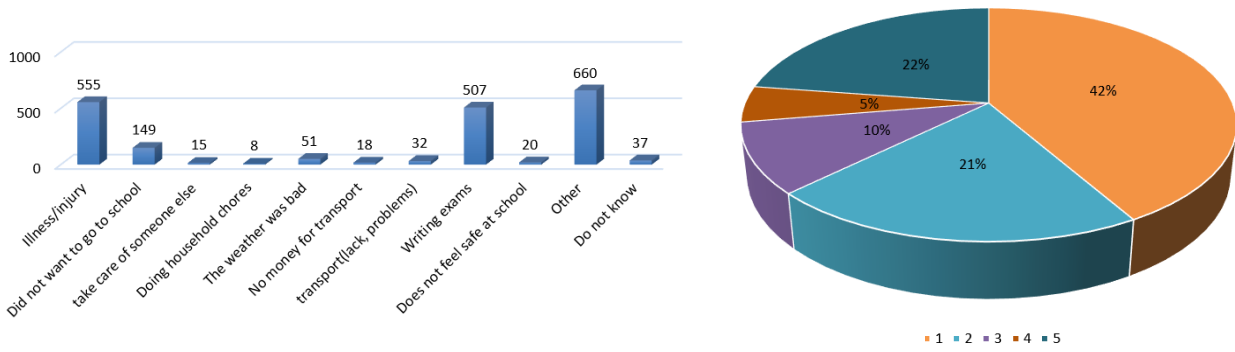


Figure 5: Absenteeism reasons (on the left) and duration (on the right)

In Figure 6, we see that the average number of days lost during the previous school week differs by province. Information on the number of days absent comes from the question "For how many days was ... absent during the past school calendar week (Monday to Friday)". The number of days lost is high in the provinces of Eastern Cape (in 2016), KwaZulu Natal (in 2015 and 2016), and Gauteng (in 2016). In these areas, students lose an average of 3 school days per year. Furthermore, data analysis shows these provinces also have high PM 2.5 concentration. These lost days negatively influence the school performance of the students and thus on South African human capital.

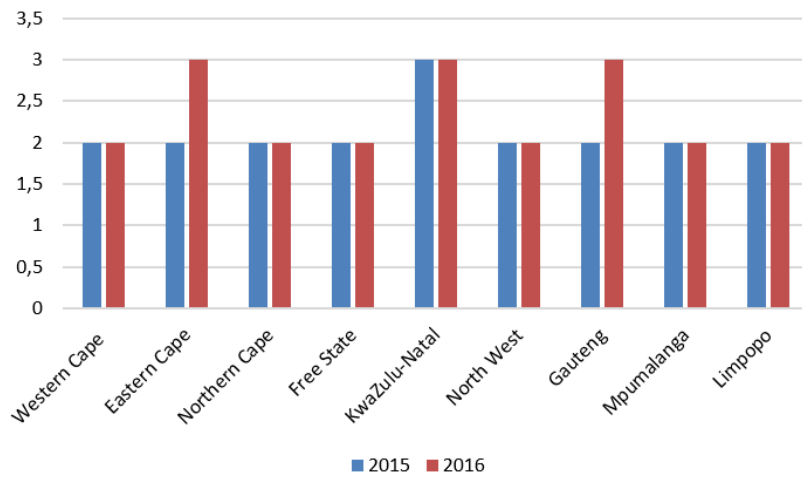


Figure 6: Days lost by year and province

After getting an idea of the different reasons for school absenteeism and duration, in Figure 7 we relate the level of air pollution (PM2.5 in $\mu g/m^3$) to absenteeism rate due to illness over the period by province. The rate of absenteeism (in %) due to illness is calculated by taking the ratio between the number of absentees due to illness and the total number of absentees. In sum, in the majority of provinces with high concentration of fine particulate matter we can notice that there is a significant rate of absenteeism due to illness. We could imagine the existence of a co-evolution between the number of sick people and pollution level.

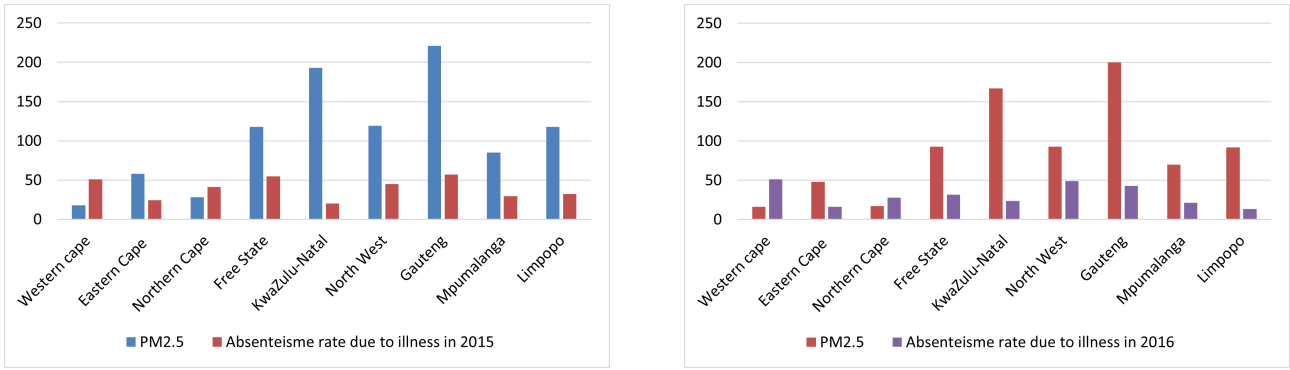


Figure 7: PM2.5, absenteeism and illness

3.4 Air pollution, absenteeism and health problems

We decide to understand this relationship by analyzing the relationship between pollution and absenteeism caused by respiratory diseases. In Figure 8, we can observe a fairly high rate of absenteeism due to flu or acute respiratory tract infection (ARTI) in some provinces in 2015. This high rate is accompanied in most cases by a fairly high level of pollution. In 2016, there was a decrease in the rate of absenteeism and also a decrease in the rate of absenteeism due to flu or acute respiratory tract infection (ARTI), but it still remains high in some provinces. In view of these figures, we could imagine with some reservations pollution affects school absenteeism through the respiratory diseases it causes. Respiratory diseases could be the transmission channel between pollution and school absenteeism in South Africa. Econometric analysis alone could tell us more about the link between air pollution and school absenteeism in South Africa.

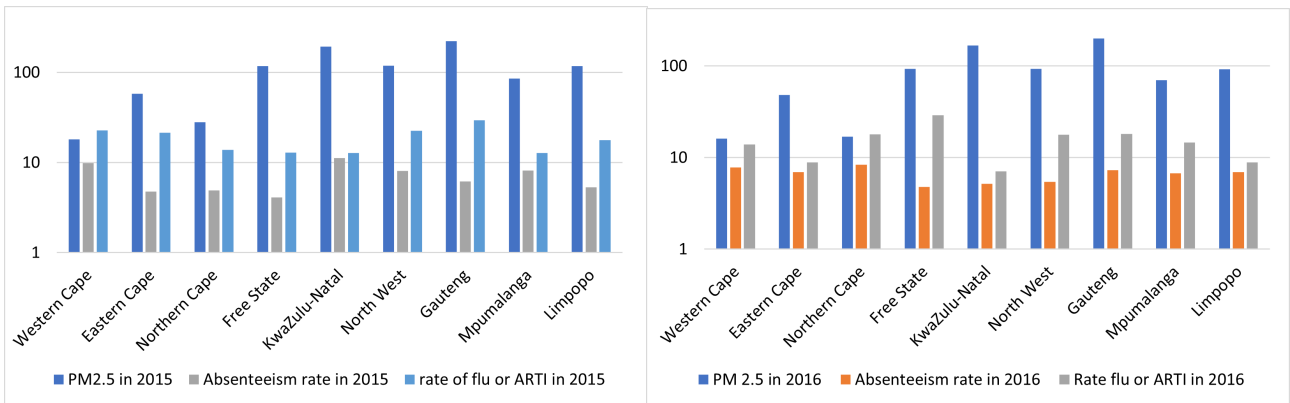


Figure 8: Pollution, absenteeism and acute respiratory tract infection

Through the stylized facts highlighted we formulate the following hypotheses which will be tested later:

- 1) Fine particulate matter pollution affects school absenteeism in South Africa.
- 2) PM2.5 affects absenteeism through the state of health of individuals.
- 3) The effect of fine particles on pupil absence is not linear. It depends on provincial and individual characteristics.

4 Econometric analysis

Exposure to air pollution depends on a number of factors which, if not taken into account, could lead to biased economic results. The aim of this section is to address the endogeneity problem and describe the instrumental variable and econometric modeling used.

4.1 Endogeneity issue and instrument choice

This section describes the endogeneity issue and the instruments used.

4.1.1 Endogeneity issue

There are three main sources of endogeneity which are simultaneity bias, omitted variables and measurement error. Air pollution variable is endogenous due to the bias of omitted variables. Our study suffers from the omission of variables that affect the residential choice of individuals. According to [Tiebout \(1956\)](#), individuals choose their residential area on the basis of several factors which may be, among others, the air quality, schools or the quantity of public goods available in the area. Agents choose a place of residence by arbitrating between the different factors. The choice of residence results in a non-random distribution of exposure to PM_{2.5}. Apart from residential choice, other key variables omitted from the study due to data unavailability are road traffic and industrial and mining production by province. Indeed, mining and industrial activity is the main source of particulate matter emissions in South Africa. A second important source is road traffic. Indeed, the use of diesel vehicles favours the emission of particles. For this reason particle sensors detect a high quantity of particles in road tunnels. Certainly, when particles are emitted in tunnels, it remain condensed and take time to dissipate due to lack of ventilation. The use of the instrumental variable method makes it possible to identify and estimate the causal relationship between air pollution and school absenteeism.

4.1.2 Instruments choice

The frequency of particulate matter variation in the atmosphere depends on its concentration and weather conditions such as wind speed and wind direction ¹. Indeed, according to [Demirci and Cuhadaroglu \(2000\)](#) wind is the main transporter of air pollutants. They find a relationship between the air pollution and wind speed. However, the relationship depends on the direction of the wind speed. Wind can carry polluting particles thousands of kilometres away one city to another or from one country to another ([Kurita et al., 1985](#); [Cox et al., 1975](#)). It really depends on its speed and direction. Moreover, [Lu and Fang \(2002\)](#); [Simpson \(1990\)](#) successively estimate PM₁₀ and PM_{2.5} concentrations from wind speed. [Lu and Fang \(2002\)](#) find that difference between the actual pollutant concentrations and those estimated is non-significant at 5%. Wind is a good predictor of pollutant levels. Also, [Chaloulakou et al. \(2003\)](#) find an inverse relationship between PM₁₀ and PM_{2.5} concentrations and wind speed. Furthermore, wind direction as well as wind speed influences PM_{2.5} concentration in a given geographical area ([Anderson, 2020](#); [Deryugina et al., 2019](#)).

Therefore we use wind speed annual average (measured in m/s) and wind direction annual average

¹U and V components of wind are used to calculate wind direction by applying the formula: $(180/\pi)atan2(v, u)$.

(measured in degrees) at provincial level as instruments. The raw data come from the the European Centre for Medium-Range Weather Forecasts (ECMWF) database and have a resolution of 0.5 (about 50km). Figure 9 shows a negative correlation between pollution level and wind speed. A low wind speed results in a high pollution level in the province. In Figure 10 we can see wind direction is also correlated to the pollution level of the considered area. These two graphs are consistent with the results of [Lu and Fang \(2002\)](#), [Simpson \(1990\)](#) and [Chaloulakou et al. \(2003\)](#).

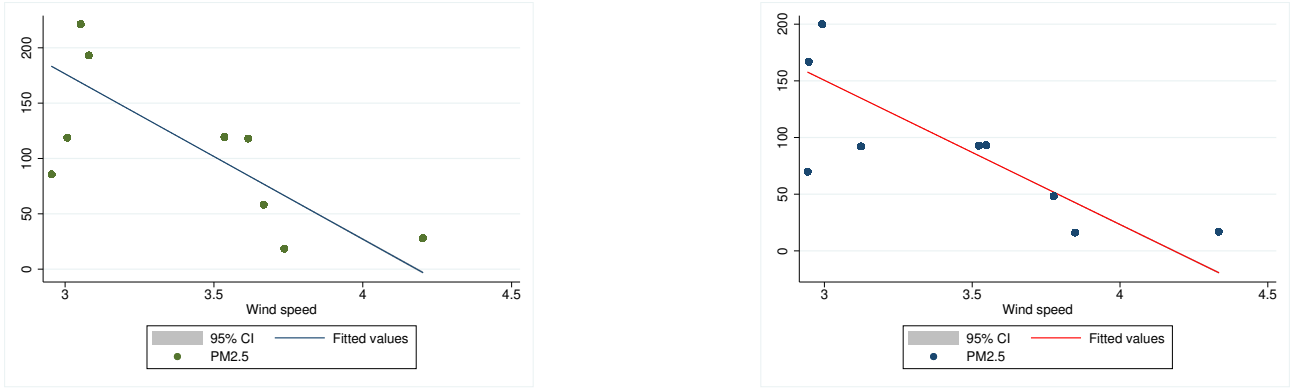


Figure 9: Relation between PM2.5 and wind speed in 2015 (left) and 2016 (right)

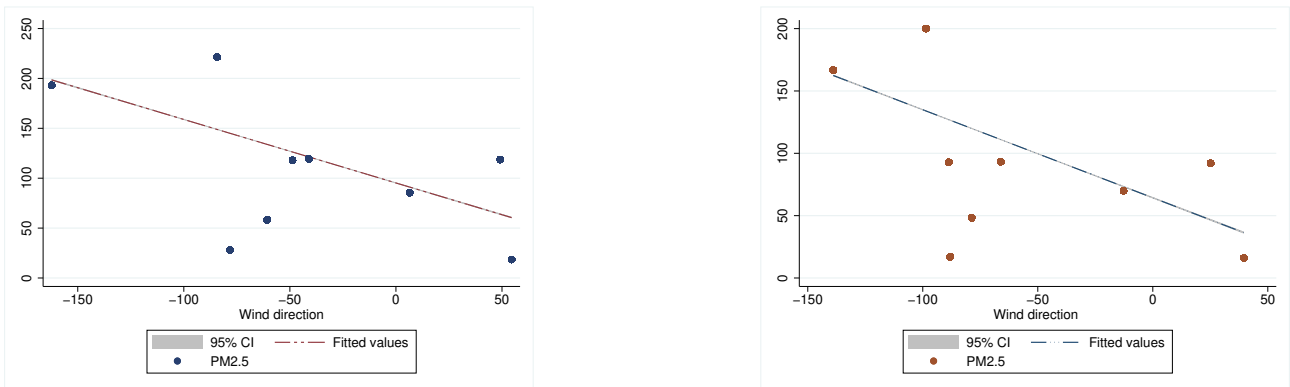


Figure 10: Relation between PM2.5 and wind direction in 2015(left) and 2016 (right)

Also, as additional instrument we add temperature inversion episodes which is a phenomenon that implies an increase in temperature with altitude while the theory indicates the opposite. [Jans et al. \(2018\)](#) and [Arceo et al. \(2016\)](#) show temperature inversions affect pollutant concentration because it leads to its concentration in the atmosphere. There is a relationship between the number of temperature inversion episodes and the level of pollutants and Figure 11 shows this correlation between the two phenomena. We can see on average the areas with a high level of pollution (PM2.5 in $\mu g/m^3$) also observe a high number of temperature inversion episodes.

To get data on temperature inversion events we use satellite data and pixels of about 25km square. For each pixel, we make the difference between the daily temperature at 1000 hpa and the temperature at 925 hpa ² ([Jans et al., 2018](#)). We then calculate for each province the number of pixels where the

²Temperature at 1000 hpa corresponds to the surface conditions and temperature at 925 hpa corresponds to the conditions at approximately 600 m above sea level

result is negative, i.e. where there has been a temperature inversion. We then obtain the number of temperature inversion episodes per province for each year. This is more accurate than calculating the temperature difference over the entire province, as a small area of inversion is likely to be compensated for by areas without inversion.



Figure 11: Relation between PM2.5 and temperature inversions

In addition, good quality instruments comply with the conditions of rank and exclusion. Wind speed, wind direction and temperature inversions affect PM 2.5 concentration in the atmosphere (Chaloulakou et al., 2003; Lu and Fang, 2002; Simpson, 1990; Jans et al., 2018; Arceo et al., 2016). The rank condition is therefore met. Moreover, these instruments do not have a direct impact on the school absence of pupils. Although it should have an effect on school absenteeism, the effect is conveyed by the quantity of pollutants in the atmosphere as a result of wind and temperature inversions. The exclusion condition is verified.

4.2 Control Function Method

Control function (CF hereafter) approach of Wooldridge (2015) is similar to two stage least squares (2SLS) approach. But, it has some interesting advantages over 2SLS. First, the CF estimator allows the comparison of OLS and 2SLS through a Hausman (1978) test, is robust to heteroscedasticity, cluster and serial autocorrelation. The second advantage is CF method allows to treat carefully nonlinear models with endogenous variables.

Indeed, there are very few methods that deal with endogeneity problems in the case of nonlinear panel models. We identify two methods that could be applied to our case: CF method and the generalized structural equation model (Drukker, 2014). But the CF method is the most intuitive because it is close to 2SLS. Details on CF method framework can be found in Part A of the Appendix.

4.2.1 Application

In this study we use a dichotomous model which is a model where the explained variable has two (02) modalities (0 and 1). The modality is 1 if pupil was absent during the last school week at school and

0 otherwise. We can choose between two types of dichotomous models: probit model and logit model. We use a logit model because it has advantages in the interpretation of marginal effects.

CF method of [Wooldridge \(2015\)](#) is applied in the following two steps:

- First stage equation: Application of fixed effect model because there are heterogeneous specific effects that have to be taken into account and also because result of the Hausman test is favorable for the fixed effects model.

$$PM_{pt} = \gamma Z_{1pt} + \theta Z_{2pt} + \lambda X_{ipt} + V_{pt} \quad (2)$$

Where Z_{1pt} and Z_{2pt} are the instruments. PM_{pt} , X_{ipt} and V_{pt} represent respectively pollution, control variables and residuals.

- Second stage equation: Application of logit model with random effects

$$P(Y_{ipt} = 1) = \phi(\alpha + \beta PM_{pt} + \lambda X_{ipt} + \mu \hat{V}_{pt}) \quad (3)$$

Where X_{ipt} and \hat{V}_{pt} are respectively control variables and estimated residuals.

4.3 Results

The main results are in table 1 and in all tables we present the marginal effects. Pollution has a positive effect on the probability of children's absence from school. An increase in PM2.5 concentration by 10 $\mu g/m^3$ results in a 1.33 to 2.46 percent risk of absenteeism. This result is consistent with previous studies ([Ransom and Pope III, 1992](#); [Makino, 2000](#); [Park et al., 2002](#); [Ransom and Pope, 2013](#); [MacNaughton et al., 2017](#)) and could be explained by the health problems induced by air pollution. According to [Umwelthilfe \(2019\)](#), children are the most exposed to pollution due to their size because pollutants are 37 times more concentrated at one metre above the ground than at two metres. This implies children inhale more pollutants than adults. Their exposure to pollution has major repercussions such as respiratory and infectious diseases, asthma and delayed development of the respiratory system and lungs that can lead to a long-term risk of developing respiratory diseases. Exposure of pupils to pollution is increased by the geographical location of schools. In Africa majority of primary schools are located close to major roads. Road traffic is one of the main sources of PM2.5 emissions in addition to the agricultural and industrial sector. As a result, children could breathe poor air quality in schools or on the road to school.

Moreover, travel time between home and school has a positive effect on the probability of children being absent from school. The greater the distance, the greater the risk of absenteeism. A school that is too far from home could demotivate the child's attendance due to the constraints it imposes. Indeed, child may not always want to spend a lot of time getting to school or the parents may not be available to accompany their child to school every time due to time constraints for example. Moreover, travel time also contributes to the exposure of children to pollution. A long commuting time to school results in a high exposure to air pollution due to the time it takes in road traffic to get to school.

In addition, results show the lack of teachers has a positive and significant effect on child's absence from school, which is logical. Teachers who are supposed to transmit knowledge are the reason for the presence of children at school. If they are absent, there is no reason why the children should not be.

Lack of teachers has a positive marginal effect of about 44% on the probability of children’s absence. Therefore this issue affects pupil’s school attendance in South Africa. This result is similar to that found by [Case and Deaton \(1999\)](#); [Visser et al. \(2015\)](#).

Furthermore, results show child’s age has effect on school attendance. A one-year increase in age reduces the risk of absenteeism by 0.18%. Indeed, age level could actually determine the child’s schooling level. The closer the child is to an exam class or is even in an exam class, some parents give more importance to the child’s schooling and thus ensure his or her school attendance.

Table 1: Baseline Results

	Absence	Absence	Absence	Absence	Absence
PM2.5	0.00246*** (3.64)	0.00242*** (3.46)	0.00151** (2.08)	0.00133* (1.87)	0.00195* (1.73)
Child’s gender(Female=1)		-0.00190 (-0.50)	-0.00420 (-0.97)	-0.00413 (-0.94)	-0.00405 (-0.92)
Child’s age		-0.000867 (-1.41)	-0.00133* (-1.88)	-0.00177** (-2.38)	-0.00177** (-2.39)
Head gender(Female=1)		0.00389 (0.77)	0.00360 (0.50)	0.00272 (0.37)	0.00281 (0.38)
Head education		-0.00175*** (-3.12)	-0.00129* (-1.73)	-0.00121 (-1.57)	-0.00124 (-1.60)
Wealth index			0.00163 (0.46)	0.00134 (0.37)	0.00132 (0.36)
Household size			0.00102 (0.84)	0.00130 (1.04)	0.00130 (1.04)
Biological father presence			0.00293 (0.39)	0.00299 (0.39)	0.00272 (0.35)
Biological mother presence			0.00853 (1.42)	0.00912 (1.49)	0.00936 (1.53)
Lack of teacher				0.0378* (1.87)	0.0374* (1.85)
Time to travel to school				0.0141** (2.45)	0.0140** (2.43)
Relative humidity					0.00450** (2.17)
Rainfall					0.000138 (1.29)
Province FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
<i>Observations</i>	19225	17535	13428	13008	13008
Cluster variable	hh	hh	hh	hh	hh
LR test P value		0.0000	0.0000	0.0000	0.0247
Converged	yes	yes	yes	yes	yes
R ²	0.006	0.10	0.32	0.33	0.34

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The table shows marginal effects

The explained variable "Absence" is 1 if the child was absent during the previous school week and 0 otherwise

In addition, the household head education level negatively influences the risk of the child’s absence from school. The fact that household head has a high level of education enables child to give more importance to school for success in life or to have a source of income in the future. Thus, the higher the level of education of the household head, the more importance he or she attaches to the educational success of the child.

Moreover, meteorological conditions more precisely relative humidity affects absenteeism. It impacts through the diseases that bad weather conditions could cause ([Currie et al., 2009](#); [Lepeule et al., 2018](#)).

5 Health transmission channels

Earlier literature links school absenteeism due to pollution to the health problems it causes (Romieu et al., 1992; Gilliland et al., 2001). Pollution affects the education of pupils through the cardio-pulmonary problems it creates. As a result, children’s attendance would depend of severity of the disease. Those who are less ill could be present but less receptive to the knowledge which are taught in class. On the other hand, those who are seriously ill would not be able to attend and would risk missing several days of classes.

To test the "health" channel revealed by previous studies (Kampa and Castanas, 2008; Pope III and Dockery, 2006), some "health" variables whose results are used and presented in Table 2. The "health" variables tested obtained from the questions *"During the past three months, did... suffer: 1=Flu or acute respiratory tract infection, 2=diarrhoea, 3= Severe cough with blood, 4=abuse of alcohol or drugs, 5=depression, 6= sexual transmitted diseases, 7=pneumonia, 8=bronchitis, 9=epilepsy"* and *"Did... consult a health worker such as a nurse, doctor or traditional healer as a result of this illness?"*

To test the transmission channels, we use stepwise regression by reference to the study of Zhou et al. (2022). This method allows us to determine the mediation effects through the following equations:

$$Absence = \alpha PM2.5 + \epsilon_1 \quad (1)$$

$$M = \beta PM2.5 + \epsilon_2 \quad (2)$$

$$Absence = \theta PM2.5 + \rho M + \epsilon \quad (3)$$

Where M is the mediation variable. The last step of the applied method differs from that of Zhou et al. (2022) because we have taken into account the control variables in equation (3). This allows us to capture the mediation effect of the transmission channels in the baseline model. A logit model is used for testing because the dependent variable and transmission channels are binary.

Results show "medical consultation" and "Flu or acute respiratory tract infection (ARTI)" are transmission channels. In table 2, we can see air pollution affects the transmission channels (columns 2 and 6) and these transmission channels also affect absenteeism (columns 3 and 7). It can be said that pollution acts on school absenteeism through "medical consultation" and "flu or ARTI". And it is important to note that this is a partial intermediary effect as the effect of pollution is still significant when transmission channels are added to the baseline model (He et al., 2022).

Results show pollution partially affects child school attendance through the diseases and more precisely respiratory diseases. Indeed, exposure to pollution causes respiratory diseases that can prevent pupils from attending school. This could have a negative effect on their school performance.

In the next section, we test the robustness of the baseline results by changing the model and the estimator.

Table 2: Transmission channel

	Medical consultation	Absence	Cough	Absence	Flu or ARTI	Absence
PM2.5	0.000108*** (2.65)	0.00197* (1.75)	0.00000859* (1.77)	0.00195* (1.73)	0.0000980** (2.07)	0.00192* (1.72)
Medical consultation		0.0593*** (7.31)				
Cough				0.0289 (0.46)		
Flu or ARTI						0.0618*** (8.90)
Child's gender(Female)		-0.00372 (-0.85)		-0.00403 (-0.91)		-0.00310 (-0.71)
Child's age		-0.00131* (-1.77)		-0.00178** (-2.40)		-0.00121 (-1.64)
Head gender(Female)		0.00240 (0.33)		0.00284 (0.39)		0.00231 (0.32)
Head education		-0.00134* (-1.74)		-0.00124 (-1.60)		-0.00139* (-1.80)
Wealth index		0.00277 (0.76)		0.00135 (0.37)		0.00290 (0.80)
Household size		0.00162 (1.32)		0.00130 (1.04)		0.00181 (1.47)
Biological father presence		0.00321 (0.42)		0.00271 (0.35)		0.00377 (0.49)
Biological mother presence		0.00935 (1.52)		0.00936 (1.53)		0.00862 (1.40)
Lack of teacher		0.0312 (1.62)		0.0375* (1.85)		0.0296 (1.59)
Time to travel to school		0.0138** (2.39)		0.0140** (2.44)		0.0133** (2.32)
Relative humidity		0.00439** (2.12)		0.00450** (2.17)		0.00443** (2.14)
Rainfall		0.000136 (1.27)		0.000138 (1.29)		0.000138 (1.29)
Province FE	No	Yes	No	Yes	No	Yes
Time FE	No	Yes	No	Yes	No	Yes
Observations	13008	13008	13008	13008	13008	13008

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6 Robustness checks

The results remain unchanged following the use of Generalized Structural Equation Modeling which is an alternative method to CF to fix endogeneity problem (Drukker, 2014) and probit model with random effects. We can see pollution always has a positive effect on the probability of school absence (Table 3).

Table 3: Robustness with probit modeling and GSEM

	Probit	GSEM
PM 2.5	0.00192* (1.70)	0.0322** (2.01)
Child's gender(Female=1)	-0.00410 (-0.93)	-0.0663 (-0.87)
Child's age	-0.00173** (-2.35)	-0.0315** (-2.44)
Head gender (Female=1)	0.00284 (0.39)	0.0538 (0.43)
Head education	-0.00123 (-1.59)	-0.0207 (-1.55)
Wealth index	0.00124 (0.34)	0.0277 (0.44)
Household size	0.00134 (1.07)	0.0216 (1.02)
Biological father presence	0.00249 (0.33)	0.0453 (0.35)
Biological mother presence	0.00947 (1.55)	0.166 (1.47)
Lack of teacher	0.0375* (1.87)	0.520** (2.21)
Time to travel to school	0.0141** (2.46)	0.247** (2.45)
Relative humidity	0.00446** (2.18)	0.0772** (2.15)
Rainfall	0.000139 (1.30)	0.00222 (1.31)
Province FE	yes	yes
Time FE	yes	yes
Wind speed		-127.0*** (-95.88)
Wind direction		-0.319*** (-48.04)
Temperature inversions		-0.00536*** (-24.91)
<i>Observations</i>	13008	15526
<i>Cluster variable</i>	hh	hh
<i>Converged</i>	yes	yes

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

7 Heterogeneity of effect

In the following lines, I test the non-linearity of the impact as a function of the characteristics of the individual, household head and the household.

7.1 Effect by age and socio-Economic Status

Income level allows household to protect itself against the effects of air pollution. Indeed, we know that according to the concept of "voting with their feet" of Tiebout (1956), individuals can move if the living conditions of their current geographical area are not convenient for them. Thus, when there is a deterioration of air quality, wealthy households prefer to move to an area with less air pollution. The geographical distribution of individuals would be such that wealthy households will tend to move to less polluted areas compared to poor households who do not have enough financial means to move to residential areas.

Table 4: Effect by individual’s age and household wealth

	age < 16	age ≥ 16	25% poorest	25% wealthiest
PM 2.5	0.00195* (1.73)	0.00434** (2.39)	0.00308 (1.24)	0.00463** (2.22)
Child’s gender(Female=1)	-0.00405 (-0.92)	-0.00884 (-0.95)	-0.0262*** (-2.94)	0.0147 (1.64)
Child’s age	-0.00177** (-2.39)	0.00441** (2.51)	-0.000728 (-0.52)	-0.00424** (-2.55)
Head gender (Female=1)	0.00281 (0.38)	0.00990 (0.69)	-0.00356 (-0.28)	0.0164 (1.06)
Head education	-0.00124 (-1.60)	0.00129 (1.02)	-0.00264* (-1.84)	-0.00145 (-0.97)
Wealth index	0.00132 (0.36)	0.00590 (0.92)		
Household size	0.00130 (1.04)	0.0000983 (0.05)	-0.00344 (-1.30)	0.00155 (0.64)
Biological father presence	0.00272 (0.35)	0.000405 (0.03)	0.00314 (0.25)	0.000987 (0.06)
Biological mother presence	0.00936 (1.53)	0.0212** (1.98)	0.00273 (0.18)	-0.00462 (-0.38)
Lack of teacher	0.0374* (1.85)	0.0118 (0.39)	-0.0188 (-0.68)	0.0308 (0.71)
Time to travel to school	0.0140** (2.43)	0.0151 (1.31)	0.0123 (1.11)	0.0166 (1.30)
Relative humidity	0.00450** (2.17)	0.0110*** (2.84)	0.000732 (0.17)	0.00403 (0.73)
Rainfall	0.000138 (1.29)	0.000513*** (2.79)	0.000316 (1.28)	0.000369* (1.79)
Province FE	yes	yes	yes	yes
Time FE	yes	yes	yes	yes
<i>Observations</i>	13008	3569	3173	3093
Cluster variable	hh	hh	hh	hh
Converged	yes	yes	yes	yes

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The results obtained are not in line with the concept of ”voting with their feet” of [Tiebout \(1956\)](#). Indeed, in table 4 (column 3 and 4), we can see pollution has not an impact on the school absenteeism of children from poorest households. According to [Hajat et al. \(2013\)](#); [Neidell \(2004\)](#); [Pratt et al. \(2015\)](#) wealthier households have more opportunity to protect themselves from air pollution because of ”avoidance behavior”. Indeed, they can change their place of residence by choosing a less polluted area and can also protect themselves by buying protective equipment such as masks. Poor households, however, do not have all these opportunities. The result we get could be explained by the failure to take environmental factors into account in the household’s choice of residence. Indeed, according to [Lesia et al. \(2023\)](#) the choice of residence of South African households is influenced by demographic characteristics, dwelling unit features, neighbourhood features, services provided by the government, household self-congruity, functional congruity, green building features and stakeholders’ relations.

Furthermore, individual’s age could lead to heterogeneity in the effect of pollution. Results in Table 4 show that individuals aged 16 years or more are more impacted compared to the younger ones and more specifically, individuals of study sample. This could be explained by the fact that older

individuals are more likely to be outside the house.

7.2 Effect by subgroup

South Africa has a diverse population, which leads to cultural diversity that may affect the lifestyles of individual households. The objective of this section is to see if there is heterogeneity in the effect depending on the cultural subgroup of the household head. Specifically, to see if the effect varies depending on whether the head of the household is black or not. According to [Heissel et al. \(2019\)](#), in Florida the effect of pollution affects the academic performance of black pupils while the effect is null on the performance of other pupils. Our goal is to see if there is a difference in the effect in South Africa.

In the sample, household heads were divided into 4 subgroups (black African, Coloured, Indian or Asian and white). Only 2 groups (black and coloured) are used because of the convergence problems encountered with the other groups. For example, the Indian/Asian and White groups were grouped together but there was still a problem of convergence. In a logistic model, lack of convergence is due to the fact that the likelihood maximization algorithm does not converge. [Table 5](#) shows very interesting results according to the group to which the household head belongs. In households with a black head, the marginal effect of pollution on the risk of absenteeism is positive, but it is negative in households with a coloured head.

This result could be explained by the socio-economic status of households with a coloured head. Indeed, a high level of pollution could approximate the level of economic activity. For these households, more pollution could mean more income. Head of the household could therefore invest better and be involved in the education of the children. This could result in a lower level of absenteeism.

Table 5: Effect by household head subgroup

	Black	Coloured
PM2.5	0.00241* (1.91)	-0.0300** (-2.37)
Child's gender(Female=1)	-0.00234 (-0.51)	-0.0194 (-1.19)
Child's age	-0.00181** (-2.30)	-0.00272 (-0.99)
Head gender(Female=1)	0.00373 (0.47)	-0.00668 (-0.30)
Head education	-0.000979 (-1.23)	-0.00490* (-1.65)
Wealth index	0.000298 (0.07)	0.0138 (1.07)
Household size	0.00210 (1.62)	-0.00717* (-1.66)
Biological father presence	0.00264 (0.31)	-0.0125 (-0.55)
Biological mother presence	0.00638 (1.00)	0.0340 (1.51)
Lack of teacher	0.0360 (1.60)	0.0190 (0.44)
Time to travel to school	0.0155** (2.51)	0.00767 (0.41)
Relative humidity	0.00450* (1.81)	-0.0190 (-1.38)
Rainfall	0.000181* (1.75)	-0.00177** (-2.06)
Province FE	Yes	Yes
Time FE	Yes	Yes
<i>Observations</i>	11303	1248
Cluster variable	HH	HH
converged	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7.3 Effect by household size and composition

Households may differ in size or composition and this could influence individual's behavior in the household. It is relevant to take this into account as it could lead to heterogeneity in the effect of air pollution on school attendance. Therefore, we first investigate the difference in the effect according to the size of the household (Table 6). Studies show that household size influences education and well-being of children (Dang and Rogers, 2016; Cáceres-Delpiano, 2006; Ponczek and Souza, 2012). Parents make a trade-off between quantity and quality of children. Instead of having many children, they focus on the quality of children by investing in their education and well-being. In this vein Dang and Rogers (2016) shows Vietnamese parents will tend to invest less in the schooling of a child with several siblings. One could imagine the size of the South African household would influence parental investment in the quality of children.

Results are shown in Table 6. We can see the fact that household is formed of at least 6 members positively affects air pollution impact on truancy. Therefore, the effect of air pollution on the risk of absenteeism in South Africa depends on the size of the household. We use the distribution of South

African households by size in 2019 (Statista 2021) to study this heterogeneity. Figure 12 shows a distribution of households into four size classes. We can see that 37.3% of the households were formed by 2 or 3 individuals. We are interested only in households with a size greater than or equal to 2 because a household of one individual means there are no children. Whereas a two-member household can be formed by one of the parents and the child.

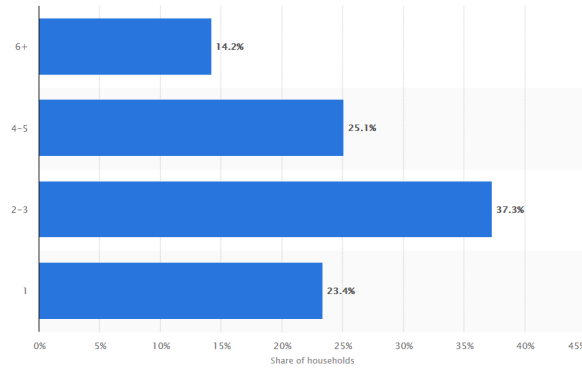


Figure 12: Distribution of households in South Africa in 2019, by household size
Source: Statista 2021

Table 6: Effect by household size

	Size=2 or 3	Size= 4 or 5	Size >= 6
PM2.5	0.00179 (0.73)	0.0000839 (0.05)	0.00395** (2.57)
Child's gender(Female=1)	0.00125 (0.12)	-0.0105 (-1.62)	-0.00406 (-0.66)
Child's age	0.0000695 (0.05)	0.000170 (0.20)	0.0000784 (0.09)
Head gender(Female=1)	0.00939 (0.51)	-0.00207 (-0.17)	0.00461 (0.44)
Head education	-0.00105 (-0.62)	-0.00111 (-1.04)	-0.000824 (-0.73)
Wealth index	0.00176 (0.25)	0.00733 (1.57)	-0.00295 (-0.45)
Biological father presence	-0.00144 (-0.07)	0.00451 (0.36)	0.00169 (0.16)
Biological mother presence	0.0155 (1.17)	0.0110 (1.13)	0.0144* (1.71)
Lack of teacher	0.00766 (0.23)	0.0576* (1.88)	0.00845 (0.32)
Time to travel to school	0.0129 (1.08)	0.00861 (1.10)	0.0192** (2.09)
Relative humidity	0.0101** (2.36)	0.00584** (2.09)	0.00284 (0.76)
Rainfall	0.0000622 (0.26)	-0.0000340 (-0.20)	0.000327** (2.17)
Province FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	2115	5629	7450
Cluster variable	hh	hh	hh
Converged	Yes	Yes	Yes
R ²	0.43	0.36	0.35

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Effect by household structure

	Bio mother in the HH	Bio mother no in the HH	Bio father in the HH	Bio father no in the HH	Bio father no in the HH & bio mother no in the HH	Bio father no in the HH & Bio mother in the HH	Bio father in the HH & Bio mother no in the HH	Bio father in the HH & Bio mother in the HH
PM2.5	0.00130 (0.99)	0.00345* (1.83)	0.00184 (1.03)	0.00197 (1.47)	0.00342* (1.75)	0.00102 (0.61)	0.00273 (0.39)	0.00201 (1.03)
Child's gender(Female)	-0.00625 (-1.22)	0.00183 (0.23)	-0.00967 (-1.34)	-0.000651 (-0.12)	0.00134 (0.16)	-0.00191 (-0.27)	0.0298 (0.68)	-0.0111 (-1.50)
Child's age	-0.00174** (-2.05)	-0.00174 (-1.17)	-0.00178 (-1.49)	-0.00179* (-1.88)	-0.000630 (-0.41)	-0.00242** (-2.03)	-0.0199*** (-2.97)	-0.000914 (-0.76)
Head gender(Female)	0.00135 (0.20)	0.0127 (1.24)	-0.00514 (-0.33)	0.00282 (0.35)	0.00915 (0.85)	0.00260 (0.23)	0.0586 (1.31)	-0.0154 (-0.77)
Head education	-0.00119 (-1.29)	-0.00135 (-1.06)	-0.000236 (-0.16)	-0.00135 (-1.64)	-0.00127 (-0.93)	-0.00190* (-1.82)	-0.00694 (-0.88)	0.000216 (0.13)
Wealth index	0.000863 (0.21)	0.00280 (0.37)	0.00482 (0.92)	-0.00109 (-0.21)	0.00927 (1.26)	-0.00702 (-1.07)	-0.0535* (-1.68)	0.00836 (1.57)
Household size	0.000994 (0.70)	0.00156 (0.81)	0.00238 (1.02)	0.000774 (0.58)	0.00106 (0.52)	0.000395 (0.25)	0.00519 (0.57)	0.000957 (0.36)
Lack of teacher	0.0265 (1.33)	0.0827 (1.56)	0.00930 (0.38)	0.0541* (1.90)	0.0746 (1.37)	0.0411 (0.42)	0.108 (0.87)	0.00372 (0.15)
Time to travel to school	0.0176*** (2.66)	0.00393 (0.38)	0.0152 (1.64)	0.0138** (1.98)	-0.00120 (-0.12)	0.0225** (2.55)	0.108* (1.80)	0.0124 (1.30)
Relative humidity	0.00383 (1.62)	0.00553 (1.17)	0.00653* (1.80)	0.00377 (1.43)	0.00512 (1.03)	0.00215 (0.67)	0.0103 (0.34)	0.00644* (1.76)
Rainfall	0.000111 (0.88)	0.000169 (0.96)	0.0000526 (0.28)	0.000163 (1.27)	0.000177 (1.00)	0.000133 (0.82)	-0.0000945 (-0.11)	0.0000660 (0.33)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9637	3371	4820	8188	3033	5118	199	4519

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Moreover, composition of the household could affect pollution's effect. Indeed, among the determinants of school absenteeism, presence of biological parents in the household plays an important role. [Case and Ardington \(2006\)](#) shows that whether parents are alive or not has a different effect on children's school performance. In order to take this aspect into account, we analyze the effect of air pollution on school attendance according to the presence of biological parents in the household. The results show that the marginal effect of pollution is zero when the biological mother is present in the household, and the effect becomes positive when the biological mother does not live in the household (Table 7 columns 1 and 2). The effect of pollution on the risk of absenteeism is zero, regardless of whether the biological father is present or not (Table 7 columns 3 and 4). On the other hand, the marginal effect is positive when none of the biological parents is present in the household (column 4). Moreover, the marginal effect is zero when the biological mother is present and the biological father is absent, or when both biological parents are present in the household. These results show the importance of the biological mother's presence on the child's school attendance. This is because the biological mother takes care of the child when she is present. She looks after the child's health and well-being.

8 Conclusion

This study focuses on the impact of pollution on human capital. It looks at absenteeism among children aged 5 to 15 years in South Africa in 2015 and 2016. The choice of age group is based on the country's education system, where primary school pupils are between 5 and 15. Results show pollution has a significant positive effect on the risk of child school absence. The marginal effect is between 1.33 and 2.46% following a $10 \mu\text{g}/\text{m}^3$ increase in the average concentration of PM 2.5. In this way, the study contributes to the literature by removing ambiguity about the effect of air pollution on school absenteeism. Other factors such as child's age, education level of the household head, lack of

teacher, travel time between home and school, annual relative humidity also affect school absenteeism. Moreover, the impact varies by individual and household characteristics.

Moreover, data analysis and previous literature (Romieu et al., 1992; Gilliland et al., 2001; Kampa and Castanas, 2008; Pope III and Dockery, 2006) show the effect of pollution on school absence is explained in particular by the health problems it induces. Pollution would therefore affect children's schooling via the "health" channel. Also, the results obtained are in line with those found by some of the previous studies (Ransom and Pope III, 1992; Makino, 2000; Park et al., 2002; Ransom and Pope, 2013; MacNaughton et al., 2017) (Table A1).

Furthermore, in terms of environmental policies, the country is lagging somewhat behind, as are most developing countries. The South African authorities first became interested in particulate matter in 2012. Under the Air Quality Act of 2004, they are adding standards for PM_{2.5} concentration. The country sets as standards an annual PM_{2.5} concentration level of 25 $\mu\text{g}/\text{m}^3$ between 2012 and 2015, then a concentration of 20 $\mu\text{g}/\text{m}^3$ between January 2016 and December 2029 and finally a concentration of 15 $\mu\text{g}/\text{m}^3$ from January 2030. To protect population from the harmful effects of air pollution, South African leaders must ensure measures are respected. Also they must play an active role to fight global warming. Indeed, global warming is the main reason for the concentration of particles in the atmosphere because it leads to wide fluctuations in rainfall and even absence of rainfall in some regions. Rain cleans the air of the particulate matter and pollutants it contains. As a result, the less rain there is, the more particulate matter there are in the atmosphere that negatively affect the health of individuals. Another important recommendation is the relocation or construction of schools away from high road traffic areas.

An important limitation of this study is the time dimension of "health" variables which serve as transmission channels. As for question asked in the questionnaire, the variables concerned identify individuals who have fallen ill in the last three months, whereas the question on absenteeism concerns child who have been absent during the previous school week. There is therefore a problem with the time dimension. It is possible pupils who have been ill in the last three months may have been present during the previous school week. This could lead to an underestimation of the effect of pollution passing through the transmission channel. Another limitation lies in the geographical precision of the households. Indeed, data are not geo-referenced. Therefore we do not have a precise idea of the location of households. The province of residence of households is the only information we have. Information on the city of residence, for example, would allow us to have more precise information on the level of pollution to which households are exposed. Also, it is implicitly assumed that the observed level of absenteeism corresponds to the annual absenteeism level of the province. The lack of air pollution data over the survey period leads us to make this implicit assumption.

Future research could attempt to address the various limitations of this paper, particularly with regard to information on household location. Another research idea would be to try to understand why the effect of air pollution differs according to the population subgroups to which the household head belongs.

References

- Abera, A., Friberg, J., Isaxon, C., Jerrett, M., Malmqvist, E., Sjöström, C., Taj, T., and Vargas, A. M. (2020). Air quality in africa: Public health implications. *Annual Review of Public Health*, 42.
- Adato, M., Devereux, S., and Sabates-Wheeler, R. (2016). Accessing the ‘right’ kinds of material and symbolic capital: the role of cash transfers in reducing adolescent school absence and risky behaviour in South Africa. *The Journal of Development Studies*, 52(8):1132–1146.
- Anderson, M. L. (2020). As the wind blows: The effects of long-term exposure to air pollution on mortality. *Journal of the European Economic Association*, 18(4):1886–1927.
- Arceo, E., Hanna, R., and Oliva, P. (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? evidence from mexico city. *The Economic Journal*, 126(591):257–280.
- Barnes, B., Mathee, A., Thomas, E., and Bruce, N. (2009). Household energy, indoor air pollution and child respiratory health in South Africa. *Journal of Energy in Southern Africa*, 20(1):4–13.
- Basic Education Department, o. (2017a). National policy on hiv, stis and tb for learners, educators, school support staff and officials in all primary and secondary schools in the basic education sector.
- Basic Education Department, o. (2017b). Protocol to deal with incidences of corporal punishments in schools.
- Bauer, S. E., Im, U., Mezuman, K., and Gao, C. Y. (2019). Desert dust, industrialization, and agricultural fires: Health impacts of outdoor air pollution in africa. *Journal of Geophysical Research: Atmospheres*, 124(7):4104–4120.
- Behrman, J. R. (1996). The impact of health and nutrition on education. *The World Bank Research Observer*, 11(1):23–37.
- Branson, N., Hofmeyr, C., and Lam, D. (2014). Progress through school and the determinants of school dropout in South Africa. *Development Southern Africa*, 31(1):106–126.
- Cáceres-Delpiano, J. (2006). The impacts of family size on investment in child quality. *Journal of Human Resources*, 41(4):738–754.
- Case, A. and Ardington, C. (2006). The impact of parental death on school outcomes: Longitudinal evidence from South Africa. *Demography*, 43(3):401–420.
- Case, A. and Deaton, A. (1999). School inputs and educational outcomes in South Africa. *The Quarterly Journal of Economics*, 114(3):1047–1084.
- Case, A. and Paxson, C. (2001). Mothers and others: who invests in children’s health? *Journal of health economics*, 20(3):301–328.
- Chaloulakou, A., Kassomenos, P., Spyrellis, N., Demokritou, P., and Koutrakis, P. (2003). Measurements of pm₁₀ and pm_{2.5} particle concentrations in athens, greece. *Atmospheric Environment*, 37(5):649–660.

- Chen, L., Jennison, B. L., Yang, W., and Omaye, S. T. (2000). Elementary school absenteeism and air pollution. *Inhalation Toxicology*, 12(11):997–1016.
- Coker, E. and Kizito, S. (2018). A narrative review on the human health effects of ambient air pollution in sub-saharan africa: An urgent need for health effects studies. *International journal of environmental research and public health*, 15(3):427.
- Cox, R., Eggleton, A., Derwent, R., Lovelock, J., and Pack, D. (1975). Long-range transport of photochemical ozone in north-western europe. *Nature*, 255(5504):118.
- Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., and Rivkin, S. G. (2009). Does pollution increase school absences? *The Review of Economics and Statistics*, 91(4):682–694.
- Dang, H.-A. H. and Rogers, F. H. (2016). The decision to invest in child quality over quantity: Household size and household investment in education in vietnam. *The World Bank Economic Review*, 30(1):104–142.
- Demirci, E. and Cuhadaroglu, B. (2000). Statistical analysis of wind circulation and air pollution in urban trabzon. *Energy and buildings*, 31(1):49–53.
- Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., and Reif, J. (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, 109(12):4178–4219.
- Drukker, D. M. (2014). Some stata commands for endogeneity in nonlinear panel-data models. In *Presentation to the German Stata Users Group meeting*.
- Education Department, O. (2001). Age requirements for admission to an ordinary public school.
- Education Department, o. (2002). Revised national curriculum statement grades r-9 (schools). *Policy: Natural Sciences*, 443.
- Elf, J. L., Eke, O., Rakgokong, M., Variava, E., Baliram, Y., Motlhaoleng, K., Lebina, L., Shapiro, A. E., Breyse, P. N., Golub, J. E., et al. (2017). Indoor air pollution from secondhand tobacco smoke, solid fuels, and kerosene in homes with active tuberculosis disease in South Africa. *BMC research notes*, 10(1):1–6.
- Gilliland, F. D., Berhane, K., Rappaport, E. B., Thomas, D. C., Avol, E., Gauderman, W. J., London, S. J., Margolis, H. G., McConnell, R., Islam, K. T., et al. (2001). The effects of ambient air pollution on school absenteeism due to respiratory illnesses. *Epidemiology*, pages 43–54.
- Graff Zivin, Joshua et Neidell, M. (2013). Environment, health, et human capital. *Journal of Economic Literature*, 51(3):689–730.
- Greenpeace (2019).
- Gustafsson, M. and al. (2011). The when and how of leaving school: The policy implications of new evidence on secondary schooling in south africa. Technical report.

- Hajat, A., Diez-Roux, A. V., Adar, S. D., Auchincloss, A. H., Lovasi, G. S., O'Neill, M. S., Sheppard, L., and Kaufman, J. D. (2013). Air pollution and individual and neighborhood socioeconomic status: evidence from the multi-ethnic study of atherosclerosis (mesa). *Environmental health perspectives*, 121(11-12):1325–1333.
- Hausman, Jerry A et Ostro, B. D. e. W. D. A. (1984). Air pollution et lost work.
- He, J., Qing, C., Guo, S., Zhou, W., Deng, X., and Xu, D. (2022). Promoting rural households' energy use for cooking: Using internet. *Technological Forecasting and Social Change*, 184:121971.
- Heissel, J., Persico, C., and Simon, D. (2019). Does pollution drive achievement? the effect of traffic pollution on academic performance. Technical report, National Bureau of Economic Research.
- Hilsdon, A.-M. and Randell, S. (2012). Joint ilo–unesco committee of experts on the application of the recommendations concerning teaching personnel (ceart). *11th Session of the CEART (Geneva, 8–12 October 2012)*.
- Hunter, N. and May, J. (2003). *Poverty, shocks and school disruption episodes among adolescents in South Africa*. School of Development Studies, University of Natal Durban.
- INSIGHTS, S. A. M. (2020).
- IQAIR, r. (2020).
- Jans, J., Johansson, P., and Nilsson, J. P. (2018). Economic status, air quality, and child health: Evidence from inversion episodes. *Journal of health economics*, 61:220–232.
- Jinabhai, C., Taylor, M., Coutsoudis, A., Coovadia, H., Tomkins, A., and Sullivan, K. (2001). A health and nutritional profile of rural school children in kwazulu-natal, South Africa. *Annals of Tropical Paediatrics*, 21(1):50–58.
- journal, M. . G. (2016).
- Kalisa, E., Clark, M. L., Ntakirutimana, T., Amani, M., and Volckens, J. (2023). Exposure to indoor and outdoor air pollution in schools in africa: Current status, knowledge gaps, and a call to action. *Heliyon*.
- Kampa, M. and Castanas, E. (2008). Human health effects of air pollution. *Environmental pollution*, 151(2):362–367.
- Katoto, P. D., Byamungu, L., Brand, A. S., Mokaya, J., Strijdom, H., Goswami, N., De Boever, P., Nawrot, T. S., and Nemery, B. (2019). Ambient air pollution and health in sub-saharan africa: Current evidence, perspectives and a call to action. *Environmental research*, 173:174–188.
- Kurita, H., Sasaki, K., Muroga, H., Ueda, H., and Wakamatsu, S. (1985). Long-range transport of air pollution under light gradient wind conditions. *Journal of climate and applied meteorology*, 24(5):425–434.
- Labour Department, o. (1995). Education labour relations council.

- Leiman, A., Standish, B., Boting, A., and van Zyl, H. (2007). Reducing the healthcare costs of urban air pollution: The South African experience. *Journal of Environmental Management*, 84(1):27–37.
- Lelieveld, J., Haines, A., and Pozzer, A. (2018). Age-dependent health risk from ambient air pollution: a modelling and data analysis of childhood mortality in middle-income and low-income countries. *The lancet Planetary health*, 2(7):e292–e300.
- Lepeule, J., Litonjua, A. A., Gasparri, A., Koutrakis, P., Sparrow, D., Vokonas, P. S., and Schwartz, J. (2018). Lung function association with outdoor temperature and relative humidity and its interaction with air pollution in the elderly. *Environmental research*, 165:110–117.
- Lesia, M. P., Aigbavboa, C. O., and Thwala, W. D. (2023). Factors influencing residential location choice in south africa: exploratory factor analysis (efa) and confirmatory factor analysis (cfa). *Journal of Housing and the Built Environment*, pages 1–28.
- Lu, H.-C. and Fang, G.-C. (2002). Estimating the frequency distributions of pm10 and pm2.5 by the statistics of wind speed at sha-lu, taiwan. *Science of the total environment*, 298(1-3):119–130.
- MacNaughton, P., Eitland, E., Kloog, I., Schwartz, J., and Allen, J. (2017). Impact of particulate matter exposure and surrounding “greenness” on chronic absenteeism in massachusetts public schools. *International journal of environmental research and public health*, 14(2):207.
- Makino, K. (2000). Association of school absence with air pollution in areas around arterial roads. *Journal of epidemiology*, 10(5):292–299.
- Mestry, R., Hendricks, I., and Bisschoff, T. (2009). Perceptions of teachers on the benefits of teacher development programmes in one province of South Africa. *South African journal of education*, 29(4).
- Mustapha, B. A., Blangiardo, M., Briggs, D. J., and Hansell, A. L. (2011). Traffic air pollution and other risk factors for respiratory illness in schoolchildren in the niger-delta region of nigeria. *Environmental health perspectives*, 119(10):1478–1482.
- Neidell, M. J. (2004). Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma. *Journal of health economics*, 23(6):1209–1236.
- Norman, R., Barnes, B., Mathee, A., Bradshaw, D., Collaboration, S. A. C. R. A., et al. (2007). Estimating the burden of disease attributable to indoor air pollution from household use of solid fuels in South Africa in 2000. *South African Medical Journal*, 97(8):764–771.
- NORTH WEST DEVELOPMENT CORPORATION, r. (2023).
- Noubiap, J. J. N., Essouma, M., and Bigna, J. J. R. (2015). Targeting household air pollution for curbing the cardiovascular disease burden: a health priority in sub-saharan africa. *The Journal of Clinical Hypertension*, 17(10):825–829.
- Operario, D., Cluver, L., Rees, H., MacPhail, C., and Pettifor, A. (2008). Orphanhood and completion of compulsory school education among young people in south africa: Findings from a national representative survey. *Journal of Research on Adolescence*, 18(1):173–186.

- Ostro, B. D. (1983). The effects of air pollution on work loss et morbidity. *Journal of Environmental Economics et Management*, 10(4):371–382.
- Park, H., Lee, B., Ha, E.-H., Lee, J.-T., Kim, H., and Hong, Y.-C. (2002). Association of air pollution with school absenteeism due to illness. *Archives of pediatrics & adolescent medicine*, 156(12):1235–1239.
- Petkova, E. P., Jack, D. W., Volavka-Close, N. H., and Kinney, P. L. (2013). Particulate matter pollution in african cities. *Air Quality, Atmosphere & Health*, 6(3):603–614.
- Ponczek, V. and Souza, A. P. (2012). New evidence of the causal effect of family size on child quality in a developing country. *Journal of Human Resources*, 47(1):64–106.
- Pope III, C. A. and Dockery, D. W. (2006). Health effects of fine particulate air pollution: lines that connect. *Journal of the air & waste management association*, 56(6):709–742.
- Pope III, C Arden et Burnett, R. T. e. T. M. J. e. C. E. E. e. K. D. e. I. K. e. T. G. D. (2002). Lung cancer, cardiopulmonary mortality, et long-term exposure to fine particulate air pollution. *Jama*, 287(9):1132–1141.
- Pratt, G. C., Vadali, M. L., Kvale, D. L., and Ellickson, K. M. (2015). Traffic, air pollution, minority and socio-economic status: addressing inequities in exposure and risk. *International journal of environmental research and public health*, 12(5):5355–5372.
- Ransom, M. and Pope, I. (2013). Air pollution and school absenteeism: Results from a natural experiment. *Manuscript, Department of Economics, Brigham Young University*.
- Ransom, M. R. and Pope III, C. A. (1992). Elementary school absences and pm10 pollution in utah valley. *Environmental research*, 58(1-2):204–219.
- Ritchie, H., Roser, M., and Rosado, P. (2022). Energy. *Our World in Data*. <https://ourworldindata.org/energy>.
- Röllin, H., Mathee, A., Bruce, N., Levin, J., and von Schirnding, Y. (2004). Comparison of indoor air quality in electrified and un-electrified dwellings in rural South African villages. *Indoor Air*, 14(3):208–216.
- Romieu, I., Lugo, M. C., Velasco, S. R., Sanchez, S., Meneses, F., and Hemandez, M. (1992). Air pollution and school absenteeism among children in mexico city. *American Journal of Epidemiology*, 136(12):1524–1531.
- Rondeau, V., Berhane, K., and Thomas, D. C. (2005). A three-level model for binary time-series data: the effects of air pollution on school absences in the southern california children’s health study. *Statistics in medicine*, 24(7):1103–1115.
- Roy, R. (2016). The cost of air pollution in africa.
- Rustein, S. O. and Johnson, K. (2004). The dhs wealth index.

- Simpson, R. (1990). A model to control emissions which avoid violations of pm10 health standards for both short and long term exposures. *Atmospheric Environment. Part A. General Topics*, 24(4):917–924.
- Spaull, N. (2015). Schooling in South Africa: How low-quality education becomes a poverty trap. *South African child gauge*, 12:34–41.
- Statistics South Africa, S. (2014). General household survey 2015.
- Statistics South Africa, S. (2015). General household survey 2015.
- Statistics South Africa, S. (2016). General household survey 2016.
- Tanmowo, M. (2000). Air pollution and respiratory health in africa: A review. *East African medical journal*, 77(2).
- Thambiran, T. and Diab, R. D. (2011). Air pollution and climate change co-benefit opportunities in the road transportation sector in durban, South Africa. *Atmospheric Environment*, 45(16):2683–2689.
- THE BORGEN PROJECT, r. (2020).
- The Education Labour Relations Council, E. (2003). Policy handbook for educators.
- Tiebout, C. M. (1956). A pure theory of local expenditures. *American Journal of Public Health*, 64(5):416–424.
- Timæus, I. M. and Boler, T. (2007). Father figures: The progress at school of orphans in South Africa.
- Timæus, I. M., Simelane, S., and Letsoalo, T. (2013). Poverty, race, and children’s progress at school in South Africa. *The Journal of Development Studies*, 49(2):270–284.
- Umwelthilfe, D. (2019). *no₂ report: Hotspots in germany, czech republic, slovenia, bulgaria and serbia and poland.*
- Vambe, M. T. (2005). Opening and transforming South African education. *Open Learning: The Journal of Open, Distance and e-Learning*, 20(3):285–293.
- van Donkelaar, A., Martin, R., Brauer, M., Hsu, N., Kahn, R., Levy, R., Lyapustin, A., Sayer, A., and Winker, D. (2018). Global annual pm_{2.5} grids from modis misr and seawifs aerosol optical depth (aod) with gwr 1998–2016. *NASA Socioeconomic Data and Applications Center (SEDAC)*.
- Vanker, A., Barnett, W., Nduru, P. M., Gie, R. P., Sly, P. D., and Zar, H. J. (2015). Home environment and indoor air pollution exposure in an african birth cohort study. *Science of the Total Environment*, 536:362–367.
- Visser, M., Juan, A., and Feza, N. (2015). Home and school resources as predictors of mathematics performance in South Africa. *South African Journal of Education*, 35(1).
- Wayne, W. S. and Wehrle, P. F. (1969). Oxidant air pollution and school absenteeism. *Archives of Environmental Health: An International Journal*, 19(3):315–322.

WHO (2016). classification of cities polluted.

Wichmann, J. and Voyi, K. (2012). Ambient air pollution exposure and respiratory, cardiovascular and cerebrovascular mortality in cape town, South Africa : 2001–2006. *International journal of environmental research and public health*, 9(11):3978–4016.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2):420–445.

Zhou, W., Qing, C., Guo, S., He, J., Deng, X., Song, J., and Xu, D. (2022). Labor off-farm employment and farmers' cooking clean energy use: Evidence from rural china. *Agriculture*, 12(7):972.

A Literature review : Air pollution and school absenteeism

Studies	Statistical / Econometric setup	Interest Variable	Dependent variable	Sample	Key findings
Wayne and Wehrle (1969)	Correlation Study	KI (potassium iodide)	Absence rates attributed to respiratory illness	Two elementary schools in Los Angeles Sept 17, 1962, to June 21, 1963 Country: USA	No significant effect of KI on school absenteeism
Ransom and Pope III (1992)	lagged moving average models Various unrestricted and polynomial distributed lag models	PM10 in /m3	Weekly and daily attendance	Valley located in Utah County 6 school districts Period of study: 1985-1990 Country: USA	Positive impact of PM10 on elementary school absences
Romieu et al. (1992)	Logistic regression mode	O3 (ozone)	respiratory related school absenteeism	The southwestern part of Mexico City 111 preschool children Country: Mexico	Positive impact of O3 on respiratory-related school absenteeism
Makino (2000)	Logistic regression Poisson regression Quintile method Correlation analyse	Suspended particulate matter (SPM) Nitrogen dioxide	Daily absence	Between 534-648 schoolchildren, Two school around arterial roads Period: 1993 to 1997 Country: Japan	Positive effect of SPM and nitrogen dioxide on school absence
Chen et al. (2000)	Time-series autoregression Analysis	PM10 CO O3	Daily elementary school absenteeism	57 elementary school of washoe county 27793 student Period: 1996 to 1998 Country: USA	Negative effect of PM10 Positive effect of CO and O3 on elementary school absenteeism
Gilliland et al. (2001)	Poisson log-linear model	PM10 O3 NO2 (nitrogen dioxide)	Daily absence	2081 children of 4th grade school Children of 12 southern California communities Period: first 6 months of 1996 Country: USA	Positive effect of O3 on respiratory illness-related absence rates No significant effect of PM10 and NO2
Park et al. (2002)	Generalized additive poisson regression	PM10 O3 NO2 SO2	Daily illness-related absence	One elementary school of Seoul, March 2, 1996 to december 22, 1999	Positive effect of PM10, O3 and SO2 No significant effect of NO2
Rondeau et al. (2005)	Logistic transition model with random effects	PM10 O3	Daily absence	12 Southern California communities 1932 children aged 9-10 years	No significant effect of PM10 No clear effect of O3
Currie et al. (2009)	Difference-in-difference-in-differences strategy	CO (carbon monoxide)	Aggregated six-week attendance period	The 39 largest school district of texas Period: 1996 to 2001 Country: USA	Positive effect of CO in elementary and middle school absence
Ransom and Pope (2013)	Wald-type estimator OLS and IV Method 2 instruments: Closure to the mill and atmospheric stagnation	PM10 CO	Weekly attendance	Provo school district and Northridge School of Utha Valley Period: 1986 to 1987 Country: USA	Negative or no significant effect of CO school absence (no clear effect) But positive effect of PM10
MacNaughton et al. (2017)	Generalized linear models Poisson function	PM 2.5	percent of students chronically absent at school	1772 public schools in Massachusetts 2012-2013 academic year Country: USA	Positive impact of PM2.5 on school absence

Table A1: Pollution and school absenteeism

B Descriptive statistics

Table B1: Descriptive statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
Absence	26143	0.06	0.24	0	1
PM 2.5	26143	117.47	67.45	16.06	221.42
Head education	23723	8.34	4.56	0	18
Head gender	26143	0.46	0.5	0	1
Child gender	26143	0.5	0.5	0	1
Child age	26143	9.89	3.08	5	15
Household size	26143	5.95	2.7	1	24
Presence of biological father=1	21505	0.39	0.49	0	1
Presence of biological mother=1	23785	0.75	0.43	0	1
Lack of teacher	25513	0.02	0.15	0	1
Wind speed in m/s	26143	3.32	0.39	2.94	4.33
Wind direction in degrees	26143	-59.87	65.08	-162.34	54.53
Temperature inversion	26143	-4.18	0.34	-4.41	-3.31
Rainfall in mm	26143	540.39	191.99	194.13	873.51
Year relative humidity	26143	51.7	8.49	39.37	68.13

C CF method framework

Consider the following structural equation:

$$y_1 = z_1\omega_1 + \gamma_1 y_2 + u_1 \quad (1)$$

z is a vector of exogenous variables and y_2 is the endogenous variable.

The exogeneity of z is given by the following orthogonality condition:

$$E(z'_j u_1) = 0 \quad (2)$$

with $j=1,2,\dots,k$ et $z=(z_1, z_2)$. Then, it is assumed that the elements of z are not perfectly collinear and thus the rank condition for (1) holds. Moreover, the rank condition is more easily stated in a linear form reduced by y_2 in the case of model 1. This linear form reduced by y_2 is as follows:

$$y_2 = z\phi_2 + v_2 = z_1\phi_{21} + z_2\phi_{22} + v_2 \quad (3)$$

$$E(z'v_2) = 0 \quad (4)$$

Furthermore, the rank condition is met in (1) if and only if $z_2\phi_{22} \neq 0$. This allows us to say there is an exogenous variable omitted from (1) which is partially correlated with y_2 .

Therefore, there is a correlation between the structural error u_1 and the error of the reduced form v_2 . One could therefore understand the "Function Control" approach through the following linear relation:

$$u_1 = \rho_1 v_2 + e_1 \quad (5)$$

$$E(v_2 e_1) = 0 \quad (6)$$

$$\text{With } \rho_1 = E(v_2 u_1) / E(v_2^2) \quad (7)$$

Furthermore, since u_1 and v_2 are not correlated with z then e_1 also is not correlated with z and y_2 .

In the end, we obtain a new following valid equation by inserting (5) into (1):

$$y_1 = z_1\omega_1 + \gamma_1 y_2 + \rho_1 v_2 + e_1 \quad (8)$$

In the CF method, v_2 in equation (8) is considered as a new explanatory variable. Its inclusion provides a new error term e_1 that is uncorrelated with other variables, including, y_2 . Indeed, the presence of v_2 in the equation allows "controlling" the endogeneity of y_2 . In other words, v_2 approximates the factors of u_1 that are correlated with y_2 .

D Wealth index construction: Principal component analysis

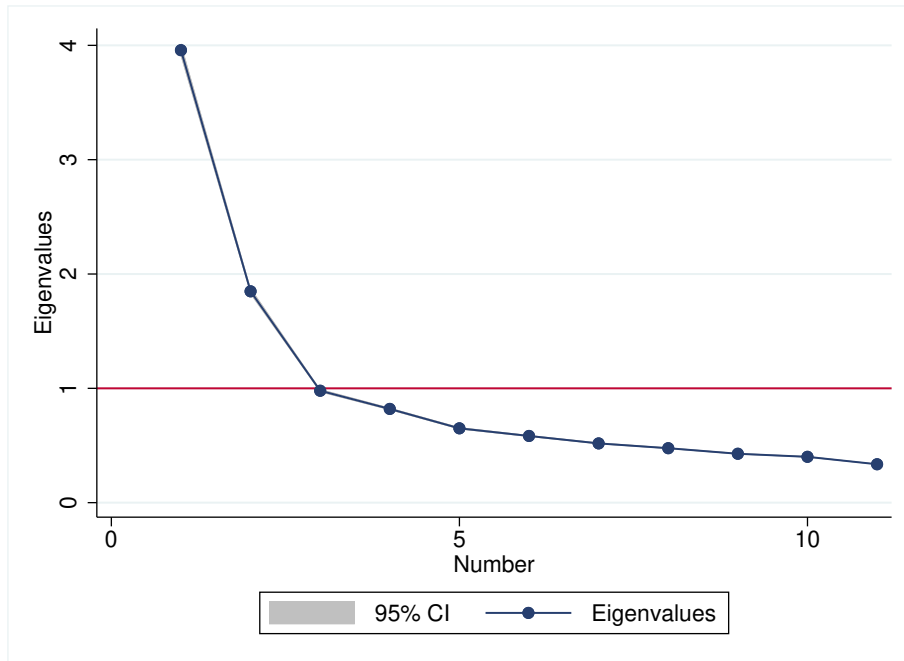


Figure C1: Components and eigenvalues

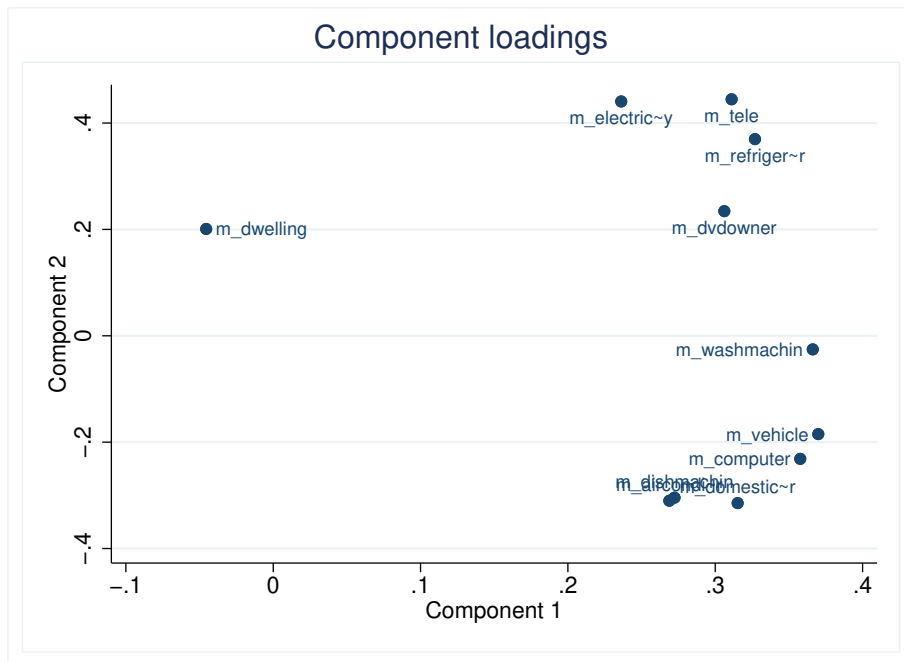


Figure C2: Selected components

E Data sources

Variables	Data sources
PM 2.5 concentration	van Donkelaar et al. (2018)
Socio-economic data	General Household Survey (Statistics South Africa, 2016, 2015, 2014) www.datafirst.uct.ac.za
Rainfall	Climatic Research Unit (CRU)
Temperature	ERA5 hourly data on pressure levels from 1979 to present. cds.climate.copernicus.eu
Humidity and Wind	ERA5 monthly mean data on single levels from 1979 to present. cds.climate.copernicus.eu