Disentangling the causal effects of air pollutants on health: when the numerous characteristics of the planetary boundary layer can help

Alexandre Godzinski,[1] Milena Suarez Castillo[2*]

[1]French Ministry of Environment and CREST
[2]Insee and CREST

May 23, 2019

[*Preliminary version - Do not circulate*]

Quasi-experimental evidence increasingly support the air pollution detrimental effect on health, but the question of which air pollutants impact which pathology remains largely open. In this paper, we exploit a novel sets of instruments: altitude weather conditions, and more specifically the characteristics of the planetary boundary layer (PBL). Numerous physical characteristics may influence differently the air pollutants' concentrations. By selecting optimal instruments from a large set of PBL characteristics with a *IV Lasso* procedure, we disentangle the impact of six air pollutants. We find that at least three air pollutants, ozone, sulfur dioxide and carbon monoxide, have a strong effect on respiratory diseases, independently of each other, even when controlling for the other pollutants in presence. Children are mostly affected. As for cardiovascular diseases and the mortality rate, our results point out to respectively carbon monoxide and particulate matters.

Key words : air pollution, health, emergency hospital admissions, planetary boundary layer, IV Lasso.

JEL code : C26; C55; I18; Q51; Q53

*We thanks the audience at the 10th French Econometrics Conference for their suggestions and useful comments. We are specially grateful to Frédérique Chéruy for the help with the LMDZ model, and to Frédéric Hourdin. We thank Pierre Bayart and Chantal Vilette at Insee for their help with the civil registry data. We thank Alireza Banaei, Fatma Kaci, Anne Bataillard, Max Bensadon and Francoise Bourgoin from ATIH for providing us the hospital admission database, Anne Laborie, Gilles Levigoureux and Frédéric Penven from ATMO France and the AASQA for providing us the air pollution datasets. Feedbacks are welcome at: alexandre *dot* godzinski *at* ensae *dot* fr & milena *dot* suarez-castillo *at* insee *dot* fr.

1 Introduction

Air pollution massively affects the environment we live in, thereby our health. Air pollution is multifaceted: the air breathed by urban population contains particulate matters of various sizes, and diverse gases. Quasi experimental evidence on mortality, in the real life setting, is accumulating (e.g. (*Currie and Neidell, 2005*), (*Chay and Greenstone, 2003*)). Morbidity has recently received significant attention.[1] But so far, the impact of the very diverse pollutants has mostly been studied separately, without accounting for the other pollutants in presence. Pollutants' concentrations often vary together: disentangling their separate impact is a difficult task. Meanwhile, some are strongly anticorrelated (e.g. nitrogen oxides and ozone), which makes the approximation of considering air pollution as a whole dubious. Moreover, from a public policy perspective, reducing one pollutant emissions or another may imply very distinct regulation or taxation, at least because they may come from distinct sources. In a context where not all pollutants' concentrations follow the same time trends and are not all tackled as efficiently, it is highly relevant to document further their separate health effects. For instance, in France over the last fifteen years, while sulfur dioxide concentration is strongly going down, followed sluggishly by particulate matters and nitrogen oxides, ozone maintained its level and even slightly increased since the early 2000s.

In this paper, we use a novel set of numerous instruments combined with optimal instruments selection to disentangle the separate effects of six pollutants from each other. Recent techniques, as put forward in (*Belloni et al., 2012*), allow us to select instruments in an optimal way, avoiding ad-hoc selection and enhancing precision in a setting where we decisively need it. Indeed, to our knowledge, no study has succeeded in precisely measuring the impact of several pollutants at the same time.[2] For instance, if (*Arceo et al., 2016*) derive convincing evidence of the impact of air pollution on infant mortality in Mexico City, they are not conclu-

[1] For example, papers leveraging some exogenous variations from the transportation sector include (*Schlenker and Walker, 2015*) and (*Moretti and Neidell, 2011*), who study pollution shocks (respectively from carbon monoxide and ozone) due to plane idling or boat traffic near airport and harbour

[2] One exception is (*Schlenker and Walker, 2015*) which find an effect of CO on respiratory health nearby Californian airports, even when controlling for NO2. We here consider on top of these two, four pollutants, importantly ozone and particulate matters.

sive on which pollutant(s) to incriminate, be it particulate matters or carbon monoxide. Going further than the impact of an aggregate measure of air pollution is still an open question in the small but growing literature linking causally health and air pollution. The main challenge lies in finding enough appropriate instruments that influence differently distinct air pollutants. The need for instrumentation comes from various sources of possible confounders to draw a causal chain from air pollution to health, such as socio-economic and economic levels or weather variations.[3] Some of this bias can be accommodated with fixed effects (seasonal, location-specific). Some of these variables (e.g. weather characteristics) are observed so can be controlled for. But without instrumentation, functional forms hypotheses and measurement errors may become a first-order problem.

The instruments we use are derived from numerous altitude weather characteristics: some of them are well known to impact air pollutant concentrations independently of human activities. Urban air pollution mainly comes from anthropogenic sources, but the atmosphere dynamics, through wind and sunlight for example, plays a key role in its mixing, chemistry and dispersion and thus in ambient air pollution inhaled by population. Instruments from altitude weather variables are not new: thermal inversion phenomena (see below for more details) were used by (*Jans et al., 2018*), (*Arceo et al., 2016*) and (*Chen et al., 2018*), while surface weather conditions as wind strength and directions were used by (*Deryugina et al., 2016*) and (*Anderson, 2015*). The originality here is to use a large set of altitude[4] weather conditions, and let the data speak on the strongest relationships. We also rely on a very simple novel instrument: the height of the planetary boundary layer, under which air pollutants are trapped:[5] roughly, the higher the height of the planetary boundary layer, the larger the air volume available for pollutants, and the lower the concentration. Indeed, in a first step and before calling selection techniques, we show thanks to this single instrument that air pollution understood as a whole, measured with

[3] Quasi experimental evidence complement other studies by paying considerable attention at dealing with confounding factors (*Currie et al., 2011*).

[4] We use altitude weather conditions and not surface weather conditions as instruments to ensure the validity of the exclusion restriction assumption.

[5] In fact, (*Schwartz et al., 2016*) use PBL height and wind strength as an instruments to study daily deaths caused by air pollution in Boston. Our first set of results is derived under a method which is quite close to this paper (a classical IV), but we here consider on top of mortality, emergency admissions, and improve on the external validity of the study by considering the ten largest urban areas of France.

an index, has a detrimental effect on health. But we might find many other and more complex phenomena linking altitude weather variables to ground-level pollution, even conditional on ground-level weather, by leveraging the rich set of instruments at hand. To that end, we use optimal instruments selection in a high dimensional set of altitude weather variables. We heavily rely on the econometric theory by (*Belloni et al., 2012*) and (*Chernozhukov et al., 2015*).[6] The instrument which has first been put forward by simple physics, the inverse of boundary layer height, should also be selected as dictated by the data. We expect, on top of its selection, to capture relationships between atmosphere parameters and pollutants in a data-driven manner, hopefully leveraging different relationships for each pollutants.

We rely on five rich datasets that provide a daily perspective from 2010 to 2015 in the ten largest urban areas of France. We observe six pollutants monitoring: particulate matters of less than 2.5 micrometers PM2.5, of less than 10 micrometers PM10, carbon monoxide CO, nitrogen dioxide NO2, ozone O3, sulfate dioxide SO2. Ground-level weather data comes from Météo France. As for health, we observe emergency admissions in hospital belonging to the urban areas under consideration. For altitude weather variables, we rely on the output of a general climate model LMDZ, from the *Laboratoire de Météorologie Dynamique*.[7] This model among others contributes to fueling Intergovernmental Panel on Climate Change (IPCC)[8] reports (See (*Hourdin et al., 2006*) and (*Dufresne et al., 2013*)).

Our analysis is twofold. First, we show that air pollution affects urban population health through several outcomes: respiratory and cardiovascular emergency admissions, but also mortality rate. Moderate variations of air pollution, at modern levels in Europe urban areas, are therefore costly. For the method, we rely on a simple instrument which is suggested by the atmosphere dynamics and has well-known effects on air pollution. In this first setting, air pollution is understood as an aggregate of distinct pollutants. Second, optimal instruments selection with LASSO leads to different selected variables per pollutants, therefore suggesting sufficiently distinct first order relationships between pollutants and atmospheric conditions. The

[6]For another application, in a different context, see (*Gilchrist and Sands, 2016*)
[7]`http://lmdz.lmd.jussieu.fr/`
[8]In French, GIEC, *Groupe d'experts intergouvernemental sur l'évolution du climat*

instrument which has first been put forward by simple physics, the inverse of boundary layer height, is also selected in the very first variables for three pollutants, as found by the data. This reduced set of instruments is then used in a two-stage least squares with several pollutants.

Our results shows that ozone, sulfur dioxide and carbon monoxide do have an effect on respiratory health, independently from each other and even after controlling for the other pollutants. Quantitatively, we find 4.5% more respiratory admissions when O3 goes up by + 10 $\mu g/m^{-3}$ (or about half a standard deviation), 5.5% more respiratory admissions when SO2 goes up by + 1 $\mu g/m^{-3}$ and 0.6% more respiratory admissions when CO goes up by + 100 $\mu g/m^{-3}$ (or about half a standard deviation). Most of these aggregate effects are driven by emergency admissions of newborns, infants and young children. In addition, we find an effect of PM2.5 on the mortality rate (+ 10 $\mu g/m^{-3}$ leads to +1.8% of the mortality rate) and of CO on cardiovascular diseases emergency admissions (+ 100 $\mu g/m^{-3}$ leads to +1.5% emergency admissions). As a falsification test, we perform similar analysis on an unrelated yet very common emergency admissions: digestive diseases. By contrast with the other health indicators, the absence of significant effects comforts the causal link.

The article proceeds as follows. In the second section, we start by presenting jointly the data and the mechanisms at work. Then, we present and discuss the empirical strategy and the instruments' selection procedure in the third section. Finally, we present our results and then conclude.

2 Data and background

In this section, we describe the data sources which have all in common the following scope: the ten most populated urban areas in France over the 2010-2015 period. Table 1 reports the population, and Figure 1 the geographical location and extension of urban areas. The largest urban area is the Paris region where more than twelve millions people live. Most of the other urban areas have about a million inhabitants. The urban areas are well spread out on the French territory.

Table 1: The ten most populated urban areas in France

Urban area	Population in thousands		
	All age	0-4	over 70
Paris	$12,470$	845	$1,203$
Lyon	$2,259$	152	249
Marseille - Aix-en-Provence	$1,744$	103	231
Toulouse	$1,312$	81	137
Bordeaux	$1,195$	67	135
Lille	$1,182$	80	111
Nice	$1,006$	52	171
Nantes	934	61	97
Strasbourg	777	45	89
Rennes	708	46	70

For Lille and Strasbourg urban area, only the French part is considered.
Source: 2013 census

Within these urban areas, many cities do have worrying air pollution levels. Figure 3 in appendix shows the annual mean of particulate matters in cities belonging to the ten urban areas, compared to WHO guidelines. The vast majority of cities and in particular the most populated do not respect the guidelines for yearly means in 2014. For instance, Rennes, the smallest urban area in our sample, do not respect the guidelines relative to particulate matters of less than $2.5 \mu m$.

2.1 Atmospheric weather characteristics

Mechanisms. The air near the ground (the planetary boundary layer, PBL) is sensitive to friction forces with the surface. These forces become negligible in the upper layers where wind circulation is global (the free atmosphere). In short, the planetary boundary layer (PBL) is the layer of air that is stuck to the ground. Pollutants are trapped within this vicinity of the earth, whose depth is defined by its height. The concentration of pollutants trapped in the layer depends on its height, with a dilution effects on the vertical axis. When higher, the volume of air available for pollutants is higher, thus their concentration lower. In this paper, we first rely on this simple physics phenomenon, namely the variations of PBL height, to causally evidence a detrimental impact of air pollution on the respiratory system of urban population. This phenomenon may be rather similar from one pollutant to the other, so one can only use it as an

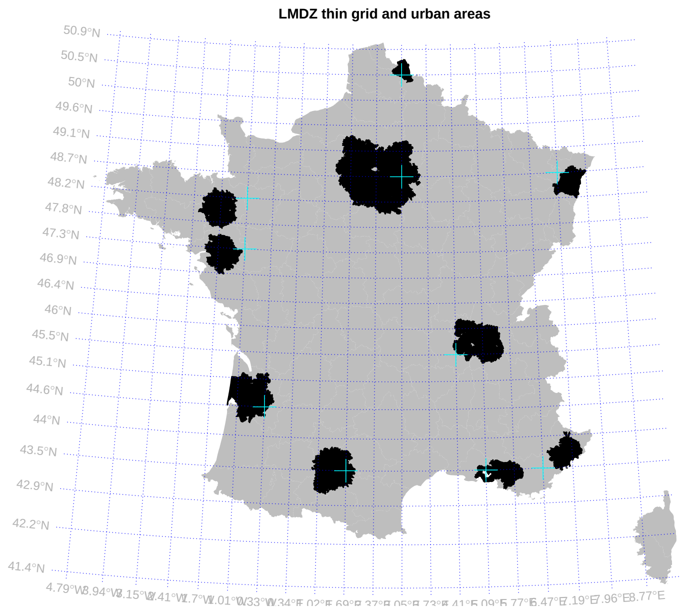

Figure 1: Geographic location of the ten most populated urban areas in France (black areas), LMDZ grid (blue dotted lines) and points representing the urban areas on this grid (cyan cross). (Source: Insee, 2010)

instrument for ambient air pollution as an aggregate statistics.

PBL height varies according to various factors. First, the height of the planetary boundary layer responds to heating flux between the sun and the earth and therefore display a diurnal pattern. Second, the planetary boundary layer height reacts to *subsidence*, which brings the top of the layer downward in a high pressure diverging area. Third, it may be modified when a horizontal movement of cold air brings it under a warmer layer of air (*frontal* inversion at the top of the planetary boundary layer).[9] If some of these phenomena do have a seasonal nature and are partially related to ground-level weather, there is no reason to expect that health would be affected by these phenomena *conditional* on seasonal and ground-level weather conditions. This makes PBL height a strong candidate for instrumentation.

Planetary boundary layer height is often defined by the presence of a thermal inversion at the top: the temperature, which usually decreases with height, sharply increases at the top of the

[9]See (*Stull, 2016*) for further details.

PBL.[10] A thermal inversion acts as a lid over the air motion beneath, because an air parcel which is cooler than its environment tends to move down. Its role over pollutants' concentrations is widely acknowledged. Thermal inversions are thus closely related to boundary layer height (during the day, it is a thermal inversion that defines the boundary layer height) but they may be multiple and varying in strength within the boundary layer height. During a thermal inversion, polluted air is trapped beneath the inversion height (a warmer layer of air blocks the vertical movement). Thermal inversions have been used by other authors to instrument air pollution (on mortality in developing countries city (*Arceo et al., 2016*) and closer to us, (*Jans et al., 2018*)). However, whereas thermal inversions may or may not happen (dummy variable), the height of the planetary boundary layer may always be defined (and is a continuous variable).

Therefore, aside its height, other characteristics related to the planetary boundary layer may influence directly pollutants' concentrations: among other, winds and thermal inversions phenomena. Interactions may play a strong role. There is a wide set of potential candidate which fulfils the conditional exclusion restriction.

Data. The data come from the LMDZ model (*Hourdin et al., 2006*), a general circulation model maintened by the *Laboratoire de météorologie dynamique* (Z is for zoom).[11] It simulates the full atmosphere over a 3D grid. The development of the model is tested and improved by comparison with atmospheric observations (field or satellite data).

We were provided the output of an hourly reconstitution of the atmosphere dynamics from 2010 to 2015 along a grid $\approx$ 50km x 50 km in which cities are located (the model is used with a zoomed version over France, see Figure 1). Many variables are present, most importantly, the boundary layer height; but also along a vertical grid parameterizing altitude through pressure levels, wind characterizations (direction, strength), humidity, temperature, and altitude corresponding to the pressure levels. In particular, the measurement of temperature at differ-

[10] During a thermal inversion, warmer air is held above cooler air; the normal temperature profile with altitude is inverted.

[11] The LMDZ model is the atmosphere component of the climate model described in (*Dufresne et al., 2013*) and used for IPCC reports; see http://lmdz.lmd.jussieu.fr/ for a general presentation

ent altitudes allows to reconstitute thermal inversions indicators within the boundary layer, as we observe temperature gradient. We aggregate the variables at the daily level (considering different statistics or hour-of-the-day) to match health data.

A large set of potential instruments. The PBL height is directly obtained from the model output at the hourly level between the 01/01/2010 to the 01/01/2015. We consider a daily measure specific to six moments of the day: 0 to 4a.m., 4 to 8 a.m, $\cdots$, until 8p.m. to midnight. As for thermal inversions, we compute thermal inversions from temperature altitude profile. At the hourly level, we define a thermal inversion when at least 50% of the layers between 10.1 hPa ($\approx$200 m) and 89,7 hPa ($\approx$ 1,2 km) have temperature above temperature at ground level (in the lower layer). We also consider a simpler measure of inversion similar to (*Jans et al., 2018*): we take only two layers (ground-level and at 98.1 hPa $\approx$450m) and define a thermal inversion when ground-level temperature is the lowest, and an additional measure of thermal inversion strength: the difference of these two temperatures. Then, we average these two hourly dummies coding the presence of an inversion, and its strength, at six moments of the day, as for PBL height. For the other weather variables varying with altitude (humidity, temperature, pressure, zonal wind, meridian wind, wind strength), we compute the average at the daily level and exclude all measures below 98.1 hPa to assert exogeneity.

2.2 Ground-level weather data

Weather conditions play a key role in human activity and air pollution formation, but also directly on health (*Deschênes and Greenstone, 2007*). Plus, ground-level weather data is likely correlated to atmospheric data, so that it is important to condition on ground-level weather in our regression, the assumption being that high altitude atmospheric variations are exogenous with respect to health, except through pollution, conditional on ground-level weather controls. We hence consider a full set of weather conditions. Data come from Météo France and are available on an hourly basis for our ten urban areas. We consider seven weather parameters : temperature, rainfall, wind speed, wind direction, insolation and humidity. We also consider three day-level dummies coding for (1) snowfall, (2) presence of a snow cover over the ground,

(3) presence of ice. Measurement stations are located at nearby airport[12], except for Paris, where the measurement station is located in a garden in the center of Paris (parc Montsouris).

2.3 Pollutant data

Air quality is measure by regional associations called AASQA (*associations agréées de surveillance de la qualité de l'air*), which are grouped in a national federation called ATMO France. They are approved by the Ministry of Environment which delegates the mission of air pollution surveillance of "regulated" pollutants. They operate numerous air quality measurement stations all over France. We consider the stations located in the 10 more populated urban areas. We focus on a rich set of air pollutants: the 6 pollutants that are widely available on an hourly basis are carbon monoxide (CO), particulate matter of less than 2.5 micrometers (PM2.5), particulate matter of less than 10 micrometers, nitrogen dioxide (NO2), ozone (O3) and sulfur dioxide (SO2). We usually have data for several measurement stations per urban area,[13] which we average at the urban area and daily level on a constant set of monitoring stations.

When using a single instrument and due to the strong correlation between air pollution emissions, we can only gauge the impact of ambient air pollution as a summary statistics of available pollutants' concentrations. We thus create a pollutant index with a principal component analysis over the 6 standardized pollutants concentration and keep the first component as the pollution index.[14] As we have missing values in pollutant concentration, the PCA is combined with an EM algorithm to deal with missing values (See (*Josse and Husson, 2012*)). The index is therefore available eventhough one or several pollutant concentrations are missing (keeping observations for which all six pollutants are observed leads to drop 80% of the sample).

Table 2 shows how pollutants are correlated. Two important points should be noted for what follows. First, PM2.5 are a subsample of PM10 (60 to 70% of PM10 particulates are PM2.5

[12] Specifically for insolation in Lille, we use the measurement station in Lillers, nearby Lille, as this parameter was not available in Lille-Lesquin airport station over the whole studied period.

[13] We have at least one measurement station for each pollution in each urban area.

[14] (*Arceo et al., 2016*) use a similar index approach when considering the joint effect of PM10 and CO.

particulates according to Airparif[15]). To preview our results, we will not be able to disentangle their effect separately as we will find no clear distinction in their response to our instruments. Second, O3 is anticorrelated with all pollutants, in particular to its precursors NO2 and CO. On average, high levels of nitrogen oxides are associated to low levels of ozone. This emphasizes the multi-dimensional aspect of air pollution, which should ideally not be treated as a whole.

Table 2: Correlation between pollutants' concentrations

	Pollution index	PM2.5	PM10	NO2	O3	CO	SO2
Pollution index	1	0.84	0.80	0.65	−0.56	0.67	0.31
PM2_5	0.84	1	0.83	0.40	−0.32	0.46	0.27
PM10	0.80	0.83	1	0.53	−0.14	0.40	0.26
NO2	0.65	0.40	0.53	1	−0.22	0.69	0.24
O3	−0.56	−0.32	−0.14	−0.22	1	−0.40	−0.08
CO	0.67	0.46	0.40	0.69	−0.40	1	0.22
SO2	0.31	0.27	0.26	0.24	−0.08	0.22	1

The dynamics of O3 is singular for several reasons. It is link to the fact that O3 is a secondary pollutant: NO2 is precursor of O3 in the reaction $NO2 + O2 \leftrightarrow NO + O3$. There is at least two effects under the anticorrelation of O3 with the other pollutant: NO2 disappears on the process of producing O3 in a slow reaction (to a lesser extent it is also the case for CO). Additionally, primary pollutant NO is unstable and reacts quickly with O3, and it usually produced in conjunction with PMs by traffic. The latter is known as the urban decrement: primary pollution can at first reduce the concentration in O3 at the local level.[16]

These elements emphasize how ambient air pollution is multifaceted. In particular, ozone is not easily captured by a pollution index.

[15] Bilan de la qualité de l'air 2017

[16] A simple way to explain it from (*Munir et al., 2012*): *At the local level freshly emitted nitric oxides (NO) produced by road-traffic react with ozone molecules and produce nitrogen dioxides (NO2). Hence road-traffic provides a local sink for ground level ozone resulting in ozone concentration in urban areas being lower than the surrounding rural areas. This phenomenon of lower ozone concentration in urban areas is referred to as ozone urban decrement.*

2.4 Health data

The data is obtained from the ATIH (*Agence Technique de l'Information Hospitalière*) that gather an administrative and exhaustive database which record all admissions in public and private hospitals. Its primary use is to compute hospitals' funding based on their activity. By final diagnostic and by urban area in which the hospital is located, we were provided the daily count of emergency admissions. Further, this information breaks down by age groups: newborns (0-28 days), infants ($\geq$ 29 days, $\leq$ 1 year-old), and a 5-years breakdown (0-4, 5-9, up to 75-79 plus over 80). More precisely, an emergency admission is an entrance through the hospital emergency unit that led to an admission from patients coming from their residence (i.e. not transferred from another hospital) or from public space.[17] Therefore, programmed admissions, long-term and recurring care are excluded. The diagnostic used here is coded at the end of the patient stay. It represents the main diagnostic which gave rise to the highest care resources. When and only when a diagnostic is not reached, the code is relative to the observed symptoms. We divide the daily count of admissions by the age-range and urban-area corresponding population (2013 legal population produced by Insee). Our variable of interest is the emergency rate of admission per 100 000 inhabitants.

In addition, we consider mortality rates constructed from daily records in civil registry (as produced by Insee at the municipality level), which are aggregated for each of our urban area and by age groups. After similarly normalizing with the legal population, our variable of interest is the mortality rate per 100 000 inhabitants.

3 Empirical strategy

We are thus faced with a large set of potential variables whose exogeneity with respect to health is asserted with the same reasoning: the maintained exclusion restriction is that unexpected altitude weather phenomena, once controlled for unexpected ground-level weather, do not influence health except through air pollution. Unexpected shocks are variations out of seasonal

[17] When due to hospital organization, emergency room is the main entry point, doctors should not use the code "emergency" systematically but only when the individual situation in the views of the patient, his relatives or his general practitioner is an emergency.

variations. These fluctuations in altitude variables are governed by air movement rules (pressure differentials, geographical landscape, heat differential absorption...) which are unlikely to affect health by themselves. Therefore, the exclusion restriction maintained here applies to a high number of instruments, which are possibly interacting (for example, solar radiation may favor ozone production all the more than atmospheric conditions are stable, e.g. no wind). Some of them are known to be related to air pollution, some of them are likely weak or close to noise when it comes to form the conditional expectation of pollutants given the instruments.

3.1 Instruments and first stage

Unpredicted shock (unsual component). First, for illustrative purpose, we define the unusual air pollution component by $\hat{p}_{ct} = P_{ct} - \hat{P}_{ct}$ the residual from the linear regression run at the city c and date t level:

$$P_{ct} = X_{ct}b + \alpha_{d,c} + \beta_{my,c} + p_{ct}$$

where P_{ct} is a given pollutant concentration, X_{ct} are ground-level variables controls[18] and $\alpha_{d,c}, \beta_{my,c}$ are respectively day-of-the-week, and month-year fixed effect which are specific to the city,[19] to capture usual pattern of pollution in our period. p_{ct} is the unusual pollution shock. We proceed the same way for the other variables. By linking the unusual components, this helps us show how the inverse of PBL height and pollutants' concentrations are strongly associated, even after partialling out seasonal mouvements and ground-level weather.

Planetary boundary layer height as an instrument. Figure 2 shows how an increase from decile to decile of the inverse of PBL height leads to a close to linear increase of 5 out of 6 pollutants' concentrations, as expected from a vertical dilution effect where pollutant concentrations would be inversely proportional to boundary layer height. These figures are built upon the unusual component of both variables, to emphasize that the link between both variables does

[18]Precisely: a polynomial of order two for temperature, rainfall and wind strength; linear controls for humidity and sunshine and 3 dummies encoding the presence of snow, if snow completely covers the ground, if there is ice.

[19]To abstract from seasonality, city-specific seasonal patterns are relevant as shown in Appendix, Figure 5. Most urban areas have a maximal PBL height in summer (which also correspond to the lowest air pollution level), but the two urban areas on the mediterranean coast display the reverse pattern. To focus on the unusual component of pollution and altitude weather, city-specific month-year effects are important.

not arise merely from seasonality. Nevertheless, without controls and seasonal fixed effects, the same patterns can be observed (see Figure 4 in appendix).

Figure 2 clearly shows how all pollutants' concentrations except O3 are driven upward when the boundary layer height goes down, conditional on weather controls and city-level temporal patterns of pollution. As mentioned above, it is not surprising that the dynamics of O3 is singular, as it is a secondary pollutant which may be consumed by nitrogen oxide in highly polluted area. For five pollutants out of six, there is a strong and physically grounded effect of PBL height on ground-level pollutants' concentrations. This instrument has very attractive features, but cannot instrument alone the independent effects of distinct pollutants.

Instrument selection. From the very large set of potential instruments, we intent to perform an optimal selection in the spirit of (*Belloni et al., 2012*). Optimal selection should be understood as unveiling a true predictive power, not as an unprincipled overfitting of the data at hand. Model selection is performed thanks to the LASSO (Least Absolute Shrinkage and Selection Operator, (*Tibshirani, 1996*)). It introduces a penalization to the OLS objective, the errors' sum of squares, by adding a scalar penalty multiplied by the l_1-norm of the (possibly high-dimensional) parameter of interest. The solution has a limited number of non-zero coefficients, whose number depend on the penalty level: as such it performs model selection. In our setting, the high dimensional parameter is the effect of *many* altitude weather characteristics on pollutants' concentrations, that is the first stage of an IV model explaining the health effects of air pollutants (the endogenous variables). (*Belloni et al., 2012*) show how to choose the penalty to insure asymptotic convergence and inference in a IV-setting where LASSO is used to select instruments in a first step. In (*Belloni et al., 2014*), the authors explain how focusing on the predictive part of the IV problem, namely the first stage, help in deriving inferential results while building on models whose initial goal was prediction and not structural estimation. In such a setting, model selection error is not a problem in itself (as long as other valid instruments are available and selected). For our problem, the method is attractive as complex relationships between altitude weather variables and air pollution may be recovered from the data. It avoids an ad-hoc choice of variables and ensure that the selection is reproducible. At the same time, a strong first stage should improve the precision of our estimates. We complement this approach

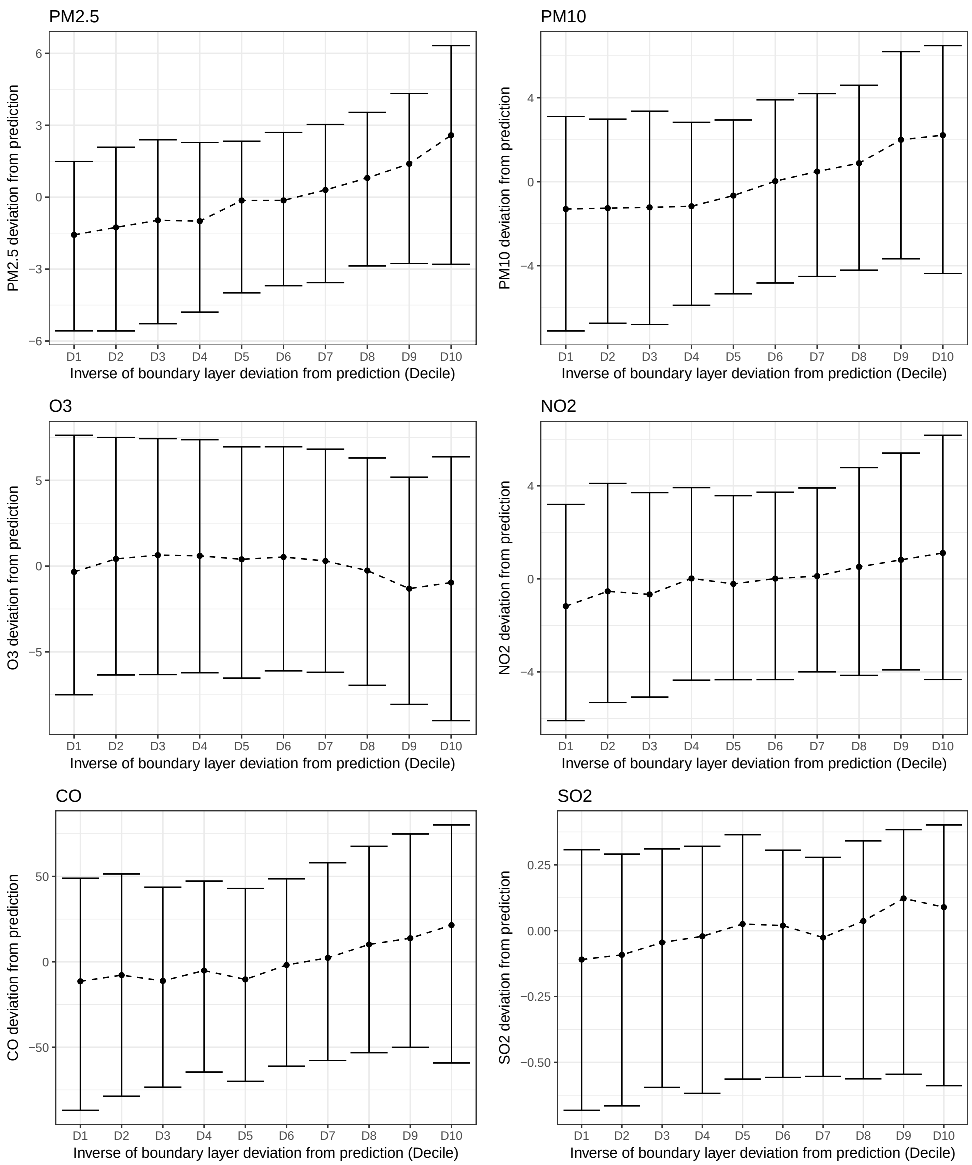

Figure 2: Unusual component of pollutants concentration and unusual component of inverse boundary layer height.

Note: ”Unusual” refers to the deviation of the variable from a set of weather and seasonal controls. For each decile of unusual inverse boundary layer height, are represented the mean and the lower and upper quartile of unusual pollutant concentration.

with a classical IV.

For each pollutant, which instruments will be selected? In a LASSO model, decreasing the penalty leads to include gradually more variables, possibly with several variables entering at the same time. We here present the first selected variables to explain distinct air pollutants, that is the first LASSO model with one or more variables selected. There is no selection on ground-level weather variables, that are forced into the model, to respect the conditional exclusion restriction. After selection, we run simple OLS on selected instruments and the same weather variables controls.[20] This is known as post Lasso estimation and alleviates the Lasso bias which shrinks point estimates toward zero. Results are presented in Table 3: each pollutant is regressed over the selected variables. From Table 3, we evidence that different instruments are selected depending on the pollutant in consideration. The exceptions are PM2.5 and PM10: the first models one obtains after Lasso are very similar for both pollutants (out of the 3 variables selected first for PM2.5, two are also selected first for PM10).

It suggests distinct relationships between these altitude weather variables and each pollutants, which is a first requirement to be able to disentangle the role of the various pollutants. Moreover, it is worth noting that the inverse of PBL height is selected within the first most predictive instruments for most pollutants out of six, although averaged at distinct time of the day, with the expected sign. Thermal inversions are found particularly predictive for nitrogen oxides and ozone and are thus also good predictors of ground-level air pollution, in line with existing literature. For particulate matters, altitude zonal wind coming from the west (resp. from east) predicts a lower (higher) concentrations, which is coherent with clean oceanic winds from the west and/or polluted air imported from the eastern regions. Note that Table 3 is not our first stage regression, which will pool together any selected instruments within the instrument set (more details are provided below).

In the following, we detail our two related IV strategies. The first builds on a classical IV with PBL height instrumenting for a pollution index (or a given pollutant) which we refer to as

[20]Beforehand, seasonal fixed effects are withdrawn from both pollutants and instruments.

Table 3: Lasso selection of instruments per pollutants and post-lasso estimation

	Dependent variable:					
	Pollutant concentration					
	PM2.5	PM10	NO2	O3	CO	SO2
Today inverse of PBL height						
... between 8p.m. and 12p.m.				$-1{,}667.248^{***}$		
				(126.565)		
... between 16p.m. and 20p.m.			$1{,}314.111^{***}$		$23{,}070.760^{***}$	
			(58.639)		(812.958)	
... between 8p.m. and 12p.m. in Paris						402.593^{***}
						(33.031)
Yesterday inverse of PBL height						
... between 8p.m. and 12p.m.	974.705^{***}					
	(64.540)					
... between 16p.m. and 20p.m.	$1{,}140.015^{***}$	$1{,}893.643^{***}$				
	(55.117)	(61.981)				
Thermal inversions						
... presence between 8 and 12a.m.				-3.410^{***}		
				(0.752)		
... strength (day average)			-0.106			
			(0.127)			
... strength between 4 and 8a.m.			0.722^{***}			
			(0.063)			
... strength between 8 and 12a.m.				-0.196		
				(0.156)		
Altitude wind						
Zonal ($W \to E$) strength at 97.5 kPa	-0.326^{***}	-0.365^{***}				
	(0.012)	(0.016)				
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,046	14,178	14,824	15,917	14,064	14,775
R^2	0.377	0.306	0.508	0.449	0.304	0.053

This table presents post-lasso models by pollutants. Before all regressions, we partial out fixed effects. For each pollutants, a first step of instruments' selection is performed with no selection on weather controls. The penalty is chosen so as to obtain the first model with one or more variables selected. Then, OLS is run per pollutant on the set of selected instruments, which are shown here. Significance: *p<0.1; **p<0.05; ***p<0.01

monopollutant models. It shows the detrimental health effect of air pollution understood as a whole. The second uses Lasso to optimally select instruments in multipollutant models. The goal is to disentangle separately the effect of distinct pollutants.

3.2 Monopollutant models

In a first step, we use an IV strategy with a single instrument. Among our set of potential instruments, the more natural and with a well-identified physical mechanism is the inverse of boundary layer height. In a first part, we use it as an instrument with the following first stage:

$$P_{ct} = IBL_{ct}\eta + X_{ct}b + \alpha_{d,c} + \beta_{my,c} + \epsilon_{ct} \tag{1}$$

P is either a given pollutant concentration or a pollution index, $X_{c,t}$ the aforementioned ground-level weather controls and finally $\alpha_{d,c}$ and $\beta_{my,c}$ the seasonal fixed effects (day-of-the-week $\times$ city and month-year $\times$ city). In practice, IBL_{ct} refers to 2 instruments: the inverse of boundary layer height at date t and at $t-1$. The first stage step is reported in Table 4, which confirms Figure 2 and the strength of the inverse of PBL height and its lag as instruments. All pollutants except SO2 respond strongly to PBL height. O3 responds in the opposite way compared to the other pollutants, probably because of the increases in concentration of nitrogen oxides.

The second stage writes as follows:

$$R_{ct}^{(p)} = P_{ct}\delta + X_{ct}d + a_{d,c} + e_{my,c} + \nu_{ct} \tag{2}$$

with R the rate of admissions in emergency per 100 000 inhabitants in city c and date t for a given pathology p and with the same set of weather controls and seasonal patterns. Importantly, our exclusion restriction condition is that altitude weather variables influence air pollution at given ground-level weather, and should not impact health directly. Therefore, we will draw particular attention to robustness checks in specifying ground-level weather controls. We estimate this model by two-stage least squares regressions with standard errors are clustered at the city level.

Table 4: Inverse of PBL height and pollutant concentrations.

	Dependent variable:						
		Concentration ($\mu g/m^{-3}$)					
	Pollution index	PM2.5	PM10	NO2	O3	CO	SO2
Instruments							
Inverse of PBL Height	218.522***	1,700.531***	1,869.372***	1,434.486***	$-$2,607.193***	28,909.270***	45.989
	(24.353)	(265.677)	(306.226)	(236.517)	(516.551)	(4,698.262)	(49.730)
Lag of inverse of PBL H	206.628***	2,365.942***	2,512.861***	468.835**	$-$1,484.590***	12,382.830***	36.367
	(30.039)	(266.420)	(321.395)	(236.677)	(309.218)	(3,637.090)	(43.691)
Observations	21,399	16,046	14,178	14,824	15,917	14,064	14,775
F-statistics	1,356	993	659	351	12	173	387
R^2	0.737	0.614	0.605	0.824	0.821	0.778	0.421

All regressions includes month-year and day-of-the-week fixed effects, interacted with city fixed effects, and weather controls. Standard errors are clustered at the city level. The F-statistics corresponds to the hypothesis of joint nullity of the two instruments. Significance: *p<0.1; **p<0.05; ***p<0.01

3.3 Multipollutant models and optimal instruments

Pollutant concentrations tends to be very correlated. An additional challenge to tackle is to measure the separate impact of distinct pollutants on health. To that end, we search for optimal instruments within our large set of potential candidates, all a priori exogenous conditional on ground-level weather, because measured in altitude. As for the econometric theory and inference, this part heavily rely on (*Belloni et al., 2012*) and (*Chernozhukov et al., 2015*).

To abstract from seasonality, we first take out the estimated seasonal fixed effects from any of the variables considered in the following equations: pollution, ground-level and altitude weather characteristics, emergency admissions. Lower case letters designate residuals from a linear regressions over month-year$\times$city and day-of-the-week$\times$city fixed effects. Selection on these effects is not appropriate: we want to maintain the conditional exclusion restriction and use identifying variations which do not come from mere seasonality. We therefore study all variables after partialling-out seasonal fixed effects, which boils down to a Frisch-Waugh transformation. This is the first step of the treatment required for panel data in such a setting (*Belloni et al., 2016*), and is similar to the IV-lasso implementation in (*Gilchrist and Sands, 2016*).

We consider the following selection equation

$$p_{kct} = aw_{ct}\eta + x_{ct}b + \epsilon_{ct} \tag{3}$$

where k indexes pollutants, and aw_{ct} is a high dimensional set of instruments built from the altitude weather variables;[21] and η is a high dimensional vector to be estimated.

We formulate the assumption that η is in fact at least approximately sparse, i.e. that only a ”small” number of dimensions of this vector is non negligible. That is, only some of the introduced instruments variables do have a non negligible impact on pollutants' concentration. With this assumption, we can avoid hand-picking instruments and run a first stage Lasso regression whose output are the non zero coefficients, one for each pollutant. In this step, there is no selection on ground-level weather controls, there are unpenalized: in the end, none of the dimension of b are at zero. Again, this is done to maintain the exclusion restrictions, for reasons similar to these applying to seasonal fixed effects.[22]

With the pooled selected instruments (one set of instruments has been selected for each pollutant), we compute a two stage least squares with multiple pollutants, taking into account our additional controls (again the same ground-level weather controls, remember that seasonal FE have been partialled out). In this post-lasso IV, the first stage writes as follows

$$p_{kct} = aws_{ct}\mu + x_{ct}c + u_{ct} \tag{4}$$

where aws are altitude weather variables selected by lasso in a first step among aw. The second stage is as follows

$$r_{act} = \sum_{k=1}^{5} \hat{p}_{ct}\delta_k + x_{ct}\beta + \nu_{ct} \tag{5}$$

[21]For inverse PBL height, its lag, and thermal inversion presence, we build many possible functions of the data at the city-date (c, t) level: averages over 24hours; averages over day hours (when pollution is mostly emitted); over 6 time windows of 4 hours and specific to the city (allowing for various predictive patterns per city). We add thermal inversions strength averaged at 6 moments of the day; evaporation; terrestrial heat latent flux; and in 14 altitude layers, humidity, zonal wind, meridian wind, wind strength. The full set of instruments comprise 322 variables: 76 related to thermal inversions, 66 related to inverse PBL height and 66 to its lag, 85 related to winds.

[22]However, seasonal fixed effects are somewhat too numerous to be forced into the model in the same way. Theoretical properties after selection from a high dimensional set of instruments, conditional on covariates, assume the later to be low dimensional.

where each $\hat{p}_{kct}$ is derived from the post-lasso first stage linear regression, and r and x are emergencies admissions (or mortality rate) and ground-level weather controls.

We employ lasso selection relying on the *hdm* package (*Chernozhukov et al., 2016*). The penalty parameter in chosen *rigorously* in the sense of (*Belloni et al., 2012*) ("rigorous lasso"), to allow for principled inference. After lasso selection, we run a two-stage least squares on the pooled set of instruments. All our regressions standard errors are clustered at the city level.

4 Results

4.1 Causal impact of air pollution

We first consider air pollution as a whole. In this section, the inverse of planet boundary layer is used to instrument an aggregate pollution index as well as pollutants taken separately.[23] We present reduced form results, causal estimates and finally explore delayed effects over a few days.

We start by showing reduced-form estimates in Table 5, relating emergencies admissions and mortality rate to inverse of the PBL height. Panel (C) corresponds to our baseline first stage, with both inverse PBL height and its lag as instruments. Panel (A) and (B) show the results when considering one or the other instrument. Conditional on weather and seasonal patterns, a lower boundary layer involves significantly more emergency admissions for respiratory diseases and a higher mortality rate on the following day. The timing is rather different for cardiovascular diseases: the contemporary PBL height affects the emergency admissions more than its lag. This table suggests that the health reaction to air pollution might be different between respiratory and cardiovascular diseases. As a falsification test, we add digestive diseases which are the other most common admissions in emergency. As expected, we find no effects of the inverse of boundary layer on these pathologies.

[23]For now, we let ozone aside. To study O3, we will need to control at least for the other pollutants involved in equilibrium with O3, that are anti-correlated. O3 is nevertheless considered when computing pollution indexes that aggregates all pollutants of our sample. We come back to O3 in the next section.

Table 5: Inverse of planetary boundary layer height, emergency admissions and mortality rate

	Dependent variable, per 100 000 inhabitants: *Emergency admissions*			*Mortality rate*
	Respiratory diseases	Cardiovascular diseases	Digestives diseases	
One instrument (A)	(1)	(2)	(3)	(4)
Lag of inverse of PBL Height	16.984***	1.980	−7.419	15.673***
	(5.372)	(3.221)	(7.775)	(5.975)
Observations	21,459	21,459	21,459	21,459
One instrument (B)	(1)	(2)	(3)	(4)
Inverse of PBL Height	5.686	7.743**	1.931	12.841**
	(5.806)	(3.766)	(7.181)	(5.057)
Observations	21,468	21,468	21,468	21,468
Two instruments (C)	(1)	(2)	(3)	(4)
Inverse of PBL Height	−2.010	7.658	6.107	6.704
	(6.259)	(5.337)	(6.773)	(6.361)
Lag of inverse of PBL Height	17.720***	−0.825	−9.656	13.218*
	(5.616)	(4.699)	(7.752)	(7.147)
Observations	21,459	21,459	21,459	21,459

All regressions includes month-year and day-of-the-week fixed effects, interacted with city fixed effects; and weather controls. Standard errors are clustered at the city level. Significance: *p<0.1; **p<0.05; ***p<0.01

Table 6 presents the IV results for respiratory emergency admissions, derived from Equations 1 and 2. Each coefficient is from a separate regression, the dependent variable being either emergency admissions for respiratory, cardiovascular, digestive diseases or the mortality rate. In our baseline specification in column (1), we instrument the pollution index. Although for now, we do not want to attribute these health effects to a particular pollutant, we provide the results when considering separately distinct pollutants instead in columns (2-6). The later estimates may be compared to the existing literature, which as here do not control for the presence of other pollutants. An increase by half a standard deviation of the air pollution index (0.65, which corresponds to the average difference in levels of this air pollution index between a Sunday and a Wednesday) leads to emergency admissions for respiratory diseases higher by 0.03 admissions per 100 000 inhabitants, that is an increase by about 2%. Similarly, the admissions for cardiovascular diseases are higher by 0.6% and mortality rate by 1.5 %. Again, we perform a falsification test of the effect of air pollution on digestive diseases which is as expected insignificant. For each specific pollutant, concentrations are expressed in $\mu g/m^{-3}$. Quantitatively, + 10 $\mu g/m^{-3}$ in PM2.5 (about a standard deviation) leads to + 3% more respiratory admissions, + 2% more admissions for cardiovascular diseases and a mortality rate higher by +3%.[24] For CO, + 200 $\mu g/m^{-3}$ (about a standard deviation) leads to +4% more respiratory admissions and +4% of the mortality rate. However, with this strategy, we cannot rule out that a single pollutant drives all the results. We come back to this issue in the next section.

Finally, we conduct a battery of robustness checks. In the second and third columns of Table 7, we show the point estimates when using alternative instruments either following (*Jans et al., 2018*), using their definition of thermal inversions, or considering PBL height and its lag averaged at 6 moments of the day (12 instruments) instead of our baseline with PBL height and its lag averaged over the full day (2 instruments). The results are very close, although of higher magnitude in column (2). Then, we test carefully the sensitivity of the results to the specification of weather controls. The following columns (4-6) address this concern: we modify the degrees of polynomials or withdraw the weather variables (temperature, rainfall, wind strength,

[24] We may compare the later estimate to that from (*Schwartz et al., 2016*), which find, for Boston over 2000-2009, that for an increase by about $6\mu g/m^{-3}$ of PM2.5 leads to an increase by 0.9% of daily deaths. Our estimate is of similar magnitude, although slightly higher.

Table 6: Causal effect of air pollution on health. Monopollutant IV models.

	Instrumented variable:					
	(1) Pollutant index	(2) PM2.5	(3) PM10	(4) CO	(5) NO2	(6) SO2
Dependent variable:						
Respiratory diseases	0.041***	0.005***	0.005***	0.0003**	0.007*	0.306
[1.53]	(0.015)	(0.002)	(0.002)	(0.0001)	(0.004)	(0.241)
Cardiovascular diseases	0.014**	0.003**	0.001	0.0003	0.006**	0.143
[1.59]	(0.006)	(0.001)	(0.001)	(0.0002)	(0.002)	(0.131)
Mortality rate	0.050***	0.006***	0.005***	0.0004***	0.014***	0.244
[2.10]	(0.016)	(0.002)	(0.002)	(0.0001)	(0.004)	(0.245)
Digestive diseases	−0.009	−0.003	−0.002	0.00001	−0.004	−0.106
[2.81]	(0.023)	(0.003)	(0.002)	(0.0002)	(0.005)	(0.184)
Observations	21,399	16,046	14,178	14,064	14,824	14,775
Concentration ($\mu g/m^{-3}$)		[16.5]	[25.8]	[407.7]	[37.4]	[1.00]

Each coefficient is from a separate regression. Regressions with the same endogenous variables share the same number of observations. Instruments are inverse PBL height and its lag. All regressions includes month-year and day-of-the-week fixed effects, interacted with city fixed effects. All pollutant concentrations are in $\mu g/m^{-3}$. The average level of the dependent variables and of the pollutants' concentrations are given within brackets. Standard errors are clustered at the city level. Significance: *p<0.1; **p<0.05; ***p<0.01

humidity and sunlight). Neither the point estimate nor the significance are altered. Finally, we consider different specifications for fixed effects. Our baseline set of fixed effects is rather restrictive and capture already a large amount of seasonal variations (in particular seasonal trend at the month-year level, specific to each cities). Column (8) consider a loose set of temporal fixed effects, not specific to the city, but taking into account aggregate shocks at the day, season (3-month periods) and year level. Respiratory diseases are somewhat less significant p-value of 0.051 instead of 0.006 in our baseline), but the results remain very similar for cardiovascular diseases and mortality. If we consider that monthly seasonal patterns are the same from one year to the next, as in column (10), the resultats are very similar. However, taking the specification further and introducing week$\times$ city fixed effects alter the significance of our results (column 11) for respiratory diseases. All in all, our baseline set of fixed effects seems a reasonable choice. The most robust result which almost never changes across any robustness tests is relative to the mortality rate.

In this first part, our results clearly evidence a detrimental short-term effect of air pollution on respiratory health. However, we cannot properly distinguish which pollutant has the strongest effect and even whether all pollutants do have an impact on health or rather that the effect is borne by a few ones, which covary with all others.

Table 7: Air pollution and health: robustness checks. Monopollutant IV models.

	Health outcome on instrumented pollutant index										
	(Baseline)	(Alternative instruments)		(Other Ground-level weather controls)			(Other seasonal fixed effects)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent											
Respiratory	0.041***	0.090***	0.037**	0.034***	0.039**	0.048***	0.209***	0.133*	0.175***	0.044**	0.006
	(0.015)	(0.029)	(0.014)	(0.011)	(0.016)	(0.015)	(0.072)	(0.068)	(0.045)	(0.018)	(0.017)
Mortality	0.050***	0.065***	0.036***	0.046***	0.051***	0.054***	0.101***	0.104***	0.106***	0.051***	0.041***
	(0.016)	(0.021)	(0.014)	(0.012)	(0.018)	(0.017)	(0.020)	(0.021)	(0.020)	(0.015)	(0.014)
Cardiovascular	0.014**	0.034**	0.020***	0.011*	0.014**	0.015**	0.045***	0.043***	0.044***	0.020**	0.017*
	(0.006)	(0.014)	(0.008)	(0.006)	(0.007)	(0.007)	(0.010)	(0.009)	(0.009)	(0.009)	(0.010)
Polynom order											
Temperature	2	2	2	1	3	2	2	2	2	2	2
Rainfall	2	2	2	1	3	2	2	2	2	2	2
Wind strength	2	2	2	1	3	2	2	2	2	2	2
Humidity	1	1	1	0	1	2	1	1	1	1	1
Sunlight	1	1	1	0	1	2	1	1	1	1	1
Snow dummies	1	1	1	0	1	1	1	1	1	1	1
Seasonal FE	Day-of-the-week (Day) × City Month-year × City						City	City Day Season Year	Day Season × City Year	Day × City Month × City Year	Day × City Week × City Year

Notes: Alternative instruments are (*Jans et al., 2018*) thermal inversions instruments in column (2) and PBL heigth and its lags at six moments of the day in column (3). Standard errors are clustered at the city-level. *p<0.1; **p<0.05; ***p<0.01

Table 8: Air pollutants and respiratory health. Post-lasso selection IVs.

	Dependent variable: *Emergency admissions for respiratory diseases per 100 000 inhabitants*												
N pollutants	1	2			3					4		5	
CO	0.0002*	0.001***			0.001*	0.0004**		0.001**		0.001**		0.0002	0.001*
	(0.0001)	(0.0001)			(0.0003)	(0.0002)		(0.0002)		(0.0003)		(0.0003)	(0.0004)
	[14,064]												
O3	0.008***	0.007***			0.006***	0.006***	0.003*			0.006***	0.004*	0.005**	0.007***
	(0.001)	(0.001)			(0.001)	(0.001)	(0.002)			(0.002)	(0.002)	(0.002)	(0.002)
	[15,917]												
SO2	0.062***		0.046***			0.058***	0.062**		0.050***	0.062***	0.037*	0.054**	0.085***
	(0.006)		(0.016)			(0.018)	(0.027)		(0.015)	(0.018)	(0.022)	(0.024)	(0.026)
	[14,775]												
PM10	0.003***		0.002**					−0.007	0.005		0.005**	0.005*	
	(0.001)		(0.001)					(0.005)	(0.010)		(0.002)	(0.003)	
	[14,178]												
PM2.5	0.002*			0.001				0.008	−0.004				0.0001
	(0.001)			(0.001)				(0.006)	(0.011)				(0.002)
	[16,046]												
NO2	0.004**			0.003**	0.003		0.004***			0.001	0.00004	−0.001	0.001
	(0.002)			(0.001)	(0.002)		(0.001)			(0.002)	(0.002)	(0.002)	(0.003)
	[14,824]												
Observations	[in bracket]	11,533	9,851	11,714	9,697	8,902	9,460	7,485	8,337	7,677	6,608	5,195	6,111

In column (1), each coefficient corresponds to a separate regression with one pollutant. Other columns correspond to a regression. Before all regressions, we partial out fixed effects. All variables are first regressed on month-year and day-of-the-week fixed effects both interacted with city fixed effects and then replaced by the corresponding residuals. A first step of per-pollutant lasso selection is performed, conditional on weather variables which are forced into the model (no selection), selected instruments are then pooled and enter a regular IV estimations. The sample reduce in size when increasing the number of pollutant due to missing values. Standard errors are clustered at the city level. Significance: *p<0.1; **p<0.05; ***p<0.01

4.2 Disentangling the impact of distinct pollutants

In Table 8 and 9 we present our main set of results, derived from Equations 4 and 5. We start with monopollutant models where instruments are selected with a Lasso step, as described in section 3.3, so we can compare how applying this new method modify the results obtained before in a classical IV framework. The first column presents the results of six separate regressions, each with one pollutant, and the other column presents the results when successively adding more pollutants. The final and complete specification comprise five distinct pollutants. We can therefore consider these results as giving the separate impact of pollutants, once controlled for the other main pollutants.[25] Although we mostly comment the specification with 5 pollutants, we note that the sample decreases in size each time that we add a pollutant (for an observation to enter the regression, all the considered pollutants should be observed in a given urban area at a given date, a condition which significantly alters the sample size). That is why we should not disregard the results with fewer pollutants, which are based on more observations.

For respiratory diseases (Table 8), we find compelling evidence of the detrimental and pollutant-specific effect of three pollutants: ozone (O3), sulfur dioxide (SO2) and carbon monoxide (CO). These three pollutants affect significantly emergency admissions in the three pollutants model (column 5 of Table 8) and in a 5 pollutants models (last column of Table 8). The impact of carbon monoxide might however been still confounded with the impact of PM10: comparing the last two columns of Table 8, introducing PM10 instead of PM2.5 alter the significance of CO, to the benefit of PM10. The most robust findings is the detrimental effect of ozone O3 and sulfur dioxide. O3 is found to have an independent effect from other pollutants, which is stable when we successively add more pollutants to the equation: a standard deviation of O3 ($24\mu g/m^{-3}$) causes between 5 (6th column) to 12 % (1rst column) more emergency admissions for respiratory diseases. A standard deviation of SO2 ($1.5\mu g/m^{-3}$) causes between 5 (last column) to 9 % (1rst column) more emergency admissions for respiratory diseases.

[25] We never consider both types of particulate matters together as we did not suceed in disentangling their separate effects, even in regressions with only these two pollutants.

Table 9: Air pollutants and health. Cardiovascular diseases and Mortality rate. Post-lasso selection IVs.

	Dependent variable, per 100 000 inhabitants:							
	Mortality rate				*Emergency admissions for cardiovascular diseases*			
N pollutants	1	2	3	5	1	2	3	5
PM2.5	0.004***	0.004***	0.005**	0.004**	0.0002		−0.002	0.0004
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)		(0.001)	(0.002)
	[16,046]				[16,046]			
PM10	0.004***				−0.0001			
	(0.001)				(0.001)			
	[14,178]				[14,178]			
CO	0.0002*		0.00002	−0.0001	0.0002*	0.0003**	0.0004**	0.0004**
	(0.0001)		(0.0002)	(0.0005)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
	[14,064]				[14,064]			
O3	−0.002			−0.003	−0.001			0.0004
	(0.001)			(0.003)	(0.001)			(0.003)
	[15,917]				[15,917]			
SO2	0.04*			0.027	−0.006			−0.006
	(0.023)			(0.030)	(0.016)			(0.023)
	[14,775]				[14,775]			
NO2	0.006***	0.004	0.003	0.001	0.002	−0.0001	−0.0002	−0.002
	(0.002)	(0.002)	(0.003)	(0.004)	(0.001)	(0.002)	(0.002)	(0.002)
	[14,824]				[14,824]			
Observations	[in bracket]	11,714	8,322	6,111	[in bracket]	10,566	8,322	6,111

In column labeled 1, each coefficient corresponds to a separate regression with one pollutant. Other columns correspond to a multipollutants regression. Before all regressions, we partial out fixed effects. All variables are first regressed on month-year and day-of-the-week fixed effects both interacted with city fixed effects and then replaced by the corresponding residuals. A first step of per-pollutant lasso selection is performed, conditional on weather variables which are forced into the model (no selection), selected instruments are then pooled and enter a regular IV estimations. The sample reduce in size when increasing the number of pollutant due to missing values. Standard errors are clustered at the city level. Significance: *p<0.1; **p<0.05; ***p<0.01

For mortality rate and cardiovascular diseases, we find strong evidence of the negative effect of respectively particulate matters PM2.5 and carbon monoxide CO, eitheir in a monopollutant model or in the 5 pollutants models (Table 9). Quantitatively, a standard deviation of PM2.5 ($11\mu g/m^{-3}$) causes an increase between 2 and 4 % of the mortality rate. A standard deviation of CO ($216\mu g/m^{-3}$) causes between 3 to 6 % more emergency admissions for cardiovascular diseases.[26]

(Preliminary) We conduct two types of heterogeneity analysis: along the age ladder and by subgroups of diseases. In the following, we focus for respiratory diseases on models with

[26]For digestive diseases, we perform similar falsification test which are available in Appendix, Table 14, and find no effects.

ozone, sulfur dioxide and carbon monoxide and for cardiovascular diseases and mortality rate on monopollutant models, a choice which preserves the sample size. We first explore the effects along the age ladder in Table 10, where both hands of the age distribution stand out in particular the youngest. Note that admissions are much more frequent as well at both extremes. A level of ozone higher by one standard deviation ($24\mu g/m^{-3}$) leads to 12 more emergency admissions per 100 000 newborns[27] for respiratory diseases, about 2.6 more emergency admissions per 100 000 infants,[28] and 0.9 more emergency admissions per 100 000 children aged less than 4.

(Preliminary) Second, we derive results by group of diseases in Table 11: acute upper respiratory system diseases (e.g. pharyngitis), influenza, bronchitis and chronic diseases (e.g. asthma, or COPD, Chronic obstructive pulmonary disease) and abnormalities of breathing.

(Preliminary) Finally, we explore delayed effects. Throughout the whole analysis, we have assumed that the impact of air pollution on short-term health indicators was exclusively contemporaneous. In this last part, we check whether we find lagged effects when introducing leads and lags in the IV, with both contemporaneous and lagged instruments. Table 15 in Appendix shows that it is probably not the case. When introducing two lags in the baseline models (columns (1)), the effect still appear as contemporaneous. When introducing a lead on top of two lags (columns (2)), results are somewhat less significant. For most of our results, these regressions suggest that the (short-term) effect is mostly contemporaneous. However, it is not as clearcut for ozone on respiratory diseases, as when introducing nitrogen oxides (which is in equilibrium with ozone), ozone appears with a two-days lagged effects on emergency admissions for respiratory diseases, but not in all regressions.

[27] Aged less than 28 days.

[28] Aged from 29 days to one year.

Table 10: Air pollutants and health: sensitivity along the age ladder. Post-lasso selection IVs.

	Emergency admissions for respiratory diseases per 100 000 inhabitants (of the age group)									
	(newborns)	(Infant)	(≤ 4)	(5-14)	(15-59)	(60-64)	(65-69)	(70-74)	(75-79)	(≥ 80)
O3	0.510**	0.112***	0.037***	0.002	0.0002	0.006	−0.002	−0.004	−0.002	0.025*
	(0.232)	(0.034)	(0.010)	(0.002)	(0.001)	(0.005)	(0.002)	(0.013)	(0.012)	(0.014)
CO	0.053***	0.005	0.003***	−0.00003	−0.00005	0.001	0.0002	−0.001	−0.0001	0.003
	(0.013)	(0.004)	(0.001)	(0.0002)	(0.0001)	(0.001)	(0.0005)	(0.001)	(0.002)	(0.002)
SO2	−0.846	2.085***	0.265**	0.020	0.011	−0.002	−0.082	−0.066	−0.018	0.281*
	(2.682)	(0.483)	(0.133)	(0.030)	(0.011)	(0.069)	(0.074)	(0.116)	(0.187)	(0.146)
Mean dep. var.	21.78	14.49	5.46	0.50	0.49	1.16	1.43	2.26	3.45	7.23
Observations	8,902	8,902	8,902	8,902	8,902	8,902	8,902	8,902	8,902	8,902

	Emergency admissions for cardiovascular diseases per 100 000 inhabitants (of the age group)									
	(newborns)	(Infant)	(≤ 4)	(5-14)	(15-59)	(60-64)	(65-69)	(70-74)	(75-79)	(≥ 80)
CO	0.003	−0.0002	−0.00004	0.00000	0.00005	0.0002	−0.001	0.001	0.0003	0.003*
	(0.002)	(0.0002)	(0.0001)	(0.00004)	(0.0001)	(0.0003)	(0.0004)	(0.001)	(0.001)	(0.001)
Observations	14,064	14,064	14,064	14,064	14,064	14,064	14,064	14,064	14,064	14,064

	Mortality rate			
	(≤ 4)	($4 \leq . < 65$)	($65 - 74$)	(≥ 75)
PM2.5	−0.0004	0.001**	0.014***	0.034***
	(0.001)	(0.0004)	(0.004)	(0.011)
Observations	16,046	16,046	16,046	16,046

Before all regressions, we partial out fixed effects. All variables are first regressed on month-year and day-of-the-week fixed effects both interacted with city fixed effects and then replaced by the corresponding residuals. A first step of per-pollutant lasso selection is performed, conditional on weather variables which are forced into the model (no selection), selected instruments are then pooled and enter a regular IV estimations. The sample reduces in size when increasing the number of pollutant due to missing values. Standard errors are clustered at the city level. Significance: *p<0.1; **p<0.05; ***p<0.01

Table 11: Air pollutants and respiratory pathologies. Post-lasso selection IVs

	Dependent variable: Emergency admissions for respiratory diseases, by pathology (ICD-10 code)				
	Acute, upper resp.	Influenza	Bronchio.	Chronic	Abnormalities of Breathing
	(J00-J06)	(J09-J18)	(J20-J22)	(J40-J99)	(R06)
O3	-0.0001	0.0005	0.004^{***}	0.001^{*}	0.0003^{**}
	(0.0004)	(0.001)	(0.001)	(0.001)	(0.0002)
CO	0.0001^{*}	0.00003	0.0003^{***}	0.0001	0.00001
	(0.00003)	(0.0001)	(0.0001)	(0.0001)	(0.00002)
SO2	-0.001	0.021^{**}	0.023^{**}	0.012	0.003
	(0.005)	(0.010)	(0.011)	(0.019)	(0.004)
Mean dep. var.	0.079	0.435	0.861	0.635	0.055
Observations	8,902	8,902	8,902	8,902	8,902

Before all regressions, we partial out fixed effects. All variables are first regressed on month-year and day-of-the-week fixed effects both interacted with city fixed effects and then replaced by the corresponding residuals. A first step of per-pollutant lasso selection is performed, conditional on weather variables which are forced into the model (no selection), selected instruments are then pooled and enter a regular IV estimations. Standard errors are clustered at the city level. Significance: *p<0.1; **p<0.05; ***p<0.01

5 Conclusion

This paper has shown how distinct pollutants have strong and independent effects on short-term respiratory health of urban population. We develop a twin strategy, showing first how air pollution is causally linked to daily emergency admissions and mortality rates and second how optimally selecting many more instruments allowed to disentangle the effects of several pollutants. To our knowledge, we are the first to provide causal evidence on the separate effects of ozone, carbon monoxide and sulfur dioxide on respiratory diseases, jointly and independently, in the real urban environment, and controlling for the other pollutants. Moreover, we find a significant impact of carbon monoxide on cardiovascular diseases as well as of particulate matters on the mortality rate, while controlling for the other pollutants in presence. In addition, we show how high dimensional data from a general climate model can be leveraged to provide a large set of instruments which prove very insightful for clean evidence of ambient pollution levels on health. Our estimates could be considered for the production of a short-term pollution index reflecting the joint and independent impact of several pollutants.

Our results point out to large effects of relatively small amounts to ozone, sulfur dioxide, carbon monoxide and particulate matters, borne in priority by children and elderly. While European norms have improved air quality as e.g. carbon monoxide is concerned, ozone concentrations are not at all decreasing in modern European cities.

References

Anderson, 2015. Anderson, M. L. (2015). As the wind blows: The effects of long-term exposure to air pollution on mortality.

Arceo et al., 2016. 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.

Belloni et al., 2012. Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Belloni et al., 2014. Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50.

Belloni et al., 2016. Belloni, A., Chernozhukov, V., Hansen, C., and Kozbur, D. (2016). Inference in high-dimensional panel models with an application to gun control. *Journal of Business & Economic Statistics*, 34(4):590–605.

Chay and Greenstone, 2003. Chay, K. Y. and Greenstone, M. (2003). The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *The quarterly journal of economics*, 118(3):1121–1167.

Chen et al., 2018. Chen, S., Oliva, P., and Zhang, P. (2018). Air pollution and mental health: Evidence from china. Technical report, National Bureau of Economic Research.

Chernozhukov et al., 2015. Chernozhukov, V., Hansen, C., and Spindler, M. (2015). Post-selection and post-regularization inference in linear models with many controls and instruments. *American Economic Review*, 105(5):486–90.

Chernozhukov et al., 2016. Chernozhukov, V., Hansen, C., and Spindler, M. (2016). hdm: High-dimensional metrics. *arXiv preprint arXiv:1608.00354*.

Currie et al., 2011. Currie, J., Heep, S., and Neidell, M. (2011). Quasi-experimental approaches to evaluating the impact of air pollution on childrens health. *Health affairs (Project Hope)*, 30(12):2391.

Currie and Neidell, 2005. Currie, J. and Neidell, M. (2005). Air pollution and infant health: what can we learn from california's recent experience? *The Quarterly Journal of Economics*, 120(3):1003–1030.

Deryugina et al., 2016. Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., and Reif, J. (2016). The mortality and medical costs of air pollution: Evidence from changes in wind direction. Technical report, National Bureau of Economic Research.

Deschênes and Greenstone, 2007. Deschênes, O. and Greenstone, M. (2007). The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *American Economic Review*, 97(1):354–385.

Dufresne et al., 2013. Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J.-P., Eth, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N. (2013). Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate Dynamics*, 40:2123–2165.

Gilchrist and Sands, 2016. Gilchrist, D. S. and Sands, E. G. (2016). Something to talk about: Social spillovers in movie consumption. *Journal of Political Economy*, 124(5):1339–1382.

Hourdin et al., 2006. Hourdin, F., Musat, I., Bony, S., Braconnot, P., Codron, F., Dufresne, J.-L., Fairhead, L., Filiberti, M.-A., Friedlingstein, P., Grandpeix, J.-Y., et al. (2006). The lmdz4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection. *Climate Dynamics*, 27(7-8):787–813.

Jans et al., 2018. Jans, J., Johansson, P., and Nilsson, P. (2018). Economic status, air quality, and child health: Evidence from inversion episodes.

Josse and Husson, 2012. Josse, J. and Husson, F. (2012). Handling missing values in exploratory multivariate data analysis methods. *Journal de la Société Française de Statistique*, 153(2):79–99.

Moretti and Neidell, 2011. Moretti, E. and Neidell, M. (2011). Pollution, health, and avoidance behavior evidence from the ports of los angeles. *Journal of human Resources*, 46(1):154–175.

Munir et al., 2012. Munir, S., Chen, H., and Ropkins, K. (2012). Modelling the impact of road traffic on ground level ozone concentration using a quantile regression approach. *Atmospheric environment*, 60:283–291.

Schlenker and Walker, 2015. Schlenker, W. and Walker, W. R. (2015). Airports, air pollution, and contemporaneous health. *The Review of Economic Studies*, 83(2):768–809.

Schwartz et al., 2016. Schwartz, J., Bind, M.-A., and Koutrakis, P. (2016). Estimating causal effects of local air pollution on daily deaths: effect of low levels. *Environmental health perspectives*, 125(1):23–29.

Stull, 2016. Stull, R. (2016). *Practical Meteorology: an algebra based survey of atmospheric science*. BC Campus.

Tibshirani, 1996. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

6 Additional Figures

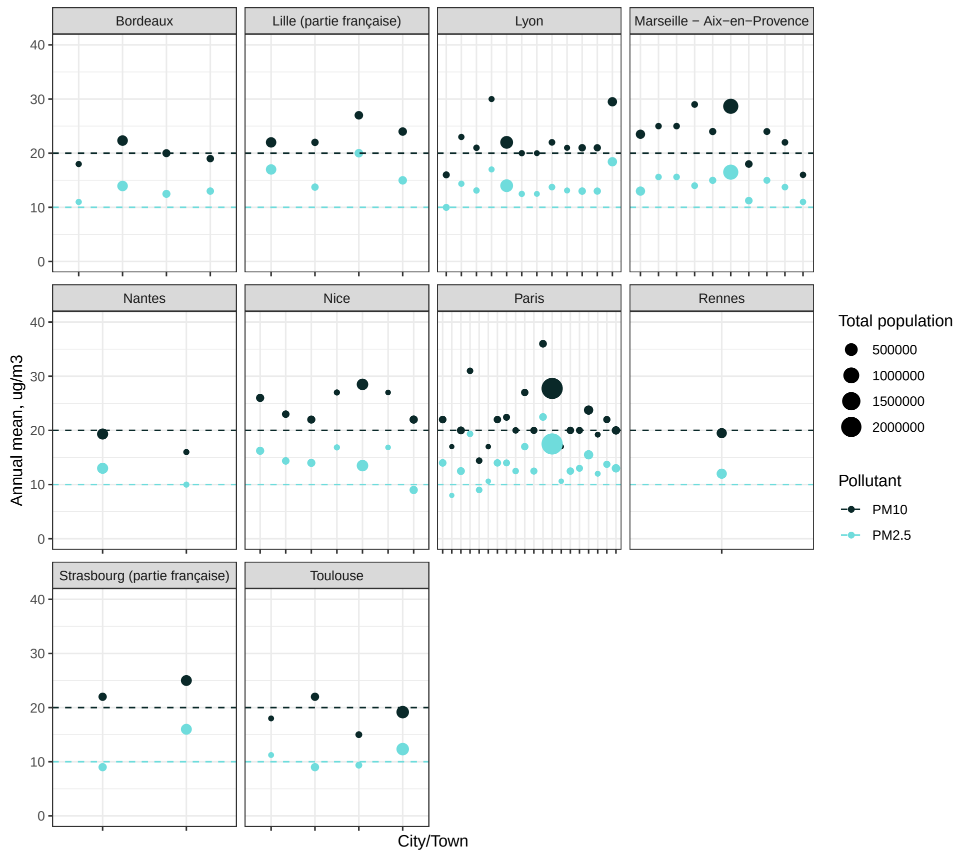

Figure 3: Annual mean of particulate matters in cities in urban areas reporting data to WHO. Dotted line represents WHO guidelines. Source: WHO Ambient (outdoor) air pollution database 2016, measurements in 2014, Census 2013

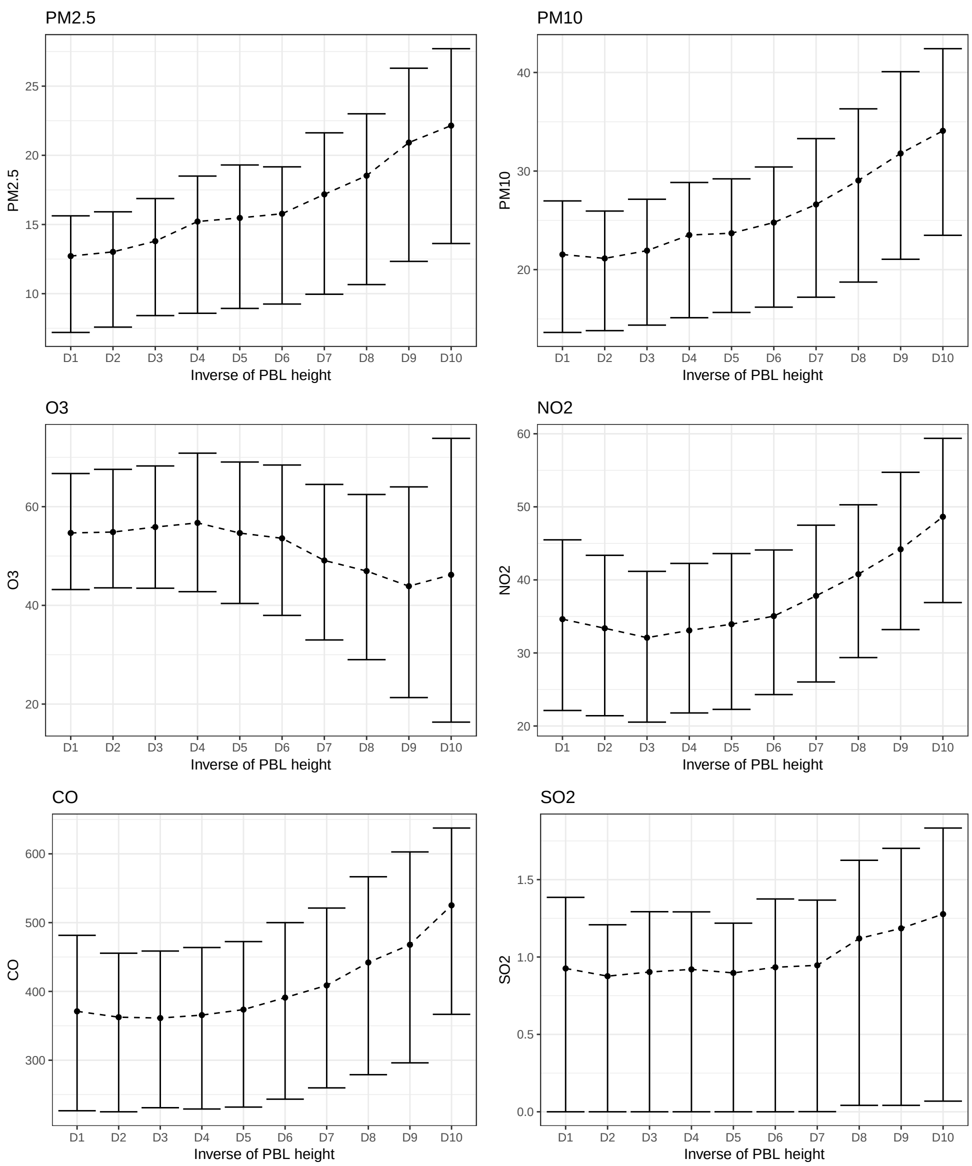

Figure 4: Pollutants concentration (non standardized) and inverse boundary layer height (raw correlation).

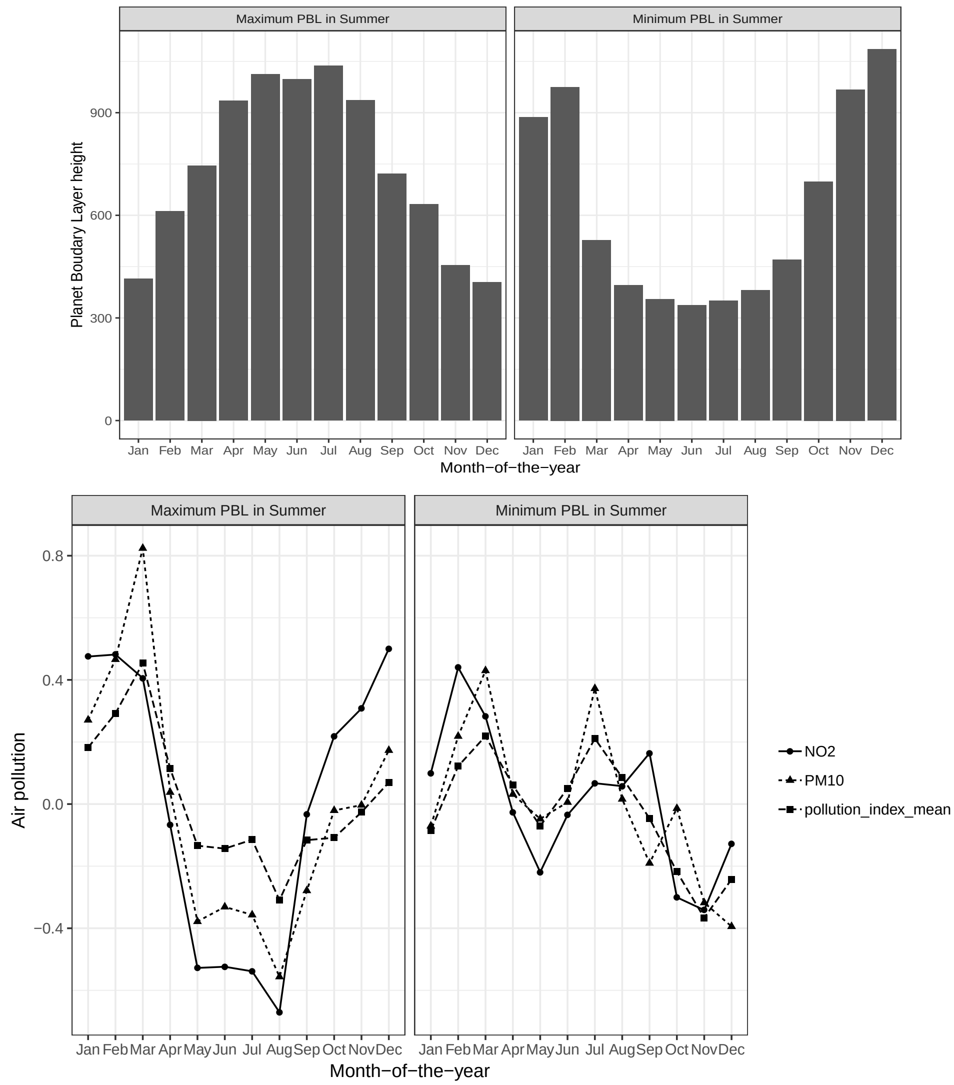

Figure 5: Most cities have a PBL seasonality with maximal height in summer (Paris, Toulouse, Bordeaux, Strasbourg, Lyon, Rennes, Nantes). Two cities have their maximal PBL height in winter (Marseille, Nice). Finally Lille PBL has a flat profile.

7 Additional Tables

Table 12: Monopollutant models with Lasso instruments selection.

	Emergency admissions for respiratory diseases per 100 000 inhabitants		
Instruments	(a',b',c,d)	(a,b,c,d)	(a",b",c,d)
N instrumental variables	330	150	1,170
PM2.5	0.002*	0.001	0.001
	(0.001)	(0.001)	(0.001)
	[30]	[25]	[39]
PM10	0.003***	0.002***	0.002**
	(0.001)	(0.001)	(0.001)
	[35]	[21]	[35]
CO	0.0002*	0.0003	0.0003
	(0.0001)	(0.0002)	(0.0002)
	[24]	[10]	[18]
O3	0.002**	0.005***	0.005***
	(0.001)	(0.001)	(0.001)
	[38]	[20]	[41]
SO2	0.062***	0.115***	0.100***
	(0.006)	(0.041)	(0.031)
	[14]	[5]	[8]
NO2	0.004**	0.004**	0.004**
	(0.002)	(0.002)	(0.002)
	[31]	[16]	[25]

Note: Each coefficient is derived from a separate IV regression on a subset of selected instruments. Standard errors are clustered at the city level. The set of available instruments comprises (a) thermal inversions, (a') thermal inversions interacted with city, (a") thermal inversions interacted with (humidity, pressure, TKE, wind strength), (b) PBL heights, (b') PBL heights interacted with the city, (b") PBL heights interacted with (humidity, pressure, TKE, wind strength), (c) zonal (W-E) and meridional(N-S) winds in several layers, (d) other variables: altitude humidity, evaporation, latent heat flux (continental) .

Table 13: From OLS to IV

	Emergency admissions for respiratory diseases					Mortality				
	(OLS)	(OLS)	(IV)	(IV)	(IV)	(OLS)	(OLS)	(IV)	(IV)	(IV)
Pollution index	0.195***	−0.003	0.041***	0.030*	0.037**	0.095**	0.012**	0.050***	0.028*	0.036***
	(0.028)	(0.008)	(0.015)	(0.017)	(0.014)	(0.041)	(0.005)	(0.016)	(0.014)	(0.014)
Instruments	.	.	(1)	(2)	(3)	.	.	(1)	(2)	(3)
Day x City	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Month-year x City	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Weather controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Observations	21,841	21,400	21,399	21,400	21,399	21,841	21,400	21,399	21,400	21,399

Note: Pollution index is derived from a PCA analysis (the first component). Instruments (1) are average inverse of PBL height and its lag; (2) are average inverse of PBL height at 6 moments of the day; (3) are average inverse of PBL height at 6 moments of the day and their lag.Standard errors are clustered at the city level. *p<0.1; **p<0.05; ***p<0.01

Table 14: Falsification test: air pollution and digestive diseases.

	Dependent variable: Emergency admissions for digestive diseases per 100 000 inhabitants					
	(1)	(2)	(3)	(4)	(5)	(6)
PM2.5	−0.002					
	(0.001)					
PM10		−0.001				
		(0.001)				
CO			−0.00004			
			(0.0001)			
O3				−0.001		
				(0.001)		
SO2					−0.012	
					(0.013)	
NO2						−0.002
						(0.002)
Observations	16,046	14,178	14,064	15,917	14,775	14,824

Before all regressions, we partial out fixed effects. All variables are first regressed on month-year and day-of-the-week fixed effects both interacted with city fixed effects and then replaced by the corresponding residuals. A first step of per-pollutant lasso selection is performed, conditional on weather variables which are forced into the model (no selection), selected instruments are then pooled and enter a regular IV estimations. The sample reduce in size when increasing the number of pollutant due to missing values. Significance: *p<0.1; **p<0.05; ***p<0.01

Table 15: Lagged effects. Post-lasso selection IVs.

	Respiratory diseases					Cardiovascular diseases		Mortality rate	
	(1)	(2)	(3)	(4)		(1)	(2)	(1)	(2)
O3 (t+1)		0.006*		0.006	CO (t-1)		−0.00004		
		(0.003)		(0.004)			(0.0001)		
O3 (t)	0.007***	0.001	0.006**	0.002	CO (t)	0.0003**	0.0003		
	(0.002)	(0.005)	(0.002)	(0.004)		(0.0001)	(0.0002)		
O3 (t-1)	−0.001	−0.001	0.0001	−0.002	CO (t-1)	0.00003	0.0001		
	(0.002)	(0.003)	(0.002)	(0.002)		(0.0001)	(0.0001)		
O3 (t-2)	0.001	0.002	0.002	0.003***	CO (t-2)	0.00001	−0.00001		
	(0.001)	(0.002)	(0.002)	(0.001)		(0.00004)	(0.00004)		
CO (t+1)		0.0002		−0.0002	PM2.5 (t+1)				−0.0002
		(0.0004)		(0.001)					(0.002)
CO (t)	0.001***	0.001**	0.001***	0.001***	PM2.5 (t)			0.004**	0.004
	(0.0002)	(0.0003)	(0.0002)	(0.0003)				(0.002)	(0.003)
CO (t-1)	−0.0001	−0.0002	0.00001	−0.0002	PM2.5 (t-1)			0.0004	0.001
	(0.0002)	(0.0002)	(0.0003)	(0.0003)				(0.002)	(0.003)
CO (t-2)	0.00001	0.0001	0.0004*	0.0003	PM2.5 (t-2)			0.001	0.0004
	(0.0002)	(0.0002)	(0.0002)	(0.0003)				(0.001)	(0.001)
SO2 (t+1)		0.006		0.037					
		(0.027)		(0.037)					
SO2 (t)	0.041*	0.032	0.023	0.010					
	(0.023)	(0.033)	(0.023)	(0.030)					
SO2 (t-1)	−0.012	−0.006	0.039	0.016					
	(0.033)	(0.042)	(0.041)	(0.044)					
SO2 (t-2)	−0.021	−0.021	0.002	−0.003					
	(0.022)	(0.029)	(0.017)	(0.025)					
NO2 (t+1)				0.004					
				(0.005)					
NO2 (t)			−0.00004	−0.0002					
			(0.004)	(0.004)					
NO2 (t-1)			−0.001	−0.001					
			(0.003)	(0.002)					
NO2 (t-2)			−0.003	−0.0002					
			(0.002)	(0.002)					
Observations	5,255	4,418	4,069	3,236		11,468	10,659	12,162	10,768

Before all regressions, we partial out fixed effects. All variables are first regressed on month-year and day-of-the-week fixed effects both interacted with city fixed effects and then replaced by the corresponding residuals. Compared to the contemporaneous IV equation, the set of instruments is the same, but is inflated with all instruments' lags (one to five periods), before selection. A first step of per-pollutant lasso selection is performed, conditional on weather variables which are forced into the model (no selection), selected instruments are then pooled and enter a regular IV estimations. Standard errors are clustered at the city-level. *p<0.1; **p<0.05; ***p<0.01