Air pollution and labor supply: Evidence from social security data

Felix Holub* Laura Hospido† Ulrich J. Wagner‡

December 31, 2016

Preliminary and incomplete - please do not quote or cite!

Abstract

We estimate the causal impact of air pollution on the incidence and duration of sickness leaves taken by a representative sample of employees affiliated to the social security system in Spain. Identification derives from day-to-day variation in air pollution concentrations to which the individuals in the sample are exposed in their place of residence. We compute local measures of air quality by interpolating geo-referenced data from almost 900 air quality monitoring stations in all of Spain. These monitoring stations measure and record, at least once per hour, the concentration of various air pollutants that are known to cause harm to human health. We estimate a linear probability model that relates the event of a worker staying at home on a given day in 2009 to the air quality experienced at the place of residence, controlling for confounding factors such as weather, season and individual effects. Our study contributes new evidence on the impact of pollution on worker productivity.

*Department of Economics, University of Mannheim, Germany
†Banco de España, Madrid, Spain
‡Department of Economics, University of Mannheim, Germany

1 Introduction

How does air quality affect human capital? Researchers have gathered a sizable body of empirical evidence on this question, especially concerning the relationship between air quality and human health. In recent years, this literature has grown increasingly sophisticated, relying on ever larger and more detailed datasets, often from administrative sources, covering outcomes relating to infant health (Chay & Greenstone, 2003; Currie & Neidell, 2005; Currie et al., 2009b), outpatient admissions in hospitals and mortality (Moretti & Neidell, 2011; Karlsson et al., 2015; Schlenker & Walker, 2016). At the same time, some recent studies have extended the scope of the analysis beyond health impacts to investigate how air quality impacts on an individuals' productivity at work or in the classroom. This line of research suggests that bad air quality lowers productivity both at the intensive margin – i.e., the performance on the job (Zivin & Neidell, 2012; Lichter et al., 2015) or in the classroom (Ebenstein et al., 2016; Roth, 2016) – and at the extensive margin – i.e., the number of hours worked (Hanna & Oliva, 2015) or spent in school (Currie et al., 2009a).

A fundamental empirical challenge in the estimation of the short-run impact of air pollution on labor supply arises from unobserved economic shocks that shift air pollution and labor demand simultaneously and thus induce bias in the estimated relationship between air pollution and labor supply (Hanna & Oliva, 2015). In this paper, we exploit rich, individual-level panel data from the Spanish social security system to circumvent this problem. Specifically, we investigate the impact of air pollution on sick leaves taken by workers with full-time employment contracts. Because the terms of these contracts are shaped by a highly rigid collective bargaining process, they are unlikely to respond to short-run economic shocks. Furthermore, the daily panel allows us to purge the estimates of possible sorting bias that would arise e.g. if less polluted places attract individuals with weaker health, and to control for unobserved shocks to local labor demand at the municipality-by-month level.

Our econometric approach fits a linear probability model for the event that an individual "calls in sick" on a given day. The model relates this decision to air quality at the place of residence, controlling for weather, individual effects, and a range of time-by-location effects. In addition, we exploit quasi-experimental variation in PM_{10} that is due to Sahara dust advection in order to instrument for local PM_{10} concentrations.

The results imply that a permanent reduction in PM_{10} concentrations by one standard deviation would reduce the sick leave rate by 0.04 percentage points relative to a mean absence rate of 3.13%. This effect is statistically and economically significant. Because social security covers more than 95% of employees in Spain, our estimates are presumably close to the population effect in Spain.[1] For the population of full-time employees, the above percentage effect implies an annual increase in labor supply by 1.9 million worker-days, with an associated productivity gain valued at more than €191 million.

2 Institutional Background

2.1 Temporary disability benefits in Spain

The vast majority of workers in formal employment relationships are entitled to temporary disability benefit. In particular, all affiliates of the social security system are entitled to sick leave benefits provided that they see a doctor affiliated with the health care system for treatment and that they have contributed to social security during a minimum contribution period of 180 days in the five years immediately preceding the illness.[2] The benefit consists of a daily subsidy the amount of which is calculated using the contribution base and percentages applicable to it. As a general rule, the regulatory base is the result of dividing the amount of the contribution base of the worker in the preceding month by the number of days worked in that month. In case of common illness, the benefit is paid from the fourth day of the leave. The benefit payment corresponds to 60% of the regulatory base from day four until day 20 of the sickness spell, and 75% from day 21 onward. During the first three days, no social security benefit is paid, although many employers have schemes in place that cover sick pay during those days. The maximum duration of the benefit is 12 months, renewable for another 6.[3]

Although the payment of benefits is always done by the employer, from the six-

[1] Previous work has estimated labor supply responses to pollution within the confines of a particular city (Hansen & Selte, 2000; Hanna & Oliva, 2015).

[2] No minimum contribution period is required in the case of an accident.

[3] In the case of an accident or occupational disease, social security pays 75% of the regulatory base starting on the day following the beginning of the sick leave. The employer pays the first day of the leave in full.

teenth day, the employer can claim reimbursement of the benefits paid from the social security administration.

In addition, some collective agreements complement the temporary disability benefit. Workers receive less than their wage during the sick leave, but if the agreement complements the amount of the disability payment, the difference to the salary may be small. Some agreements grant matching funds to achieve 100 % of the salary during the temporary disability from day one.

Example A worker earns a monthly base salary of €1,340.54 which amounts to €44.68 per day. He has been sick at home for 22 days and his collective agreement does not complement the temporary disability benefit. During days 1 to 3 of the sick leave, the worker earns €0. During days 4 through 15, the company pays a benefit of 60% of the base salary, i.e.

$$\text{€}44.68 \cdot 60\% \cdot 12 = \text{€}321.73$$

During days 16 through 20, the social security administration pays a benefit of 60%

$$\text{€}44.68 \cdot 60\% \cdot 5 = \text{€}134.05$$

Finally, the benefit paid by the social security administration rises to 75% during days 21 and 22 (2 days): 75% paid by Social Security

$$\text{€}44.68 \cdot 75$$

All amounts are before taxes.

2.2 Air quality standards in Europe

In recent years, the European Parliament and the Council have passed a series of directives aimed at harmonizing air quality standards across EU member states (Council of the European Union, 1999; Council of the European Union and Parliament of the European Union, 2000, 2002, 2004, 2008). The directives have established legally binding limits on ambient concentrations for a variety of air pollutants. The most recent one,

Table 1: Air quality standards for selected air pollutants

Pollutant	Concentration (per m^3)	Averaging Period	Legal Nature	Permitted exceedances each year
Sulfur dioxide	125 µg	24 hours	Limit	3
(SO_2)	350 µg	1 hour	Limit	24
Nitrogen dioxide	200 µg	1 hour	Limit	18
(NO_2)	40 µg	1 year	Limit	-
Particulate Matter	50 µg	24 hours	Limit	35
(PM_{10})	40 µg	1 year	Limit	-
Carbon Monoxide (CO)	10 mg	Max. daily 8-hour mean	Limit	-
Ozone (O_3)	120 µg	Max. daily 8-hour mean	Target	25 days averaged over 3 years

Source: Abridged from European Environment Agency, `http://ec.europa.eu/environment/air/quality/standards.htm`

Directive 2008/50/EC establishes limit values that apply to pollutant concentrations during different time intervals, i.e. a daily mean, the maximum daily 8-hour mean or an hourly mean, and prescribes the maximum number of permitted exceedances during the course of a year.

Table 1 summarizes the limit values for the pollutants we study in this paper, namely particulate matter smaller than 10 micrometers (PM_{10}) , nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO) and ozone (O3).[4] For example, the daily mean of SO_2 shall not surpass 125 µg/m^3 more than 3 times a year. In addition, the 1-hour mean may not exceed 350 µg/m^3 more than 24 times a year. Similarly, the 24-hour daily mean of PM_{10} must not exceed 50 µg/m^3 more than 35 times and the 1-hour mean concentration of NO_2 may not exceed 200 µg/m^3 more than 18 times a year.[5] For pollutants such as CO and O_3, the limits apply to average concentrations calculated over the preceding 8 hours. The maximum of these 8-hour means for CO must not exceed 10mg/m^3. The corresponding limit for O_3 is 120 µg/m^3 and may not be exceeded on more than 25 days per year (this standard must be met only over a three-year average).

[4]The EU directive also regulates particulate matter smaller than 2.5 micrometers ($PM_{2.5}$). This pollutant is not considered in the subsequent analysis because of insufficient coverage of $PM_{2.5}$ measurements in the dataset.

[5]We also report the annual standards for the three pollutants, though this will not be pursued in the analysis below.

3 Data

For the analysis in this paper we have merged four large datasets that we describe in more detail in this section.

3.1 Employment histories

Our primary data come from the Spanish social security administration (*seguridad social*) which administrates both health insurance and pension benefits for more than 95% of the workforce in Spain. Since 2004, the administration maintains a research dataset, the Muestra Continua de Vidas Laborales (MCVL). The MCVL is a representative sample of anonymized individual work histories drawn from the universe of individuals who were affiliated with the social security at some point during the reporting year. An individual record contains information on both current-year and historical employment relations, dating back to the time when the administration began to keep computerized records.

3.2 Sick leaves

Information on sickness leaves taken by social security affiliates are first gathered and processed by the employer's mutual indemnity association which relegates the information back to the social security administration when reimbursements are claimed. While sickness leaves are not contained in the MCVL, it is possible to link this information for the individuals sampled in the MCVL, as demonstrated by Alba (2009) and Malo et al. (2012). The linking is done by staff members of the social security administration so as to ensure confidentiality. In this paper, we use information on sickness leaves during the year 2009 which were matched to social security affiliates in the MCVL samples between 2004 and 2009. For these individuals, we combine information on employment status and contribution bases in 2009 with the corresponding information on sick leaves for that year.

3.3 Air quality

Data on air quality come from AirBase, an extensive database of measurements of air quality in the Member States of the European Union (EU) and other countries working with the European Environment Agency (EEA). Data are collected annually by the EEA under a mandate from the Council of the European Union.[6]

With AirBase the European Topic Centre for Air and Climate Change provides a unified interface for accessing these data through the EEA website.[7] The database is comprised of time series data on ambient concentrations of a variety of air pollutants with up to hourly resolution as well as meta-data on monitoring stations. In its current version 8, AirBase contains data of almost 900 air quality measurement stations across in Spain between January 1986 and December 2012. Figure 1a shows a map with the exact location of each air quality monitor in the sample. Apart from location, the monitors differ in terms of the set of air pollutants they monitor and the time window of measurement (the vast majority of stations is still active). Information on the GPS-coordinates of the station allows us to link them to the nearest municipality in order to construct a dataset of air quality across Spanish towns. When more than one air quality station is located in a town, the readings are averaged across stations.

3.4 Weather

Meteorological data were downloaded from the website of the European Climate Assessment & Dataset project (ECA&D).[8] The ECA&D project collects daily data on twelve essential climate variables provided by national meteorological institutes and research institutions. For Spain, historical information is available from 1896 onward. The number of variables and geographical coverage has been increasing steadily until today. The data are delivered at the level of the weather station (of which there are 117) and include the geographic coordinates of its location and other relevant meta-

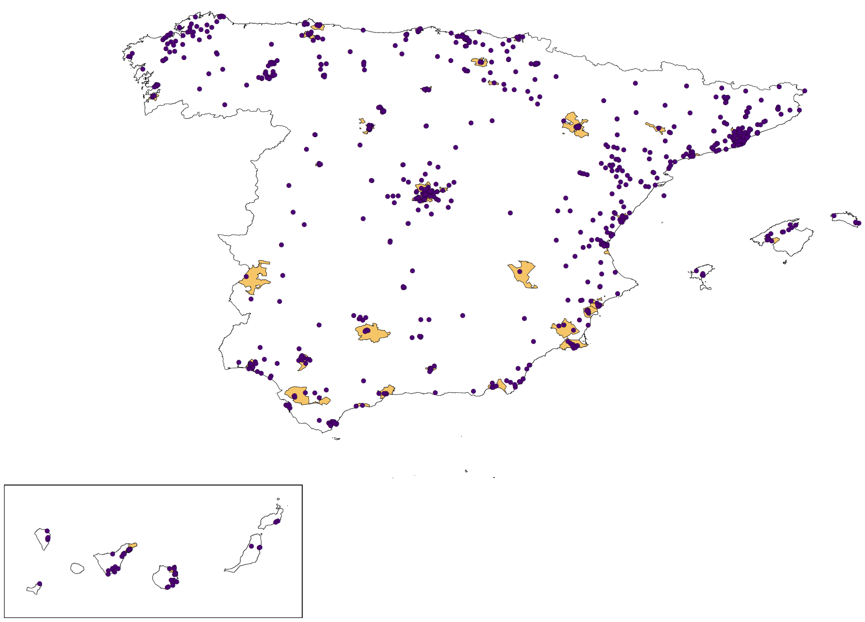
[6] Council Decision 97/101/EC of 27 January 1997 establishing a reciprocal exchange of information and data from networks and individual stations measuring ambient air pollution within the Member States, OJ L 35, 5.2.1997, p. 14–22.

[7] Available online at `http://acm.eionet.europa.eu/databases/airbase/`

[8] The ECA&D project was initiated by the European Climate Support Network of GIE-EUMETNE, an association of 31 European national meteorological agencies currently coordinated by the Royal Netherlands Meteorological Institute. The project website is available online at `http://eca.knmi.nl/`

Figure 1: Geographic coverage

(a) Location of air quality monitors



(b) Municipalities in the sample

Table 2: Descriptive Statistics: Individuals

	mean	sd	min	max	count
Absence ($\cdot 10^4$)	312.9	1167.6	0	10,000	261,545
Age	39.0	11.3	16	65	261,545
Female	0.47	0.50	0	1	261,545
Income	1,681.9	935.5	4.4	37,324.8	261,545

data. Based on the geographic coordinates, we assign to each of municipality weather conditions at the station that is located closest to the municipality's centroid and has non-missing data. This means that more the assigned weather station is not necessarily located within the boundaries of the municipality. In this way, we assign daily measurements for the key meteorological variables to more than 50 Spanish cities.

3.5 Descriptive Statistics

We merge the worker data to daily pollution and weather data on the basis of the 5-digit municipality code of the worker's primary residence. Because the MCVL provides the municipality of residence only for individuals living in municipalities with more than 50,000 inhabitants,[9] the matched dataset is representative only of the more densely populated municipalities in Spain. Figure 1b displays the location of the 49 municipalities contained in the dataset.

The estimation sample comprises 261,545 workers, 47% of which are female (cf. Table 2). Worker age ranges from 16 to 65 years and averages at 39 years. The average propensity to take a sick leave on a given day 3.13% (in Table 2 and in the regressions reported below, we have scaled the frequency of sick leaves by a factor of 10^4). Figure 2 plots the duration of these sick leaves, and Figure 3 lists the main diagnosis codes reported.

Table 3 summarizes the covariates in the merged dataset which is organized as a worker-by-day structure. Panel A reports daily mean values for PM_{10}, NO_2, SO_2 and maximum daily 8-hour means for CO and O_3. These measures closely correspond to the legally binding limits on short-term concentrations summarized in Table 1, and are reported in the corresponding units of measurement.

[9]For individuals living in smaller municipalities, only the province of residence is provided in the MCVL.

Figure 2: Duration of sick leaves

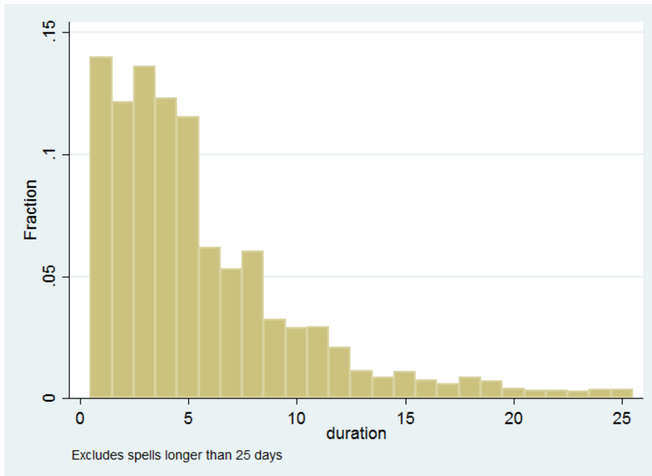


Figure 3: Most frequent diagnosis codes

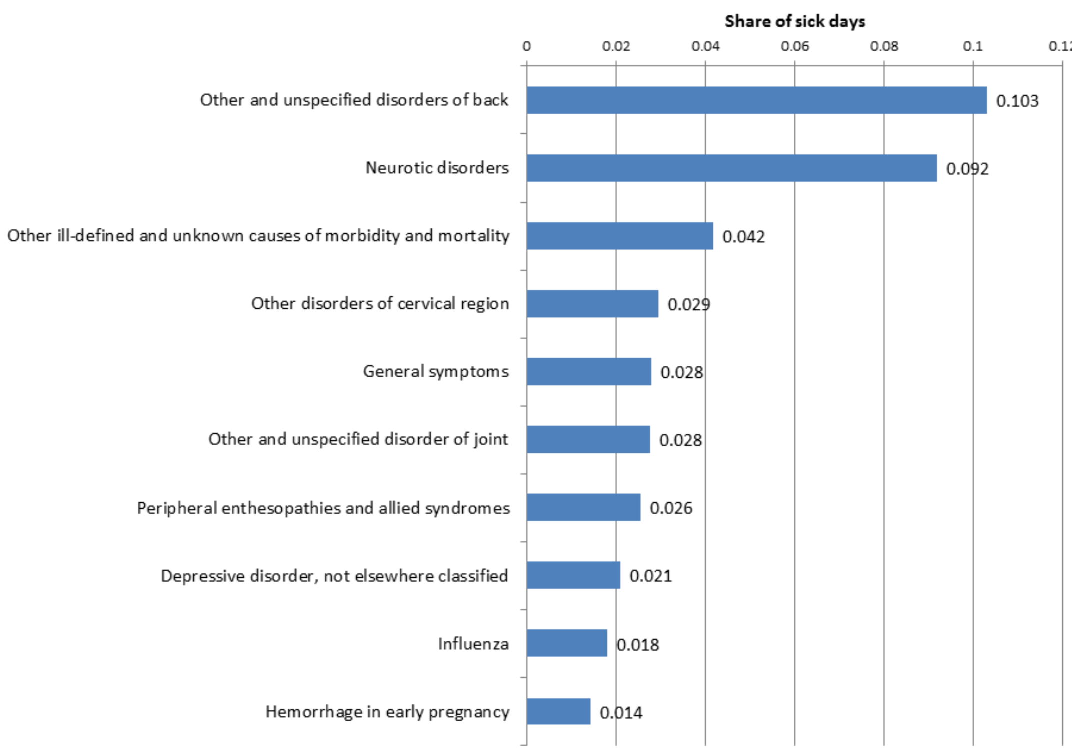


Table 3: Descriptive statistics: Pollution and weather

Variable	mean	sd	min	max	count
A. Pollution: Mean concentrations					
PM_{10} daily mean	25.14	14.85	0	151.66	16060
PM_{10} discount (Sahara dust)	2.24	6.97	0	220	16060
SO_2 daily mean	5.41	3.85	0	62.28	16060
CO maximum 8-hour mean	.52	.35	0	4.30	16060
O_3 maximum 8-hour mean	71.46	25.71	2	162.25	16060
NO_2 daily mean	29.52	17.63	.58	121.17	16060
B. Pollution: Intervals relative to EU standard					
$PM_{10} \geq 25\%$ of limit	.83	.38	0	1	16060
$PM_{10} \geq 50\%$ of limit	.46	.50	0	1	16060
$PM_{10} \geq 75\%$ of limit	.18	.38	0	1	16060
$PM_{10} \geq 100\%$ of limit	.06	.23	0	1	16060
$SO_2 \geq 25\%$ of limit	.002	.043	0	1	16060
$SO_2 \geq 50\%$ of limit	0	0	0	0	16060
$SO_2 \geq 75\%$ of limit	0	0	0	0	16060
$SO_2 \geq 100\%$ of limit	0	0	0	0	16060
$CO \geq 25\%$ of limit	.004	.060	0	1	16060
$CO \geq 50\%$ of limit	0	0	0	0	16060
$CO \geq 75\%$ of limit	0	0	0	0	16060
$CO \geq 100\%$ of limit	0	0	0	0	16060
$O_3 \geq 25\%$ of limit	.94	.25	0	1	16060
$O_3 \geq 50\%$ of limit	.67	.47	0	1	16060
$O_3 \geq 75\%$ of limit	.25	.43	0	1	16060
$O_3 \geq 100\%$ of limit	.02	.15	0	1	16060
$NO_2 \geq 25\%$ of limit	.13	.33	0	1	16060
$NO_2 \geq 50\%$ of limit	.00	.03	0	1	16060
$NO_2 \geq 75\%$ of limit	0	0	0	0	16060
$NO_2 \geq 100\%$ of limit	0	0	0	0	16060
C. Weather					
Mean temperature	16.1	6.9	-6.0	32.1	15847
Wind speed	30.0	20.4	0	153	16060
Precipitation	13.4	46.1	0	1327	16060
Cloud cover	3.7	2.5	0	8	16060
Sunshine	74.6	41.5	0	149	16060
Humidity	65.2	15.7	17	100	16060
Pressure	1015.6	6.4	972.2	1067.3	16060

Table 4: Correlation of pollution measures

	PM_{10}	SO_2	CO	O_3	NO_2
PM_{10}	1				
SO_2	0.09	1			
CO	0.25	0.42	1		
O_3	-0.01	-0.25	-0.33	1	
NO_2	0.35	0.48	0.48	-0.40	1

In some of the regressions below, we examine a possible nonlinear relationship between health and air quality. Following Currie et al. (e.g. 2009b), we partition the support of the distribution of pollution measurements using quartiles of the EU limit value (between 0% and 25% of the limit, between 25% and 50% of the limit, between 50% and 75% of the limit, between 75% and 100% of the limit, and above the limit). Panel B of Table 3 reports descriptive statistics of dummy variables that we define for each partition. This exercise shows that EU air quality standards were exceeded only for particulate matter (with a frequency of 6%) and ozone (with a frequency of 2%). In contrast, mean concentrations for carbon monoxide, sulfur dioxide and nitrogen oxide were much lower than their annual standards. [10]

Panel C summarizes the weather variables. Daily average temperature is measured in degrees Celsius, wind speed in 0.1 meters per second, precipitation in 0.1 millimeters and cloud cover in integer-valued oktas ranging from 0 (sky completely clear) to 8 (sky completely cloudy). Sunshine is measured in 0.1 hours per day, humidity in per cent and pressure in hectopascals.

Table 4 shows that various pollution measures are strongly correlated with one another.

4 Empirical model

We aim to model the propensity to take a sick leave in response to poor air quality. Our econometric approach is a close analogue to the one taken in the literature on the impact of pollution on health outcomes, in that it relates ambient pollution concentrations

[10]This is not to say that the ambient concentrations of carbon monoxide, sulfur dioxide and nitrogen oxide are innocuous. In fact, the World Health Organization has recommended much stricter air quality standards than the EU to avoid health problems (WHO, 2006).

to a binary health outcome.[11]

4.1 Baseline specification

Our baseline specification is a linear probability model (LPM) for the event that individual i living in municipality m takes a sick leave on day t as

$$SICK_{imt} = x'_{mt}\alpha + w'_{mt}\beta + \gamma_i + \mu_{mk} + \tau_{mt} + \varepsilon_{imt} \tag{1}$$

where x_{mt} is a vector of ambient pollution concentrations in municipality m, vector w_{mt} contains second-order polynomials of the weather variables, γ_i is an individual effect, and μ_{mk} is a municipality-by-month fixed effect. In addition, we include a vector τ_{mt} which includes fixed effects for day of week, bank holidays, school vacation.

For computational ease, we implement this regression at the municipality level using a two-stage approach similar to the one used in Currie et al. (2015). In the first stage, we estimate the absence rate Λ_{mt} on day t in municipality m as the prediction of the following linear regression with worker fixed-effects

$$SICK_{imt} = \Lambda_{mt} + \gamma_i + \nu_{imt} \tag{2}$$

In the second stage, we regress the daily absence rate $\hat{\Lambda}_{mt}$ on all the municipality level covariates as in eq. (1) above,

$$\hat{\Lambda}_{mt} = x'_{mt}\tilde{\alpha} + w'_{mt}\tilde{\beta} + \tilde{\mu}_{mk} + \tilde{\tau}_{mt} + \nu_{imt} \tag{3}$$

4.2 Instrumental variable estimation

The literature is concerned about the possible endogeneity of air pollution for a number of reasons. First, economic fluctuations that affect both employment and pollution might confound the estimates. For example, an unobserved shock to labor demand might induce both an increase in local pollution while also increasing labor supply. Second, traffic induced air pollution might be endogenous to sick leaves, simply because sick people do not drive to work. Third, individuals that are more susceptible

[11] In this draft we do not exploit information on the length of a sickness spell, conditional on having taken a leave.

to adverse health impacts of air pollution might choose to live in less polluted areas. The direction of the bias depends upon which of these possible sources of endogeneity prevails. Our research design limits the possible consequences of such endogeneity in various ways. For instance, by restricting the sample to full-time employees, we consider a sub-population with limited possibilities of intensive-margin adjustments labor supply in response to a supply shock. In addition, our focus on sick leaves discards variation from extensive-margin adjustments to labor supply which are not related to a temporary disability. What is more, the various time effects in the estimation, in particular the month-by-municipality effects control for unobserved shocks to labor demand, but also for shocks to the absence rate induced by, for example, a flue epidemic. Finally, the within-municipality-and-month estimation purges the estimates from the effects of locational sorting by individuals or firms. Nonetheless, to investigate if there are remaining sources of endogeneity, we estimate equation (1) by two-stage least squares (2SLS) using quasi-experimental variation in the part of PM_{10} that is due to Sahara dust advection.

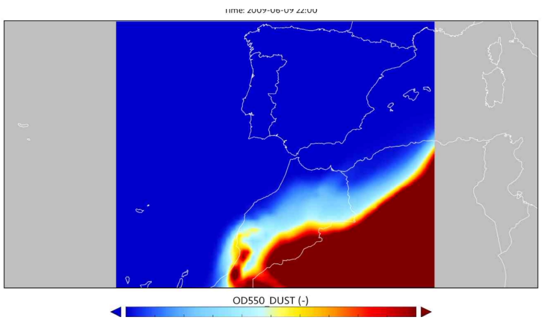
4.2.1 Sahara dust

Under certain meteorological conditions, storms in the Sahara desert stir up dust into high altitudes. These dust clouds can travel very long distances and they reach European territory several times a year. In Spain, the arrival of Sahara dust is popularly known as a “Calima” episode. These episodes are more frequent in Southern Spain and on the islands than in the North of the country. They typically last a few days and are accompanied by specific weather patterns.

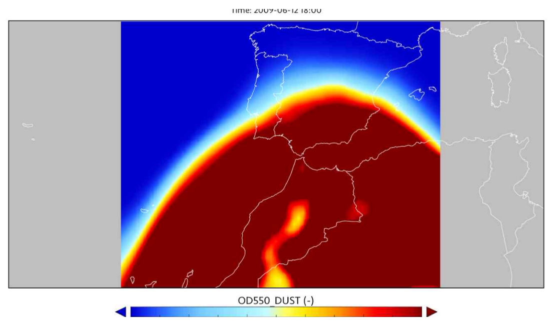
Figure 4 shows, based on data from the Barcelona Supercomputing Centre, how Sahara dust affected different regions in Spain during the month of June 2009. At the onset of the Calima episode, on June 9, the dust plume was located over North Africa only (Figure 4a). Three days later, on June 12, the plume had extended to cover the Canary islands and Southern Spain, but not the Balearic islands (Figure 4b). On June 14, all of Spain was exposed to Sahara dust, though the intensity varied across regions (Figure 4c). Figures 4d and 4e show how the effect levels off over the next two weeks.

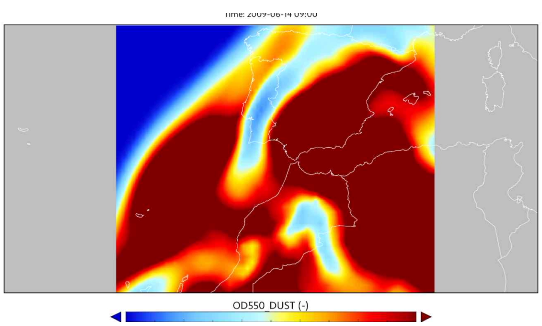
Figure 4: Calima episode in June 2009

(a) June 9, 10pm

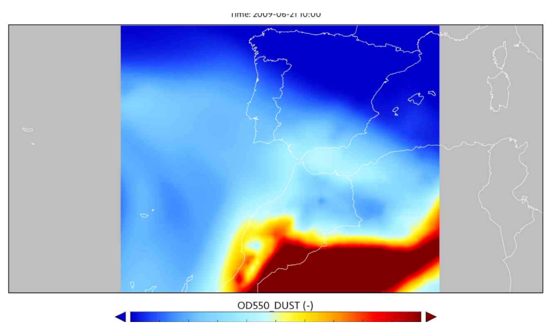


(b) June 12, 4pm

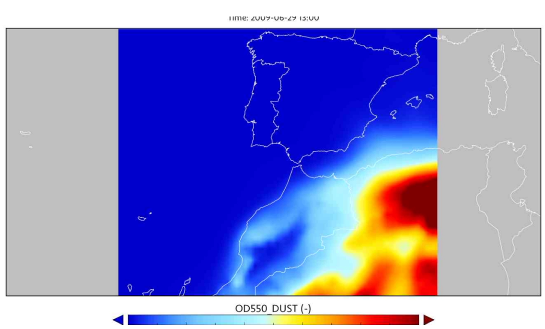




(c) June 14, 9am



(d) June 21, 10am



(e) June 29, 1pm

Source: Based on data from the BSC Dreams model, Barcelona Supercomputing Centre.

4.2.2 Sahara dust as an instrumental variable for local PM_{10}

The concentration of PM_{10} from Sahara dust is not measured at regular air quality monitors. Rather, for each Calima day, Sahara dust concentrations are attributed ex post following a standardized procedure approved by the European Commission, which is described in detail by Escudero et al. (2007) and Querol et al. (2013). A standardized procedure was designed in order to ensure a level playing field when determining whether municipalities are in compliance with the EU standard for PM_{10} concentrations. Because Calima events substantially increase non-anthropogenic PM_{10} concentrations, the affected municipalities are allowed to discount the measured PM_{10} concentration by an estimate of the Calima induced increase in PM_{10} concentrations. This PM_{10} discount varies by day and across space. We use the discount as an instrumental variable for PM_{10} concentrations because (i) it shifts local PM_{10} concentrations and (ii) it depends only on meteorological circumstances and is thus orthogonal to local labor market conditions, conditional on weather.

If PM_{10} blown in from the Sahara has a substantially different effect on human health than PM_{10} from local sources, then the IV estimate would be biased. The epidemiological literature has studied the chemical composition and human-health consequences of PM_{10} from the Sahara vs. that of non-desert PM_{10} in European cities. With regard to the former, Perez et al. (2008, Fig. 3) compare mass-adjusted concentrations of the four group elements in PM_{10} and find that crustal elements are more frequent during Saharan dust days whereas carbon, secondary aerosols and marine aerosols show no difference. In regards to the latter, a recent study concludes that "the health effects of dust-derived PM_{10} are of the same (or similar) magnitude as those reported for anthropogenic sources of air pollution" (Stafoggia et al., 2016, p. 418). We therefore assume that the instrumental variable has no direct effect on human health except through raising overall PM_{10} concentrations.

4.2.3 Data

Data on PM_{10} discounts were downloaded from the website of the Spanish Ministry of the Environment and Agriculture (`www.mapama.gob.es`). The data provided daily discounts for 29 locations in Spain. To each municipality, we link the closest station with available data. Summary statistics for the instrumental variable are reported in

Table 5: Particulate Matter: Daily mean effects

	(1)	(2)	(3)	(4)
	Dep. var.: Absence rate in municipality			
PM_{10}	0.341***	0.280***	0.285***	0.260***
	(0.0994)	(0.0532)	(0.0581)	(0.0439)
Observations	16,060	15,847	15,847	15,847
Weather controls		X	X	X
Day-type FE		X	X	X
Bank-holiday FE		X	X	X
School-holiday FE		X	X	X
Month FE		X	X	-
Municipality FE			X	-
Municipality-by-month FE				X

Notes: All coefficients are scaled by a factor of 10,000 for better readability. Weather controls are second-order polynominals of weather variables. Fixed effects for day type include three groups: weekdays non-adjacent to weekend, weekdays adjacent to weekend , and weekend days. Robust standard errors in parentheses are clustered by municipality and by day. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

the second row of Panel A of Table 3 above.

5 Results

5.1 OLS estimates

Estimation of the baseline specification (1) at the individual level based on 68 million worker-day observations proved to be computationally expensive. To avoid this, and because the identifying variation in pollution is not measured at the individual level, we implement all regressions using the two-step specification described above. In what follows, we report the results of the second-stage equation (3). Standard errors are clustered both at the municipality level to control for arbitrary serial correlation within municipalities and at the day level to account for common shocks across municipalities on a given day.

Table 5 shows how the estimated impact of particulate matter on the absence rate in a municipality changes as more controls are included in the regression. The point

Table 6: Particulate matter: Intervals

	(1)	(2)	(3)	(4)
	Dep. Var.: Absence rate in municipality			
$PM_{10} \geq 25\%$ of limit	5.081	4.640**	4.699**	4.139**
	(3.286)	(1.797)	(2.068)	(1.569)
$PM_{10} \geq 50\%$ of limit	-1.943	1.918*	1.890*	2.884***
	(1.680)	(1.083)	(1.077)	(0.597)
$PM_{10} \geq 75\%$ of limit	9.564***	4.889***	4.667***	3.434***
	(2.522)	(1.224)	(1.244)	(0.635)
$PM_{10} \geq 100\%$ of limit	9.714***	3.693***	3.790***	2.656***
	(2.859)	(1.050)	(1.142)	(0.785)
Observations	16,060	15,847	15,847	15,847
Weather controls		X	X	X
Day-type FE		X	X	X
Bank-holiday FE		X	X	X
School-holiday FE		X	X	X
Month FE		X	X	-
Municipality FE			X	-
Municipality-by-month FE				X

Notes: All coefficients are scaled by a factor of 10,000 for better readability. Weather controls are second-order polynominals of weather variables. Fixed effects for day type include three groups: weekdays non-adjacent to weekend, weekdays adjacent to weekend, and weekend days. Robust standard errors in parentheses are clustered by municipality and by day. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

estimate is always positive and statistically significant throughout, but its magnitude shrinks by roughly one third as controls for weather, location and time are included. According to the most conservative estimate of 0.000026, a reduction of the daily average concentration by one standard deviation (14.85µg) would reduce the absence rate by 0.04 percentage points from the mean of 3.13%.

Table 6 shows the results from the alternative estimation with dummies for five intervals of PM_{10} concentrations. These results confirm that PM_{10} has a positive and significant impact on the propensity to take a sick leave. The effect monotonically increases with concentrations. Although the increase is not exactly linear, the linear specification seems to provide a reasonably good approximation of the effect of increase PM_{10} concentrations on the absence rate. In what follows, we focus on the

Table 7: All pollutants: Daily mean effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Absence rate in municipality					
PM_{10}	0.260***					0.128***
	(0.0439)					(0.0444)
SO_2		0.574***				-0.394
		(0.157)				(0.241)
CO			10.64***			-1.624
			(3.214)			(1.855)
O_3				-0.299***		-0.131***
				(0.0444)		(0.0310)
NO_2					-0.197***	0.310***
					(0.0357)	(0.0463)
Observations	15,847	15,847	15,847	15,847	15,847	15,847
Weather controls	X	X	X	X	X	X
Day-type FE	X	X	X	X	X	X
Bank-holiday FE	X	X	X	X	X	X
School-holiday FE	X	X	X	X	X	X
Municipality-month FE	X	X	X	X	X	X

Notes: All coefficients are scaled by a factor of 10,000 for better readability. Weather controls are second-order polynominals of weather variables. Fixed effects for day type include three groups: weekdays non-adjacent to weekend, weekdays adjacent to weekend , and weekend days. Robust standard errors in parentheses are clustered by municipality and by day. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

linear specification which is easier to interpret.

Table 7 investigates the sensitivity of the results from the specification with municipality-by-month to including pollutants other than PM_{10}. When estimating the model for individual pollutants, a statistically significant association with the absence rate is found for each of them. However, when all pollutants are included in the regression, the coefficients on CO and SO_2 become statistically insignificant, and the sign of the coefficient on NO_2 flips. The coefficient on PM_{10} halves in magnitude. This illustrates the problem of collinear pollutants documented in Table 4 and emphasizes the need for variation in PM_{10} that is uncorrelated with concurrent variation in other pollutants.

A number of empirical studies on air quality and health have emphasized the importance of lagged pollution concentrations on health outcomes Currie et al. (e.g. 2009a).

Table 8: All pollutants: Daily mean effects and first lags

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Absence rate in municipality					
PM_{10}	0.260***		0.232***	0.128***		0.116***
	(0.0439)		(0.0309)	(0.0444)		(0.0289)
SO_2				-0.394		-0.348*
				(0.241)		(0.205)
CO				-1.624		-1.834
				(1.855)		(1.610)
O_3				-0.131***		-0.0941***
				(0.0310)		(0.0269)
NO_2				0.310***		0.320***
				(0.0463)		(0.0422)
$PM_{10}(t-1)$		0.170***	0.0676		0.108	0.0566
		(0.0536)	(0.0492)		(0.0764)	(0.0521)
$SO_2(t-1)$					-0.239	-0.0930
					(0.185)	(0.136)
$CO(t-1)$					-1.547	-0.189
					(2.018)	(1.656)
$O_3(t-1)$					-0.105***	-0.0687**
					(0.0305)	(0.0263)
$NO_2(t-1)$					0.117*	-0.0317
					(0.0598)	(0.0487)
Observations	15,847	15,468	15,468	15,847	15,468	15,468
Weather Controls	X	X	X	X	X	X
Day-type FE	X	X	X	X	X	X
Bank-holiday FE	X	X	X	X	X	X
School-holiday FE	X	X	X	X	X	X
Municipality-by-month FE	X	X	X	X	X	X

Notes: All coefficients are scaled by a factor of 10,000 for better readability. Weather controls are second-order polynominals of weather variables. Fixed effects for day type include three groups: weekdays non-adjacent to weekend, weekdays adjacent to weekend, and weekend days. Robust standard errors in parentheses are clustered by municipality and by day. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In Table 8 we report the results of specifications including only contemporaneous pollution, only lagged pollution, and both contemporaneous and lagged ambient pollution concentrations. The results in columns 1-3 indicate that lagged concentrations of PM10 have no significant impact on the absence rate. The results in columns 4-6 lend further support to this while also showing that, conditional on contemporaneous pollution, the only statistically significant coefficient on lagged pollution is obtained for ozone. In view of this, we maintain a baseline specification without lagged pollution.

5.2 IV estimates

The IV estimation is implemented as a 2-stage-least-squares (2SLS) estimator where presumably endogenous local PM_{10} is first regressed on the PM_{10} discount variable and controls. In the second stage, the absence rate is regressed on predicted PM_{10} and controls. The results are reported in Table 9.[12] The first-stage regression with an $R^2 = .47$ shows that the PM_{10} discount is a strong predictor of ambient PM_{10} concentrations. The point estimates, however, hardly differ between the OLS and 2SLS estimations. This suggests that we do not need to worry about a possible endogeneity of PM_{10} to sick leaves in our research design, where day-to-day variation in the variables of interest allows us to control for unobserved local shocks using municipality-by-month fixed effects and other time dummies.

5.3 Calculating the benefits of air quality improvements

The estimation results are suitable for calculating the monetary benefit of air quality improvements in Spain. Consider a permanent reduction in PM_{10} concentrations by one standard deviation, or 14.85μg. Among the total of 13,805,950 full-time employees in 2009, we calculate the number of additional worker days induced by this air quality improvement as

$$0.257 \cdot 10^{-4} \cdot \frac{\text{days}}{\mu\text{g}} \cdot 14.85\mu\text{g} \cdot 13,805,950 = 5,269\,\text{days}$$

[12] We do not have suitable instrumental variables for concentrations of pollutants other than PM_{10} and hence exclude those from the regression. This omission does not bias the results as long as the PM_{10} discount is exogenous to the other pollutants.

Table 9: IV estimation: PM_{10} only

	(1) OLS	(2) Reduced form	(3) First stage	(4) 2SLS
	Dep. var.: Absence rate in municipality			
PM_{10}	0.260***			0.257**
	(0.0439)			(0.102)
PM_{10} discount		0.172**	0.668***	
		(0.0741)	(0.0854)	
Observations	15,847	15,847	15,847	15,847
R^2	0.807	0.802	0.468	0.807
Adjusted R^2	0.799	0.794	0.446	0.799
Weather Controls	X	X	X	X
Day-type FE	X	X	X	X
Bank Holiday FE	X	X	X	X
School Holiday FE	X	X	X	X
Municipality-by-month FE	X	X	X	X

Notes: All coefficients are scaled by a factor of 10,000 for better readability. Weather controls are second-order polynominals of weather variables. Fixed effects for day type include three groups: weekdays non-adjacent to weekend, weekdays adjacent to weekend, and weekend days. Robust standard errors in parentheses are clustered by municipality and by day. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

where we have used the IV point estimate. We value the aggregate daily productivity gain using the average daily wage of 99.49€[13]

$$5,269\,\text{days} \cdot 99.49\frac{€}{\text{day}} = €524,252$$

Scaled up to the annual level, this air quality improvement would result in a substantial gain of more than €191 million.

6 Conclusions

TO BE COMPLETED

[13]This is net of employer contributions to social security. Using producer wages would result in a higher benefit estimate.

References

Alba, A. (2009). *La incapacidad temporal para el trabajo: análisis economómico de su incidencia y su duración*. Technical report, Universidad Carlos III de Madrid.

Chay, K. Y. & Greenstone, M. (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *Quarterly Journal of Economics*, 118(3), 1121–1167.

Council of the European Union (1999). Council Directive (EU) 1999/30/EC.

Council of the European Union and Parliament of the European Union (2000). Directive 2000/69/EC of the European Parliament and of the Council.

Council of the European Union and Parliament of the European Union (2002). Directive 2002/3/EC of the European Parliament and of the Council .

Council of the European Union and Parliament of the European Union (2004). Directive 2004/107/EC of the European Parliament and of the Council .

Council of the European Union and Parliament of the European Union (2008). Directive 2008/50/EC of the European Parliament and of the Council.

Currie, J., Davis, L., Greenstone, M., & Walker, R. (2015). Environmental health risks and housing values: Evidence from 1,600 toxic plant openings and closings. *American Economic Review*, 105(2), 678–709.

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., & Rivkin, S. G. (2009a). Does pollution increase school absences? *Review of Economics and Statistics*, 91(4), 682–694.

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

Currie, J., Neidell, M., & Schmieder, J. F. (2009b). Air pollution and infant health: Lessons from New Jersey. *Journal of Health Economics*, 28(3), 688 – 703.

Ebenstein, A., Lavy, V., & Roth, S. (2016). The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics*, 8(4), 36–65.

Escudero, M., Querol, X., Ávila, A., & Cuevas, E. (2007). Origin of the exceedances of the European daily PM limit value in regional background areas of Spain. *Atmospheric Environment*, 41(4), 730 – 744.

Hanna, R. & Oliva, P. (2015). The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City. *Journal of Public Economics*, 122, 68 – 79.

Hansen, A. C. & Selte, H. K. (2000). Air pollution and sick-leaves. *Environmental and Resource Economics*, 16(1), 31–50.

Karlsson, M., Schmitt, M., & Ziebarth, N. R. (2015). The short-term population health effects of weather and pollution. *Working Paper*.

Lichter, A., Pestel, N., & Sommer, E. (2015). *Productivity Effects of Air Pollution: Evidence from Professional Soccer*. IZA Discussion Papers 8964, Institute for the Study of Labor (IZA).

Malo, M. A., Cueto, B., García Serrano, C., & Pérez Infante, J. I. (2012). *La medición del absentismo: Estimaciones desde la perspectiva de las empresas y de las vidas laborales*. Technical report, Ministerio de Empleo y Seguridad Social.

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

Perez, L., Tobías, A., Querol, X., Künzli, N., Pey, J., Alastuey, A., Viana, M., Valero, N., González-Cabré, M., & Sunyer, J. (2008). Coarse particles from Saharan dust and daily mortality. *Epidemiology*, 19(6), 800–807.

Querol, X., Alastuey, A., Pey, J., Escudero, M., Castillo, S., Orío, A., Pallarés, M., Jiménez, S., Ferreira, F., Marques, F., Monjardino, J., Cuevas, E., Alonso, S., nano, B. A., Salvador, P., & de la Rosa, J. (2013). *Procedimiento para la identificación de episodios naturales de PM10 y PM2.5 y la demostración de de causa en lo referente* *a las superaciones del valor límite diario de $PM_1 0$*. Technical report, Ministerio de Medio Ambiente, y Medio Rural y Mariona - España.

Roth, S. (2016). *The Contemporaneous Effect of Indoor Air Pollution on Cognitive Performance: Evidence from the UK*. Mimeograph, LSE, London, UK.

Schlenker, W. & Walker, W. R. (2016). Airports, air pollution, and contemporaneous health. *Review of Economic Studies*, 83(2), 768–809.

Stafoggia, M., Zauli-Sajani, S., Pey, J., Samoli, E., Alessandrini, E., Basagaña, X., Cernigliaro, A., Chiusolo, M., Demaria, M., Díaz, J., Faustini, A., Katsouyanni, K., Kelessis, A., Linares, C., Marchesi, S., Medina, S., Pandolfi, P., Pérez, N., Querol, X., Randi, G., Ranzi, A., Tobias, A., Forastiere, F., & MED-PARTICLES Study Group (2016). Desert dust outbreaks in Southern Europe: contribution to daily $PM_1 0$ concentrations and short-term associations with mortality and hospital admissions. *Environmental Health Perspectives*, 124(4), 413–419.

WHO (2006). *WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide*. Geneva, Switzerland: World Health Organization.

Zivin, J. G. & Neidell, M. (2012). The impact of pollution on worker productivity. *American Economic Review*, 102(7), 3652–3673.