The Effect of Air Pollution on Fertility in 657 European Regions

Árpád Stump[a], Ágnes Szabó-Morvai[b,c]

[a] *Ministry of Finance, 1051 Budapest, József nádor tér 2-4.*
[b] *HUN-REN KRTK KTI, 1097 Budapest, Tóth Kálmán u. 4.*
[c] *University of Debrecen, 4032 Debrecen, Egyetem tér 1, Hungary*

Abstract

This study examines the impact of ambient air pollution on birth rates in Europe. We estimate the causal effect of air pollution on fertility by utilizing variations in wind speed and the number of heating days as instrumental variables for air quality. Our analysis encompasses 657 NUTS-3 regions, with each region having 2 to 6 years of observations between 2015 and 2020. Thus, our study is the first to extend this analysis to multiple countries, pollutants, and years. Our findings indicate that a one standard deviation increase in particulate matter concentration levels leads to a 5.1% decrease in birth rates the following year and an additional 5.9% decrease two years later. Moreover, a similar increase in air pollution has a more pronounced adverse effect on fertility in countries with lower GDP. Other pollutants have little role in shaping fertility outcomes. This result is important for environmental policies with limited resources.

Keywords: Ambient Air Pollution, Fertility, Instrumental Variables, Particulate Matter
JEL: Q53, J13, I14

[0]**Corresponding author**: Agnes Szabo-Morvai (ORCID: 0000-0003-3629-6352). HUN-REN KRTK KTI and University of Debrecen, email: szabomorvai.agnes@krtk.hun-ren.hu, Fax: +36-1-224 6700, Phone: +36-30-407-1224, Postal Address: HUN-REN KRTK KTI, 1097 Budapest, Tóth Kálmán u. 4. Árpád Stump email: stumparpad@gmail.com. **Acknowledgment**: We thank Anna Adamecz, Bíró Anikó and the participants of the Health and Society Lendület Group Meeting and the ESPE Annual Meeting for their insightful comments. We are grateful to Bálint Herczeg for the data collection. **Statements and Declarations Competing interests**: None. **Funding**: The authors gratefully acknowledge financial support from the Hungarian National Scientific Research Program (OTKA), Grant no. FK131422, the Lendület programme of the Hungarian Academy of Sciences (LP2018-2/2018). Supported by the University of Debrecen Program for Scientific Publication. The sponsors did not have any role during the design, analysis, and the writing of this study. **Author contributions**: Szabó-Morvai: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Supervision; Validation; Roles/Writing - original draft; Writing - review editing. Stump: Methodology; Formal analysis; Validation; Visualization; Roles/Writing - original draft; and Writing - review editing.

1. Introduction

In the past 70 years, fertility rates have been falling in most developing countries, raising concern for the sustainability of pension and healthcare systems. At the same time, air pollution has become a leading environmental and health concern all around the world (Fowler et al., 2020). In this research, we examine whether and how much air pollution affects fertility rates. The results of the analysis have direct policy relevance for developed countries aiming to combat low fertility.

Air pollution is significant health concern in modern societies. The World Health Organization (WHO) ranked ambient air pollution among the ten most important threats to global health in 2019 (WHO, 2019). According to the 2018 Special Report of the European Court of Auditors (ECA, 2018), lost years of healthy life from ambient air pollution is on average 0.75 per one hundred inhabitants in Europe. In 2021, 97% of the urban population in Europe was exposed to particulate concentrations exceeding the WHO guidelines (European-Council, 2024). More specifically for this study, air pollution significantly affects fertility, as shown by several studies summarized by the meta-studies of Frutos et al. (2015), Conforti et al. (2018), Jahnke et al. (2022), and Siegel et al. (2023). These meta-analyses present several studies examining the effects of air pollution concentrations on fertility, fecundability, sperm count, miscarriages, and still births. Most studies include only one pollutant at a time and thus may suffer from omitted variables bias. Moreover, some articles include multiple pollutants and some of them also employ instrumental variables methods to circumvent omitted variables bias stemming from omitted pollutants and unmeasured factors (Slama et al., 2013; Nieuwenhuijsen et al., 2014; Zanobetti et al., 2014; Nobles et al., 2018; Godzinski and Suarez Castillo, 2021). But even these latter studies, using typically small sample sizes, refer to only one country (or state, or city) or cover a short period of time. Our study uses a multiple pollutant method combined with an instrumental variable strategy, using wind speed and heating days as instruments for pollution concentrations. We analyze yearly birth rates and the air pollution data of 36 countries in Europe and its neighborhood at the NUTS 3 region level, including the concentration levels of the ten most important pollutants. To our knowledge, this is the first article to study the fertility effects of air pollution using an extended number of countries and years and, at the same time, including multiple pollutants. As a result, our results have strong external validity.

There is a large body of literature that studies how air pollution affects various fertility outcomes in the short and long run. Levine et al. (2017) documents a dramatic, more than 50% decrease in sperm count between 1973-2011 worldwide, and there is direct evidence on the causal relationship between air pollution and declining semen quality such as concentration, count, and motility (Qian et al., 2022). Furthermore, particulate matter may induce inflammatory processes and hormonal disruption which may have a detrimental effect on fecundability (Siegel et al., 2023). Additionally, according to recent evidence, particulate matter may also reach the placenta which can increase risks to the fetus (Bové et al., 2019). Pollutants also have a negative effect on ovarian reserve and increase the risk of endometriosis and PCOS which reduce fecundability (Siegel et al., 2023).

In the next section, we discuss our data collection. Section 3 describes our empirical method and provides details of robustness checks. We present our main results and heterogenous effects, and the related robustness checks in Section 4. In Section 5, we present the results of linear simulation results using our main model. Finally, Section 6 concludes the article.

2. Data

First, we collect air quality data from the European Environment Agency (EEA) using a web scraping technique in order to gather the air quality data collected using a representative sample of measuring stations that member states upload to the Internet. Note that the most frequently studied pollutants in the literature are particulate matter, carbon-monoxide, sulfur-dioxide, and nitrogen-dioxide. We extend the list of pollutants and collect information about nitrogen dioxide (NO_2), nitrogen monoxide (NO), different nitrogen oxides (NOx), ozone (O_3), sulfur dioxide (SO_2), different sizes of particulate matter ($PM_{2.5}$ and PM_{10}), benzene (C_6H_6), lead (Pb), and carbon monoxide (CO). We downloaded more than 1.1 billion data points (see Table A.5 in the Appendix). CO pollution is measured in mg/m^3, while all the other pollutants are measured in µg/m^3.

We first clean the database and delete observations that are not on hourly or daily frequency, observations with negative concentration values, and all non-validated observations (mostly missing values). Second, we calculate daily averages from the hourly data. We connect the stations to NUTS-3 regions using the GPS coordinates of the measurement stations. Since countries only report a representative selection of their air quality data to the EEA, we do not have stations in each NUTS-3 region. When there are more stations in a region, we calculate the average daily concentration across stations for each NUTS-3 region.

We aggregate the daily NUTS-3 average concentration levels to the yearly frequency using three methods. First, we calculate the yearly mean for each pollutant and for each NUTS-3 region. Second, we calculate deciles of the daily pollution levels across every NUTS-3 region for the whole observation period. Then for each year and NUTS-3 region, we count the number of days when the pollution concentration was in or above the given decile. For instance, $D9_{rtp}$ shows the number of days when the concentration level of pollutant p was in the 9^{th} or the 10^{th} pollution decile in region r and year t. Third, we also examine the concentrations relative to the European Air Quality Standards as of 2023 (ACS, see Table A.6 and Figures A.7 and A.8 in the Appendix)[1]. For example, ACS $125\%_{\mathrm{rtp}}$ is the number of days when the concentration of pollutant p exceeded 125% of the relevant concentration limit in year t and region r.

Birth rates are based on Eurostat data (EUROSTAT, 2022), and are calculated as the ratio between the number of live births and the number of women of reproductive age (15-44) on 1^{st} January. The total female population in the NUTS-3 regions is used as weights in the regressions. We also use EUROSTAT data to include NUTS-3-level GDP per person.

The NUTS-3 level yearly heating degree days (HDD) data are provided by the Joint Research Centre's AGRI4CAST Resources Portal (EUROSTAT, 2021). HDD is a weather-based technical index which is higher if there is more need for heating, taking into account the outdoor temperature, the usual indoor temperature, and technical details of the buildings.

We also use NUTS-2 level daily wind speed data (measured in km/h) from the Copernicus Climate Change Service (Commission, 2020). From the daily observations, we calculate the yearly mean wind speed.

Tables A.7 and A.8 in the Appendix show the overall coverage of the variables. Note that we only include the NUTS-3 regions that have at least one pollution data observation. HDD has good coverage, as there is data available for every NUTS-3 region in the EU, but, in general, there are no observations for countries outside of the EU (e.g, the UK). The birth rate is available for most of the regions, however, demography structure indicators (e.g., the female population aged 15-44) were only available from 2014 and on.

As Table A.7 in the Appendix indicates, there is a trade-off between using many pollutants and many regions in our regressions. As a solution, we include pollutants with a low number of observations (NO, C_6H_6, and Pb) in the LASSO regressions, but we omit them from the main regression analyses. Still, there is a strong correlation between the yearly levels of the remaining pollutants (Table A.9 in the Appendix). For example, the correlation coefficient between $PM_{2.5}$ and PM_{10} pollution levels is 0.81. NO_2 has a very strong correlation with NOx, the correlation coefficient is 0.74.

To circumvent this issue, we use principal factor analysis to combine the highly correlated pollutants. As a result, we are left with three pollutant variables in the main regressions: *PM Factor*, *NO Factor*, and *SO_2*. *PM Factor* includes PM_{10}, $PM_{2.5}$, and CO. The primary sources of particulate matter pollution are local combustion (e.g., traffic, metal industry plants), residential heating with solid fuels in cold seasons and biological material (e.g.,

[1]For some pollutants, such as NO and NOx, no daily mean pollution threshold values are set by the EU. For these pollutants, we use the annual target value or the maximum daily 8-hour mean value.

vegetative debris, spores and pollen) in warm and dry seasons (Sillanpää et al., 2006). The largest part of CO emissions come from the incomplete combustion of vehicle fuels. *NO Factor* includes NO_2, NOx stemming from the combustion of fossil fuels, like car emissions and O_3 which is produced as a result of a chemical reaction of NO_2 and NOx with oxygen in the presence of heat and sunlight. *SO_2* is included in the regressions by itself. This pollutant is mainly created when electric utilities and power plants burn coal and oil. For the details of the factorization, see Table A.10 in the Appendix.

3. Empirical method

3.1. Naive OLS

Our goal is to estimate the effect of air pollution on fertility. First, we estimate naive regressions of pollution indicators in the previous year on the natural logarithm of the birth rate. In our main specification, ambient air pollution is measured by the number of days in a year when the concentration of a pollutant exceeded 125% of the air quality standards. It is important that all pollutants are included in the regression at the same time because they are correlated and many of them may affect fertility. Examining only one pollutant at a time would likely cause the estimates to suffer from omitted variables bias.

The observations are aggregated to the year (t) and NUTS-3 region (r) level. We include year fixed effects (η_t) to control for any general shock that affected the regions at the same time, such as Europe-wide economic cycles. We also include region fixed effects (λ_r) to control for unobserved differences between regions that are unchanged in a few years of time, such as social norms that influence environmental consciousness and fertility decisions. Finally, we allow for region-specific linear time trends ($\lambda_r \times t$) of fertility in the model. Throughout the analysis, we use robust standard errors clustered at the NUTS-3 level. We estimate the following model:

$$ln(Y_{rt}) = \sum_{i=1}^{5} \beta_i P^i_{rt-1} + \eta_t + \lambda_r + \lambda_r \times t + \varepsilon_{rt} \tag{1}$$

where Y_{rt} is the birth rate, the number of births per 1000 women of age 15 to 44 in region r and year t, P^i_{rt} is the concentration level of pollutant i in region r and year t, and ε_{rt} is the error term. We calculate robust standard errors clustered at the NUTS-3 region level.

Air pollution can possibly affect fertility in the longer run as well. To test this, we include 2-year lags of the pollutants in our second specification.

$$ln(Y_{rt}) = \sum_{i=1}^{5} \beta_i P^i_{rt-1} + \sum_{i=1}^{5} \gamma_i P^i_{rt-2} + \eta_t + \lambda_r + \lambda_r \times t + \varepsilon_{rt} \tag{2}$$

Next, we include regional-level GDP as an additional control variable.

$$ln(Y_{rt}) = \sum_{i=1}^{5} \beta_i P^i_{rt-1} + \sum_{i=1}^{5} \gamma_i P^i_{rt-2} + \tau GDP_{rt-1} + \eta_t + \lambda_r + \lambda_r \times t + \varepsilon_{rt} \quad (3)$$

3.2. Instrumental variables approach

There can be other region-specific time-variant variables that we cannot observe, such as future expectations or regional variations in spending on public services (health services and public transport). Not controlling for them in the analysis may lead to a bias of unknown direction and size in our point estimates.

To circumvent this source of bias, we follow an instrumental variables design. Our instruments are wind speed and the number of heating days. These variables have been used as instruments for pollution in the literature before. Knittel et al. (2016) use local weather conditions, Schwartz et al. (2015), Schwartz et al. (2017), Zabrocki et al. (2022), Godzinski and Suarez Castillo (2021), and Deryugina et al. (2019) use wind direction and speed, and Arceo et al. (2016) use temperature (thermal inversions) to instrument endogenous ambient air pollution concentrations.

In the bulk of the previous literature, only one or just a few pollutants have been included in the regressions. In this case, even using an instrumental variable design does not provide unbiased point estimates, because the exclusion restriction likely does not hold when the instrument affects the pollutants omitted from the regressions (Benmarhnia et al.).

We can use wind speed and the number of heating days and their nonlinear functions as instruments because they affect ambient air pollution concentration and composition. Higher wind speed helps to dissipate high concentrations of ambient air pollution. Whereas, on cold winter days, the emissions increase as a result of the heating activity. The number of heating days captures this relation. Figure A.6 in the Appendix depicts the associations between the pollutants and the instruments.

The instrumental variables strategy provides unbiased estimates if the exclusion restriction holds. This ultimately consists of two parts. First, the instrumental variables should be exogenous to fertility rates in the sense that these are not affected by any other factors that may correlate with fertility rates, such as economic cycles. In the case of wind and weather, it is safe to assume that the daily and short term yearly deviations from the average are not affected by any of these factors.

Second, it is important that these weather conditions only affect fertility through air pollution and no other channels. In the previous literature, we know of no evidence that wind speed or the number of cold days would directly affect fertility rates (see e.g. Lam and Miron, 1996). Nevertheless, it might be a concern that a high number of cold days could affect agricultural activities negatively; however, the number of heating days occurs mostly in winter, thus it has a moderate effect on agriculture. Moreover, the share of agriculture in the GDP is very low in the EU, 1.6% on average in 2021 (WorldBank, 2024). Thus, it is unlikely that the number of heating days could have a meaningful effect on fertility through employment or GDP. Likewise, as the food markets of the EU are integrated, it is not likely that cold days in one region could significantly affect food prices and through them fertility.

As mentioned before, we include 3 pollution variables (PM Factor, NO Factor, and SO_2) in the main regressions, thus we need at least 3 instruments. As Figure A.6 shows, there is a nonlinear relationship between the pollutants and the instruments. Thus, we use non-linear combinations of two instruments, including squared and cubic values, interactions and indicator functions, altogether 22 instruments[2]. We include the same lagged values of the instruments as of the pollutants. In our main specification, one and two-year lags of the pollutants and the instruments are included. We run two-stage least squares (2SLS) regressions. The first-stage results show how strong and significant the relationship is between the instruments and pollution concentrations. The first stage for the pollution concentrations one year before birth is:

$$P^i_{r,t-1} = \sum_{j=1}^{2}\sum_{k=1}^{22}(\pi_{k,t-j}Z_{k,t-j}) + \tau GDP_{rt} + \eta_t + \lambda_r + \lambda_r \times t + \varepsilon_{rt} \tag{4}$$

where subscript j denotes the number of lags, and k is the k$^{\text{th}}$ instrument from the list of instruments, r denotes region, and t stands for year. The first stage for the pollutants two years before birth is:

$$P^i_{r,t-2} = \sum_{j=1}^{2}\sum_{k=1}^{22}(\pi_{k,t-j}Z_{k,t-j}) + \tau GDP_{rt} + \eta_t + \lambda_r + \lambda_r \times t + \varepsilon_{rt} \tag{5}$$

The reduced-form equations are the following:

$$ln(Y_{rt}) = \sum_{j=1}^{2}\sum_{k=1}^{22}(\pi_{k,t-j}Z_{k,t-j}) + \tau GDP_{rt} + \eta_t + \lambda_r + \lambda_r \times t + \varepsilon_{rt} \tag{6}$$

[2] Instrument list: Mean wind speed (WS), WS^2, WS^3, Number of heating days (HDD), HDD^2, HDD^3, $HDD \times WS$, $HDD \times WS^2$, $HDD \times WS^3$, $HDD \times WS$, $HDD^2 \times WS$, $HDD^3 \times WS$, $Days(WS > 4km/h)$, $Days(WS > 5km/h)$, $Days(WS > 6km/h)$, $Days(WS > 7km/h)$, $Days(WS > 8km/h)$, $[Days(WS > 4km/h)]^2$, $[Days(WS > 5km/h)]^2$, $[Days(WS > 6km/h)]^2$, $[Days(WS > 7km/h)]^2$, $[Days(WS > 8km/h)]^2$.

3.3. LASSO estimations

In the OLS and 2SLS regressions, we simplified the estimations in two ways. First, we omitted NO, C_6H_6, and Pb due to the low number of observations. If these pollutants affect fertility, our regressions suffer from omitted variables bias. Second, we combined individual pollutants into factors, but we may want to know how much each of these affect fertility.

We use the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996) which allows us to include each pollutant that we observe in the data. Thus, we can get an idea about how important those factors are that we omitted from the main specifications, and whether we need to worry about them. Moreover, LASSO also permits us to separately evaluate the importance of each pollutant in the same regression, without combining them in factor variables.

LASSO is very similar to ordinary least squares (OLS) regression, except that the minimand function of the optimization does not only include the residual sum of squares (RSS), but also a penalty term (λ) that increases with larger absolute values of the regression coefficients (see equation 7). In practice, this optimization method finds the curve that fits the data best, using as small a number of variables with $\beta \neq 0$ as possible.

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}|\beta_j| = RSS + \lambda\sum_{j=1}^{p}|\beta_j| \tag{7}$$

In other words, the lasso technique uses shrinkage and thus offers a simple way to select a model with reasonably few variables, which performs the best out-of-sample prediction of the dependent variable (James et al., 2013). Recently, it has become an accepted method to use machine learning techniques to select control variables, see Angrist and Frandsen (2019), Böheim and Stöllinger (2020), and Fluchtmann et al. (2020), for instance. In the baseline LASSO specification, we use the cross-validation function to select λ and we use a linear LASSO model.

3.4. Robustness checks

First, we include other measures of ambient air pollution concentrations. Besides the number of days when the concentrations exceeded 125% of the European air quality standard concentration limits, we measure the pollution concentration with the number of days when the concentrations exceeded 125% of the European air quality standard concentration limits; the number of days when the concentration levels reached the 10^{th} decile of pollution; and the mean pollution concentrations, as described in Section 2.

Second, we use different factorization methods to generate PM Factor and NO Factor variables. In the baseline specification, we use the principal factor method, and as a robustness check, we use the principal-component factor method, iterated principal-factor method, and maximum-likelihood factor method.

Third, we check whether the results of the LASSO estimation vary with the specification. In the baseline result we use the minimum of the CV function to select λ. In the robustness checks, we use "one-standard-error rule" (Hastie et al., 2015); the minimum value of the BIC function; the minimum of the BIC function where models are fit for all lambdas in the grid until the tolerance value is reached; adaptive ridge (adaptive lasso, using the ridge estimator to construct the initial weights in the first lasso); adaptive steps (adaptive lasso with 100 lassos); and adaptive power 1.5 (adaptive lasso, where weights are raised to the 1.5$^{\text{th}}$ power). Additionally, we check alternative seeds.

3.5. Heterogeneity

In our dataset, we included many EU regions, and thus are able to present a heterogeneity analysis. We divide the sample by the average levels of PM concentrations through the observation period. The high pollution subsample includes NUTS-3 regions with higher than median PM pollution levels and the low pollution subsample includes those with lower than median levels. Next, we do the same with GDP and run the 2SLS regressions on these subsamples. These two dimensions are somewhat correlated ($\rho = -0.3$), as the wealthier regions are less polluted in general. Still, about 30% of the regions are in the "high pollution - high GDP" or the "low pollution - low GDP" categories.

4. Results

4.1. Descriptive results

First, in Table 1 the descriptive statistics of the main variables are presented. The reported statistics refer to yearly values by NUTS-3 regions, except wind speed statistics which refer to the NUTS-2 level.

There is substantial variability in the birth rates and the pollution concentration levels not only at the country level, but also at the regional level. This is shown by the maps in Figures 1 and 2. The maps showing the rest of the pollutants are presented in Figures A.9 to A.17 in the Appendix.

The yearly observations reported in the Appendix (Table A.11) show that the concentration of NO_2 substantially decreased in the observation period, whereas the concentration of other pollutants such as PM_{10} and O_3 remained more or less unchanged.

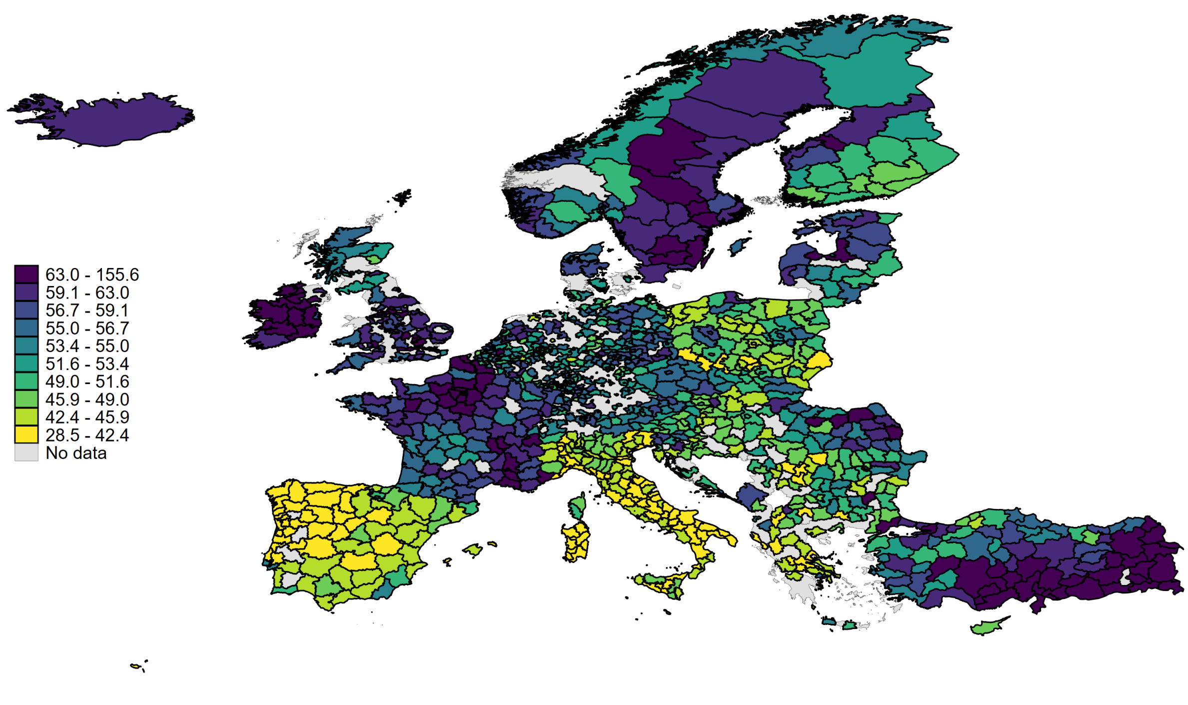
Figures A.4 and A.5 in the Appendix depict the raw associations between the concentration levels of each pollutant in year t and the log birth rates in year t+1 at the NUTS-3 level. The raw associations between birth rates and pollution concentrations show diverse

Table 1: Summary statistics

	Mean	SD	Min	Max	N
Birth rate	53.51	10.98	21.67	167.41	7,842
PM_{10}	18.76	13.93	0.00	137.63	9,392
$PM_{2.5}$	7.99	8.18	0.00	85.00	9,392
CO	0.17	0.27	0.00	4.23	9,392
SO_2	2.70	5.14	0.00	157.44	9,392
NO_2	15.96	11.94	0.00	118.20	9,392
NOx	19.80	26.04	0.00	214.39	9,392
O_3	38.78	24.54	0.00	111.56	9,392
Wind speed	3.03	0.88	1.05	5.72	9,200
Heating Degree Days	2620.17	867.19	266.55	6836.56	7,456

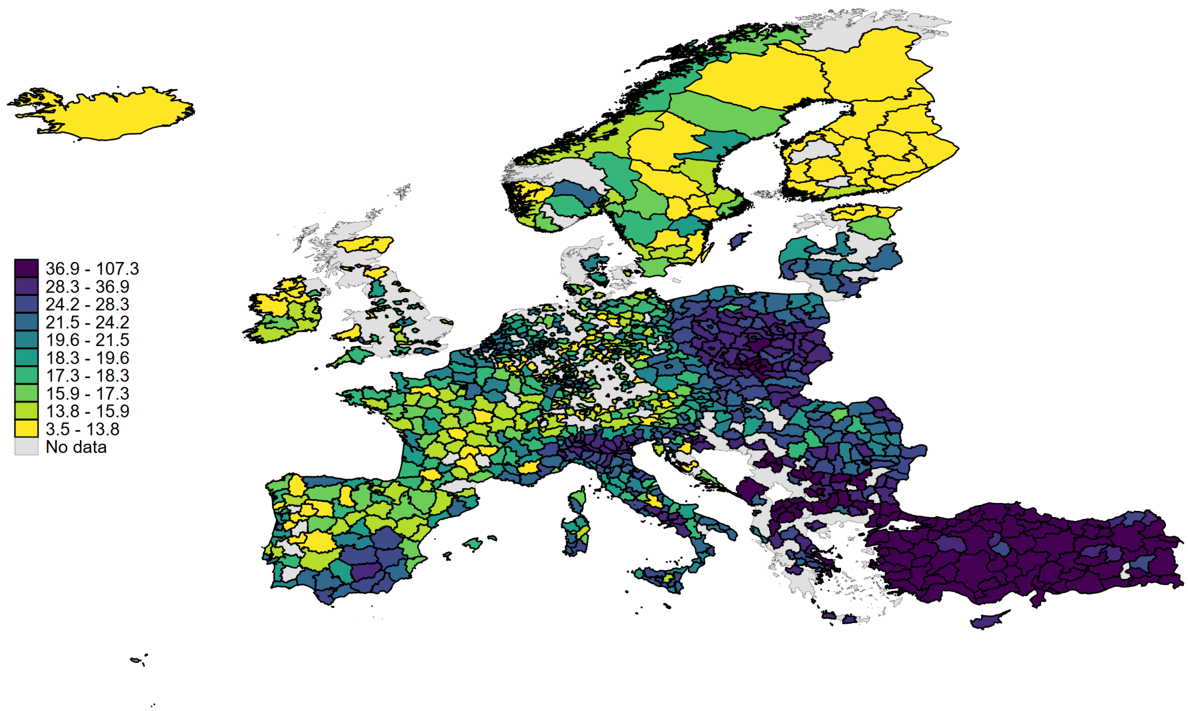
Unit of measurement: Birth rate: numre of births per 1000 women of age 15 to 44
CO - mg/m^3; other pollutants - $\mu g/m^3$; Wind speed - km/h; Heating Degree Days - none

Figure 1: Average birth rate in NUTS-3 regions (2013-2020)



Data source: EUROSTAT. Birth rate: number of births per 1000 women of age 15 to 44.

Figure 2: Average PM_{10} pollution in NUTS-3 regions (2013-2020)



Data source: European Environment Agency. Unit of measurement: $\mu g/m^3$

nonlinear patterns, including strong negative, strong positive, roughly zero, and varying associations between birth rates and pollution concentrations. These patterns reflect not only the direct negative effect of pollution on birth rates, but also the various indirect effects of secondary factors such as industrial activity and economic prosperity. We provide further descriptive figures about pollution concentrations in the Appendix; see Figure A.6 in the Appendix on the associations between instruments and pollution concentration levels and Figures A.7 and A.8 in the Appendix on the pollution distributions relative to the European air quality standards.

4.2. Main results

4.2.1. OLS and 2SLS

The main results from regression equations 1, 2, 3, and the second stage results of the 2SLS regressions are reported in Table 2. In these regressions, the air pollution concentrations are measured as the number of days when the concentration exceeds 125% of the European air quality standards (see Table A.6 in the Appendix). The pollution variables are standardized to have zero mean and the standard deviation equals one for easier interpretation of the results. In the first, second, and third columns, the simple OLS regression results are reported and suggest a moderate and significant effect of the PM Factor. The coefficients of the NO

Factor and the SO_2 are insignificant in all specifications. Nevertheless, the OLS estimates are biased, thus we do not interpret these coefficients.

In Column 4, we report the second-stage results of the 2SLS regression. Here the point estimate of the PM Factor is larger compared to the OLS estimate and is significant at the 1% level. The estimate shows that if the PM Factor increases by 1 standard deviation, the birth rates decrease by 5.1% the next year and another 5.9% two years later. 1 SD difference of the PM Factor equals approximately the difference between the PM Factor in the La Spezia region, Northwest Italy (PM Factor = -0.72; PM_{10} pollution exceeded 125% of the European air quality standards for 0 days, PM_2 for 4 days, and CO for 0 days) and the Altötting region, Bavaria, Germany (PM Factor: -0.33; Days: 15, 31, and 0 days) in 2017. The rest of the pollutants have an insignificant or a slightly significant effect on fertility in the 2SLS regression.

Table A.12 in the Appendix summarizes the first stage results of the 2SLS regression presented in Table 2. The F-statistic is meaningfully large for the PM and the NO Factors, but very low for the SO_2. At the same time, for each pollutant, the Sanderson-Windmeijer (SW) F-test statistics (Sanderson and Windmeijer, 2016) indicate that the endogenous regressors are not weakly identified. These tests show that the instruments are able to capture a meaningful part of the variation of the PM and the NO Factors. In the case of SO_2, the F-test indicates that it is weakly identified, whereas the SW F-tests suggest it is not. The SW first-stage chi-squared statistics significantly reject the null hypotheses that the particular endogenous regressors are unidentified. According to Table A.13 in the Appendix many of the coefficients are significant in the first stage regression where the PM Factor (t-1) is the dependent variable[3]. These results support that the instruments are good enough to capture the variation in the pollution concentrations.

Lastly, for comparison, we run regressions which include only one pollutant at a time, otherwise we run the exact same regressions as in Columns 3 and 4 of Table 2. The results of these regressions are reported in Table A.14 in the Appendix. As a result, nearly all the pollutants produce a significant effect in these regressions. This striking difference demonstrates how measuring the effect of only one pollutant may lead to biased estimations.

4.2.2. LASSO

The results of the baseline LASSO estimations are reported in Table 3. λ is selected by using a cross-validation function, and the grid for λ is set to 10,000. The dependent variable is log birth rate and the explanatory variables are each single air pollutants (in years t-1

[3]The rest of the first stage regressions are not reported here, but are available from the authors.

Table 2: OLS and 2SLS regression estimates

	(1)	(2)	(3)	(4)
L.PM Factor	-0.005***	-0.005***	-0.005***	-0.051***
	(0.001)	(0.002)	(0.002)	(0.010)
L2.PM Factor		-0.002	-0.002	-0.059***
		(0.002)	(0.002)	(0.013)
L.SO_2	0.027	0.009	0.009	0.299
	(0.035)	(0.043)	(0.043)	(1.198)
L2.SO_2		-0.052	-0.051	1.683
		(0.038)	(0.038)	(1.686)
L.NO Factor	-0.000	0.000	0.000	-0.005
	(0.001)	(0.001)	(0.001)	(0.011)
L2.NO Factor		0.001	0.001	0.018*
		(0.001)	(0.001)	(0.010)
Observations	5320	5320	5320	5320
Prob > F	0.010	0.000	0.000	0.000
Clusters	889.000	889.000	889.000	889.000
Model	OLS	OLS	OLS	2SLS
NUTS-3 FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
NUTS-3 linear trend	Yes	Yes	Yes	Yes
Other controls	No	No	Yes	Yes

Standard errors in parentheses

Notes: Regressions based on Eq. 1, 2, 3 and 2SLS.

Dependent variable : log birth rate.

Air quality measure: number of days when the pollution concentrations exceeded 125% of the European air quality standards, standardized

L.: first lagged values; L2.: second lagged values.

PM Factor: PM_{10}, PM_2, CO; NO Factor: NO_2, NO_X, O_3 (Principal factor method).

Robust standard errors clustered at the NUTS-3 region level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

and t-2) that we have data of (including NO, Pb, and C_6H_6), year fixed effects, country fixed effects, and country linear trends. Only the results for the air pollutants are reported in the table. Based on the absolute value of the LASSO coefficients, the first, second, and third most important pollutants are indicated in the models. Less important pollutants that were also kept in the model are marked with an x. The coefficients are not reported nor interpreted here because these are subject to omitted variables bias such as the OLS regression results. We report results for various measures of air pollution. In Column 1 we report the LASSO results with pollution measures referring to the number of days when the pollution exceeded 125% of the EU air quality standards (ACS 125%). In Column 2, ACS 175% is reported which indicates days with very heavy pollution concentrations. In Column 3, the pollution is measured with the number of days when the concentrations were in the highest, 10^{th} decile. In Column 4 the means of the pollution concentrations are used.

Table 3: LASSO results (λ selection with cross-validation)

	ACS 125%	ACS 175%	D10	Mean
PM_{10}	1	1	1	1
$PM_{2.5}$	3	3	3	3
CO			x	x
NO_2	x	x	x	x
NOx	2	2	2	x
O_3	x		x	2
SO_2	x	x	x	x
Pb	x	x	x	x
C_6H_6	x	x	x	x
NO	x	x		

Notes: The grid for λ is set to 10,000. Seed: 1234. Dependent variable: log birth rate. Independent variables: air pollutants (in years t-1 and t-2), year FE*, country FE*, country linear trend*. * Omitted from the Table. The rule used to select λ: CV - minimum of the CV function. Measures of air pollution: (1) Number of days concentration exceeds 125% of the Air quality standard, (2) Number of days concentration exceeds 175% of the Air quality standard, (3) Number of days concentration in Decile 10, (4) Mean. Results: 1: included in the model with the highest importance (highest lasso coefficient in absolute value), 2: second, 3: third, x: included in the model with lower importance. The results of other LASSO model specifications are reported in Tables A.15 and A.16 in the Appendix.

Among the three most important variables, PM_{10} and $PM_{2.5}$ are included in four, NOX is included in three, and O_3 is included in one of the four models. The CO concentrations are included in only 2 of the models which indicates that the 2SLS results on the PM Factor are most likely driven by the PM_{10} and $PM_{2.5}$ concentrations.

The LASSO estimates are suitable to judge the importance of the three variables that we omitted from the OLS and the 2SLS regressions due to a low number of observations (NO, Pb, and C_6H_6). These variables are either not included or are among the less important variables in each LASSO model. This result suggests that we most likely have not excluded any crucial pollution variables from the OLS and the 2SLS regressions.

4.3. Robustness checks

In our first robustness check, we use various measures of air pollution concentrations, and report the 2SLS estimation results in Table A.18 in the Appendix. In Column 1 we repeat the results reported in Column 4 of Table 2 for comparison. Column 2 shows the point estimates for the number of days when the pollution concentration exceeded 175% of the concentration limits. In this specification, the SO_2 variable is dropped because this pollutant never exceeds this threshold. Column 3 shows the number of days when the pollution concentration reached the 10^{th} decile, and Column 4 reports the mean concentrations. These results are very similar to the main regression results. Only the PM Factor concentrations have a significant and negative effect on birth rates one and two years later. These results are significant at the 1% level in most specifications. NO Factor and SO_2 have no significant effect on fertility.

Second, we check whether using different methods of variable reduction methods in the factor variables could affect our results. Table A.19 in the Appendix reports the results for various methods of factorizing. Using the principal-component factor and the iterated principal-factor method, our results remain similar. Whereas, using the maximum likelihood factor method, the NO Factor appears to also be an important pollutant.

Last, we check whether alternative LASSO model specifications lead to different results than the baseline specification. We get very similar results, if we apply different methods to select the λ parameter. Instead of using the minimum of the CV function, we use the "one-standard-error rule" (Hastie et al., 2015); the minimum value of the BIC function; the minimum of the BIC function where models are fit for all lambdas in the grid until the tolerance value is reached; adaptive ridge (adaptive lasso, using the ridge estimator to construct the initial weights in the first lasso); adaptive steps (adaptive lasso with 100 lassos); and adaptive power 1.5 (adaptive lasso, where weights are raised to the 1.5^{th} power). All 7 of these specifications are combined with four air pollution measures, which gives altogether 28 LASSO models. Among the three most important variables, PM_{10} is included in 14, $PM_{2.5}$ is included in 13, NOx is included in 11, and O_3 is included in 10 of the 28 models. NO, Pb, and C_6H_6 are not included among the 3 most important variables in any of these models (see Tables A.15, A.16 and A.17 in the Appendix). We also run the LASSO models with other seeds, and the results are very similar. The results are omitted from the article, but are available upon request.

4.4. Heterogeneity analysis

Finally, we present a heterogeneity analysis, with the results reported in Table 4. In the first two regressions (Columns 1 and 2) we included NUTS-3 regions of which the mean yearly average PM_{10} concentrations are above or under the overall median yearly average.

The point estimates of PM_{10} are not statistically different in the two groups, but the results are more significant in the low pollution areas. The other pollutants are insignificant, similar to the main specification (see Column 4 of Table 2).

We divide regions by GDP in a similar fashion (see Figure A.18 in the Appendix). PM_{10} one year ago is weakly or not significant. PM_{10} two years ago has a significant negative effect in the low GDP regions, but not in the high GDP regions. This result is probably due to the higher quality of health services or to the generally better health status of the population in high GDP regions. Similarly to our main result, the NO Factor is not significant in either of the two groups. However, the point estimate of SO_2 in the high GDP group suggests a large negative effect and it is highly significant.

5. Simulation

Finally, we calculate predicted birth rate differences between the actual data and a hypothetical scenario with improved particulate matter pollution levels. We assume that in the improved pollution scenario the NO Factor and SO_2 concentrations remain at the original level. We predict birth rates for each NUTS-3 region and each year, based on our main point estimates (Column 4 of Table 2). In the simulation, we assume that particulate matter pollution levels decrease such that PM_{10} exceeds ACS 125% at most 2 days a year and $PM_{2.5}$ at most 3 days a year. We limit the pollution improvement to be at most a one standard deviation decrease so that linear projections remain reasonable. Figure 3 shows the resulting differences compared to the actual birth rates. These results show that in most European regions birth rates would likely increase if the regions complied with the stricter regulations adopted in February 2024 by the European Council (EuropeanCouncil, 2024). Eastern European regions and Northern regions of Italy would likely benefit most from reducing PM_{10} and PM $_{2.5}$ pollution levels in terms of fertility rates.

6. Discussion

In this paper, we investigate the impact of different types of ambient pollutants on birth rates in Europe. Estimating the causal effect presents challenges due to omitted variable bias resulting from unaccounted pollutants and unmeasured factors. We address this by incorporating the ten most important pollutants into our analysis and employing wind speed and heating days as instrumental variables for pollution concentrations. Previous estimates regarding the influence of air pollution on fertility have typically been confined to limited geographical and temporal scopes. Our study expands upon this by examining 657 regions across Europe and its neighboring areas, drawing from up to six years of data, thereby

Table 4: Heterogeneity by PM_{10} and GDP (2SLS)

	(1) High PM_{10}	(2) Low PM_{10}	(3) High GDP	(4) Low GDP
L.PM Factor	-0.006	-0.016	-0.027*	0.010
	(0.017)	(0.010)	(0.014)	(0.016)
L2.PM Factor	-0.028	-0.018**	-0.014	-0.040**
	(0.017)	(0.009)	(0.015)	(0.019)
L.NO Factor	0.013	0.009	-0.006	0.008
	(0.015)	(0.008)	(0.012)	(0.021)
L2.NO Factor	-0.003	0.014	0.015	0.002
	(0.016)	(0.009)	(0.012)	(0.021)
L.SO_2	-0.060*	-0.011	-0.027	-0.025
	(0.037)	(0.018)	(0.036)	(0.029)
L2.SO_2	-0.035	0.005	-0.128**	-0.031
	(0.037)	(0.017)	(0.061)	(0.028)
Observations	3557	1763	2976	2344
Prob > F	0.148	0.048	0.001	0.041
Clusters	594.000	295.000	497.000	392.000
Model	2SLS	2SLS	2SLS	2SLS
NUTS-3 FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
NUTS-3 linear trend	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes

Standard errors in parentheses

Notes: Dependent variable : log birth rate. Pollution measure: mean.

L.: first lagged values; L2.: second lagged values.

PM Factor: PM_{10}, PM_2, CO; NO Factor: NO_2, NO_X, O_3 (Principal factor method).

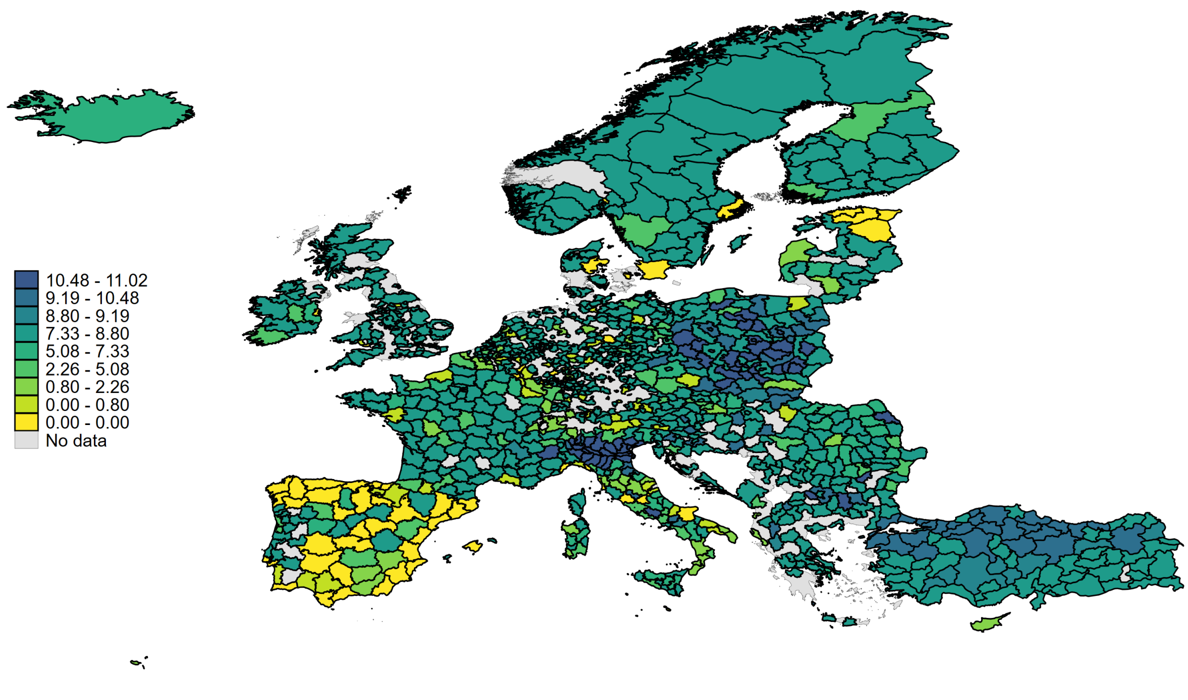
Robust standard errors clustered at the NUTS-3 region level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

enhancing the external validity of our findings. Our results remain robust across various specifications and robustness tests.

We find that it is the particulate matter concentrations, specifically $PM_{2.5}$ and PM_{10} that have a significant effect on birth rates. After controlling for these effects, the rest of the pollutants are found to exert an insignificant effect on fertility. The PM Factor coefficient is significant at the 1% level, and it suggests that an increase by 1 SD would result in a 5.1% drop in birth rates the next year and 5.9% two years later. These results are robust across various specifications. The effects of other pollutants on birth rates are insignificant in most specifications.

Figure 3: Predicted birth rate growth (%) as a result of a substantial decrease in particulate matter pollution



Notes: Linear prediction of birth rate changes if PM_{10} exceeds ACS 125% at most 2 days per year and $PM_{2.5}$ at most 3 days a year. (Assuming at most one standard deviation decrease so that linear projections remain reasonable.)

Our heterogeneity analysis shows that air pollution concentrations have a much larger effect in NUTS-3 regions with lower GDP levels.

Our results are comparable to previous results in the literature in terms of the effects of PM_{10} and $PM_{2.5}$. We find similarly strong negative effects for live births as the studies reported by Frutos et al. (2015). However, according to our results, the other pollutants such as NOx, NO_2, and SO_2 play a much smaller role in shaping birth rates across European regions. These results are similar to the results of previous multi-pollutant studies on birth rates, for instance Nieuwenhuijsen et al. (2014), who find that PMcoarse (PM larger than 2.5 but smaller than 10µg/m^3) drives the fertility effects when NOx, NO_2, $PM_{2.5}$, PM_{10} and PMcoarse are jointly included in the regressions.

We add to the previous multi-pollutant literature by extending the analysis from one region or country and (or) a short time period to the whole European region and 6 years, and show that the results previously found hold to a multitude of regions and years.

Our findings hold significance for governments striving to address air pollution amidst resource constraints. They indicate that environmental policies could synergize with population policies, such as government investments in child benefits. Therefore, policymakers can enhance the efficiency of expenditure structures by considering these complementary effects.

References

David Fowler, Peter Brimblecombe, John Burrows, Mathew R. Heal, Peringe Grennfelt, David S. Stevenson, Alan Jowett, Eiko Nemitz, Mhairi Coyle, Xuejun Liu, Yunhua Chang, Gary W. Fuller, Mark A. Sutton, Zbigniew Klimont, Mike H. Unsworth, and Massimo Vieno. A chronology of global air quality. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378(2183):20190314, September 2020. doi: 10.1098/rsta.2019.0314. URL https://royalsocietypublishing.org/doi/10.1098/rsta.2019.0314. Publisher: Royal Society.

WHO. Ten threats to global health in 2019, 2019. URL https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019.

ECA. Air pollution: Our health still insufficiently protected. Special Report 23/2018, European Court of Auditors, 2018. URL https://www.eca.europa.eu/en/Pages/DocItem.aspx?did=46723.

EuropeanCouncil. Air quality, February 2024. URL https://www.consilium.europa.eu/en/policies/air-quality/.

Víctor Frutos, Mireia González-Comadrán, Ivan Solà, Benedicte Jacquemin, Ramón Carreras, and Miguel A. Checa Vizcaíno. Impact of air pollution on fertility: a systematic review. *Gynecological Endocrinology*, 31(1):7–13, 2015. ISSN 0951-3590. doi: 10.3109/09513590.2014.958992. URL https://doi.org/10.3109/09513590.2014.958992.

Alessandro Conforti, Marika Mascia, Giuseppina Cioffi, Cristina De Angelis, Giuseppe Coppola, Pasquale De Rosa, Rosario Pivonello, Carlo Alviggi, and Giuseppe De Placido. Air pollution and female fertility: a systematic review of literature. *Reproductive biology and endocrinology: RB&E*, 16(1):117, 2018. ISSN 1477-7827. doi: 10.1186/s12958-018-0433-z.

Johanna R. Jahnke, Kyle P. Messier, Melissa Lowe, and Anne Marie Jukic. Ambient Air Pollution Exposure Assessments in Fertility Studies: a Systematic Review and Guide for Reproductive Epidemiologists. *Current Epidemiology Reports*, 9(2):87–107, June 2022.

ISSN 2196-2995. doi: 10.1007/s40471-022-00290-z. URL https://doi.org/10.1007/s40471-022-00290-z.

Eva L Siegel, Akhgar Ghassabian, Alison E Hipwell, Pam Factor-Litvak, Yeyi Zhu, Hannah G Steinthal, Carolina Focella, Lindsey Battaglia, Christina A Porucznik, Scott C Collingwood, Michele Klein-Fedyshin, Linda G Kahn, and on behalf of program collaborators for Environmental influences on Child Health Outcomes. Indoor and outdoor air pollution and couple fecundability: a systematic review. *Human Reproduction Update*, 29(1):45–70, February 2023. ISSN 1460-2369. doi: 10.1093/humupd/dmac029. URL https://doi.org/10.1093/humupd/dmac029.

Rémy Slama, Sébastien Bottagisi, Ivo Solansky, Johanna Lepeule, Lise Giorgis-Allemand, and Radim Sram. Short-term impact of atmospheric pollution on fecundability. *Epidemiology (Cambridge, Mass.)*, 24(6):871–879, 2013. ISSN 1531-5487. doi: 10.1097/EDE.0b013e3182a702c5.

Mark J. Nieuwenhuijsen, Xavier Basagaña, Payam Dadvand, David Martinez, Marta Cirach, Rob Beelen, and Bénédicte Jacquemin. Air pollution and human fertility rates. *Environment International*, 70:9–14, September 2014. ISSN 0160-4120. doi: 10.1016/j.envint.2014.05.005. URL https://www.sciencedirect.com/science/article/pii/S0160412014001470.

Antonella Zanobetti, Elena Austin, Brent A. Coull, Joel Schwartz, and Petros Koutrakis. Health effects of multi-pollutant profiles. *Environment International*, 71:13–19, October 2014. ISSN 0160-4120. doi: 10.1016/j.envint.2014.05.023. URL https://www.sciencedirect.com/science/article/pii/S0160412014001731.

Carrie J Nobles, Enrique F Schisterman, Sandie Ha, Germaine M Buck Louis, Seth Sherman, and Pauline Mendola. Time-varying cycle average and daily variation in ambient air pollution and fecundability. *Human Reproduction*, 33(1):166–176, January 2018. ISSN 0268-1161. doi: 10.1093/humrep/dex341. URL https://doi.org/10.1093/humrep/dex341.

Alexandre Godzinski and Milena Suarez Castillo. Disentangling the effects of air pollutants with many instruments. *Journal of Environmental Economics and Management*, 109:102489, September 2021. ISSN 0095-0696. doi: 10.1016/j.jeem.2021.102489. URL https://www.sciencedirect.com/science/article/pii/S0095069621000668.

Hagai Levine, Niels Jørgensen, Anderson Martino-Andrade, Jaime Mendiola, Dan Weksler-Derri, Irina Mindlis, Rachel Pinotti, and Shanna H. Swan. Temporal trends in sperm count: a systematic review and meta-regression analysis. *Human Reproduction Update*, 23(6):646–659, 2017. ISSN 1460-2369. doi: 10.1093/humupd/dmx022.

Hong Qian, Qiaoqiao Xu, Wenkai Yan, Yun Fan, Zhi Li, Chengzhe Tao, Feng Zhang, and Chuncheng Lu. Association between exposure to ambient air pollution and semen quality in adults: a meta-analysis. *Environmental Science and Pollution Research*, 29(7):10792–10801, February 2022. ISSN 1614-7499. doi: 10.1007/s11356-021-16484-9. URL https://doi.org/10.1007/s11356-021-16484-9.

Hannelore Bové, Eva Bongaerts, Eli Slenders, Esmée M. Bijnens, Nelly D. Saenen, Wilfried Gyselaers, Peter Van Eyken, Michelle Plusquin, Maarten B. J. Roeffaers, Marcel Ameloot, and Tim S. Nawrot. Ambient black carbon particles reach the fetal side of human placenta. *Nature Communications*, 10(1):3866, September 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-11654-3. URL https://www.nature.com/articles/s41467-019-11654-3. Number: 1 Publisher: Nature Publishing Group.

EUROSTAT. Fertility (National level) (demo_fer), January 2022. URL https://ec.europa.eu/eurostat/web/main/data/database.

EUROSTAT. Heating and cooling degree days - statistics, February 2021. URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Heating_and_cooling_degree_days_-_statistics#Heating_and_cooling_degree_days_by_NUTS_3_regions.

European Commission. Copernicus Climate Change Service, February 2020. URL https://climate.copernicus.eu/.

Markus Sillanpää, Risto Hillamo, Sanna Saarikoski, Anna Frey, Arto Pennanen, Ulla Makkonen, Zoya Spolnik, René Van Grieken, Martin Braniš, Bert Brunekreef, et al. Chemical composition and mass closure of particulate matter at six urban sites in europe. *Atmospheric Environment*, 40:212–223, 2006.

Christopher R. Knittel, Douglas L. Miller, and Nicholas J. Sanders. Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health. *Massachusetts Institute of Technology Press*, May 2016. ISSN 0034-6535. URL https://dspace.mit.edu/handle/1721.1/113913. Accepted: 2018-03-02T19:11:39Z Publisher: MIT Press.

Joel Schwartz, Elena Austin, Marie-Abele Bind, Antonella Zanobetti, and Petros Koutrakis. Estimating Causal Associations of Fine Particles With Daily Deaths in Boston. *American Journal of Epidemiology*, 182(7):644–650, October 2015. ISSN 0002-9262. doi: 10.1093/aje/kwv101. URL https://doi.org/10.1093/aje/kwv101.

Joel Schwartz, Marie-Abele Bind, and Petros Koutrakis. Estimating Causal Effects of Local Air Pollution on Daily Deaths: Effect of Low Levels. *Environmental Health Perspectives*, 125(1):23–29, 2017. ISSN 1552-9924. doi: 10.1289/EHP232.

Léo Zabrocki, Anna Alari, and Tarik Benmarhnia. Improving the design stage of air pollution studies based on wind patterns. *Scientific Reports*, 12:7917, May 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-11939-6. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9106699/.

Tatyana Deryugina, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif. The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction. *American Economic Review*, 109(12):4178–4219, December 2019. ISSN 0002-8282. doi: 10.1257/aer.20180279. URL https://www.aeaweb.org/articles?id=10.1257/aer.20180279.

Eva Arceo, Rema Hanna, and Paulina Oliva. 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, 2016. ISSN 1468-0297. doi: 10.1111/ecoj.12273. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecoj.12273. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecoj.12273.

Tarik Benmarhnia, Prashant Bharadwaj, and Mauricio Romero. Using Instrumental Variables under Partial Observability of Endogenous Variables for Assessing Effects of Air Pollution on Health. URL https://mauricio-romero.com/pdfs/papers/partialiv.pdf.

D. A. Lam and J. A. Miron. The effects of temperature on human fertility. *Demography*, 33 (3):291–305, 1996. ISSN 0070-3370.

WorldBank. Agriculture, forestry, and fishing, value added (% of GDP) - European Union, 2024. URL https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?locations=EU.

Robert Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. ISSN 0035-9246. URL https://www.jstor.org/stable/2346178. Publisher: [Royal Statistical Society, Wiley].

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

Joshua Angrist and Brigham Frandsen. Machine labor. Technical report, National Bureau of Economic Research, 2019.

René Böheim and Philipp Stöllinger. Decomposition of the gender wage gap using the lasso estimator. *Applied Economics Letters*, pages 1–12, 2020.

Jonas Fluchtmann, Anita M Glenny, Nikolaj Harmon, and Jonas Maibom. The gender application gap: Do men and women apply for the same jobs? Technical report, Mimeo University of Copenhagen, 2020.

T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Boca Raton, FL: CRC Press, 2015.

Eleanor Sanderson and Frank Windmeijer. A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221, February 2016. ISSN 0304-4076. doi: 10.1016/j.jeconom.2015.06.004. URL https://www.sciencedirect.com/science/article/pii/S0304407615001736.

A1. Supplementary tables and figures

Table A.5: Number of data points before and after aggregating.

Pollutant	Raw data by station hour	 by station day	Data after aggregating by nuts3 day	 by nuts3 year
C_6H_6	31 256 798	224 539	754 028	2 542
CO	72 715 222	10 248	1 331 598	3 866
NO	142 830 222	7 672	1 771 263	5 058
NO_2	231 221 335	52 483	2 742 982	7 749
NO_X as NO_2	136 133 784	10 460	1 988 002	5 688
O_3	163 347 466	2 190	2 510 812	7 098
Pb in PM_{10}	26 739	1 047 140	474 413	2 225
PM_{10}	150 259 482	3 542 713	2 718 911	7 731
$PM_{2.5}$	71 888 620	1 656 595	1 893 156	5 590
SO_2	122 923 200	66 823	1 847 375	5 319

Table A.6: EU air quality standards

Pollutant	Concentration limit (CL)	Averaging period
Fine particles ($PM_{2.5}$)	20 µg/m^3	1 year
Sulphur dioxide (SO_2)	125 µg/m^3	24 hours
Nitrogen dioxide (NO_2)	40 µg/m^3	1 year
Particulate matter (PM_{10})	50 µg/m^3	24 hours
Lead (Pb)	0.5 µg/m^3	1 Year
Carbon monoxide (CO)	10 mg/m^3	Maximum daily 8 hour mean
Benzene (C_6H_6)	5 µg/m^3	1 year
Ozone (O_3)	120 µg/m^3	Maximum daily 8 hour mean

Note: EU air quality standards according to the Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe.

Table A.7: Number of pollution observations by country tab:means_by_year}

Country	PM_{10}	SO_2	O_3	NO_2	NOx	CO	C_6H_6	NO	Pb	$PM_{2.5}$
Albania	20	28	31	31	31	32	28	0	0	21
Andorra	7	7	7	7	7	7	0	7	0	0
Austria	264	213	268	264	79	113	37	254	10	159
Belgium	210	162	220	239	28	97	166	0	114	212
Bulgaria	176	144	109	120	7	93	86	120	52	62
Croatia	70	43	81	74	73	28	24	0	14	65
Cyprus	8	8	8	8	8	8	8	0	8	8
Czechia	112	106	112	112	110	80	9	112	94	112
Denmark	32	27	47	47	41	34	8	47	15	30
Estonia	32	39	40	39	40	32	24	0	32	40
Finland	124	54	86	99	97	15	7	57	1	76
France	742	327	750	732	188	165	78	636	26	612
Germany	1685	794	1475	1683	1545	572	274	1710	418	1074
Greece	90	54	70	71	5	44	28	72	8	46
Hungary	100	96	80	96	88	90	66	0	26	53
Iceland	13	14	2	14	13	1	0	14	0	14
Ireland	58	43	57	50	50	27	17	0	0	48
Italy	795	582	757	797	387	663	626	219	366	719
Kosovo	7	6	7	5	5	6	0	3	0	7
Latvia	32	32	40	32	6	11	26	10	0	23
Lithuania	48	47	64	55	55	40	17	0	0	27
Luxembourg	8	8	8	8	7	8	8	8	2	8
Malta	16	16	16	16	7	16	13	16	10	16
Montenegro	8	8	8	8	8	8	0	8	0	1
Netherlands	201	78	202	213	213	39	25	213	6	138
North Macedonia	52	53	52	45	45	52	2	47	0	5
Norway	107	65	70	110	98	11	0	98	0	81
Poland	572	522	450	535	530	391	291	134	449	443
Portugal	142	94	141	141	140	49	23	139	7	97
Romania	257	281	274	270	37	279	188	0	164	124
Serbia	35	65	31	58	57	90	2	57	14	1
Slovakia	64	63	56	64	64	64	64	0	30	61
Slovenia	78	33	64	64	57	24	16	0	36	28
Spain	414	415	418	418	418	347	303	418	304	369
Sweden	152	50	104	112	55	20	16	13	0	83
Switzerland	68	42	68	68	68	42	17	68	0	42
Turkey	477	477	236	261	260	195	0	0	0	183
United Kingdom	425	187	486	657	656	49	33	573	17	453

Table A.8: Number of yearly NUTS-3 level observations by country

Country	Birth rate per 1000 females (15-44)	Wind	HDD
Albania	35	40	0
Andorra	0	0	0
Austria	245	280	280
Belgium	232	264	272
Bulgaria	154	176	176
Croatia	83	104	104
Cyprus	7	8	8
Czechia	98	112	112
Denmark	42	48	48
Estonia	33	40	40
Finland	126	144	144
France	686	744	744
Germany	1799	2048	2056
Greece	140	160	160
Hungary	105	120	120
Iceland	14	16	0
Ireland	56	64	64
Italy	748	864	864
Kosovo	0	0	0
Latvia	35	40	40
Lithuania	56	64	64
Luxembourg	7	8	8
Malta	14	0	16
Montenegro	7	8	0
Netherlands	203	232	232
North Macedonia	49	56	0
Norway	94	128	120
Poland	511	584	584
Portugal	140	152	144
Romania	294	336	336
Serbia	56	112	0
Slovakia	56	64	64
Slovenia	70	80	80
Spain	357	416	408
Sweden	147	168	168
Switzerland	70	80	0
Turkey	560	640	0
United Kingdom	513	800	0

Table A.9: Correlation matrix of the pollution variables

	PM_{10}	$PM_{2.5}$	CO	SO_2	NO_2	NOx	O_3
PM_{10}	1.00						
$PM_{2.5}$	0.81	1.00					
CO	0.68	0.59	1.00				
SO_2	0.66	0.54	0.55	1.00			
NO_2	0.49	0.40	0.45	0.35	1.00		
NOx	0.26	0.26	0.26	0.12	0.74	1.00	
O_3	-0.35	-0.34	-0.25	-0.21	-0.53	-0.49	1.00

Table A.10: Factor loadings of PM Factor and NO Factor

	PF	PCF	IPF	ML
PM Factor				
PM_{10}	0.718	0.903	0.798	0.632
PM_2	0.718	0.903	0.799	1.000
CO	0.008	0.017	0.008	0.019
NO Factor				
NO_2	0.141	0.745	0.218	1.000
NOx	0.127	0.670	0.189	0.000
O_3	-0.033	-0.178	-0.037	0.000

Note: PF: Principal factor; PCF: Principal-component factor; IPF: Iterated principal factor; ML: Maximum-likelihood factor. Pollution measurement method: ACS 125% - number of days when the pollution exceeded 125% of the European air quality standard.

Table A.11: Mean pollution concentrations by year

	PM_{10}	SO_2	O_3	NO_2	NOx	CO	C_6H_6	NO	Pb	$PM_{2.5}$
2013	23.205	4.410	52.131	21.254	37.182	0.405	1.319	11.615	0.018	15.175
2014	22.201	4.434	48.850	20.439	36.653	0.396	1.199	12.353	0.036	14.297
2015	25.060	5.456	52.530	20.808	37.008	0.441	1.233	13.022	0.031	14.359
2016	23.470	4.986	49.529	20.613	34.374	0.432	1.235	13.309	0.016	13.624
2017	23.509	4.851	52.107	20.293	31.545	0.417	1.206	12.072	0.015	13.684
2018	23.322	4.923	54.571	19.321	32.642	0.418	1.158	10.213	0.024	13.785
2019	21.441	4.649	54.268	18.914	32.247	0.406	0.997	9.604	0.014	12.196
2020	20.723	4.561	52.312	15.916	26.336	0.409	0.940	7.580	0.018	11.870

Table A.12: First stage results of the 2SLS regressions

Variable	F(42, 888)	P-val	SW Chi-sq(37)	P-val	SW F(37, 888)	P-val
L.PM Factor	4.150	0.000	157.940	0.000	3.510	0.000
L2.PM Factor	3.350	0.000	100.090	0.000	2.230	0.000
L.SO_2	0.060	1.000	82.000	0.000	1.820	0.002
L2.SO_2	0.060	1.000	75.100	0.000	1.670	0.008
L.NO Factor	1.580	0.012	94.020	0.000	2.090	0.000
L2.NO Factor	1.630	0.008	81.730	0.000	1.820	0.002

Notes: These statistics are reported from Eq. 4.

Table A.13: First stage results (Eq.4) with PM Factor (t-1) as the dependent variable

Dep. var.: PM Factor (t-1)	Coefficient	Robust SE	t	$P > t$	95% Conf int	
WS (t-1)	11.103	3.933	2.820	0.005	3.392	18.813
WS^2 (t-1)	-3.305	1.268	-2.610	0.009	-5.790	-0.819
WS^3 (t-1)	0.357	0.133	2.680	0.007	0.095	0.618
HDD (t-1)	0.004	0.002	1.550	0.122	-0.001	0.008
HDD^2 (t-1)	0.000	0.000	0.370	0.709	0.000	0.000
HDD^3 (t-1)	0.000	0.000	-0.580	0.565	0.000	0.000
$HDD \times WS$ (t-1)	-0.003	0.001	-2.520	0.012	-0.006	-0.001
$HDD \times WS^2$ (t-1)	0.001	0.000	2.910	0.004	0.000	0.002
$HDD^2 \times WS$ (t-1)	0.000	0.000	-0.220	0.822	0.000	0.000
$HDD \times WS^3$ (t-1)	0.000	0.000	-3.150	0.002	0.000	0.000
$HDD^3 \times WS$ (t-1)	0.000	0.000	0.280	0.781	0.000	0.000
$Days(WS > 8km/h)$ (t-1)	0.018	0.010	1.800	0.072	-0.002	0.037
$Days(WS > 7km/h)$ (t-1)	-0.004	0.006	-0.590	0.558	-0.016	0.009
$Days(WS > 6km/h)$ (t-1)	0.000	0.004	-0.090	0.929	-0.008	0.007
$Days(WS > 5km/h)$ (t-1)	-0.005	0.004	-1.360	0.173	-0.012	0.002
$Days(WS > 4km/h)$ (t-1)	-0.008	0.004	-2.130	0.033	-0.015	-0.001
$Days(WS > 8km/h)^2$ (t-1)	-0.001	0.000	-2.300	0.022	-0.001	0.000
$Days(WS > 7km/h)^2$ (t-1)	0.000	0.000	0.980	0.326	0.000	0.000
$Days(WS > 6km/h)^2$ (t-1)	0.000	0.000	-0.800	0.424	0.000	0.000
$Days(WS > 5km/h)^2$ (t-1)	0.000	0.000	1.210	0.226	0.000	0.000
$Days(WS > 4km/h)^2$ (t-1)	0.000	0.000	1.550	0.122	0.000	0.000
WS (t-2)	-0.312	3.164	-0.100	0.921	-6.515	5.891
WS^2 (t-2)	-0.960	1.094	-0.880	0.380	-3.104	1.185
WS^3 (t-2)	0.193	0.122	1.580	0.114	-0.046	0.432
HDD (t-2)	-0.003	0.002	-1.670	0.096	-0.006	0.001
HDD^2 (t-2)	0.000	0.000	0.740	0.462	0.000	0.000
HDD^3 (t-2)	0.000	0.000	-0.280	0.777	0.000	0.000
$HDD \times WS$ (t-2)	0.000	0.001	0.360	0.717	-0.002	0.002
$HDD \times WS^2$ (t-2)	0.000	0.000	0.850	0.394	0.000	0.001
$HDD^2 \times WS$ (t-2)	0.000	0.000	-0.220	0.826	0.000	0.000
$HDD \times WS^3$ (t-2)	0.000	0.000	-1.580	0.113	0.000	0.000
$HDD^3 \times WS$ (t-2)	0.000	0.000	-0.140	0.888	0.000	0.000
$Days(WS > 8km/h)$ (t-2)	0.015	0.007	1.990	0.047	0.000	0.029
$Days(WS > 7km/h)$ (t-2)	-0.005	0.006	-0.840	0.403	-0.018	0.007
$Days(WS > 6km/h)$ (t-2)	-0.003	0.004	-0.700	0.481	-0.011	0.005
$Days(WS > 5km/h)$ (t-2)	0.005	0.003	1.510	0.130	-0.002	0.012
$Days(WS > 4km/h)$ (t-2)	0.001	0.003	0.370	0.715	-0.005	0.008
$Days(WS > 8km/h)^2$ (t-2)	-0.001	0.000	-2.320	0.021	-0.001	0.000
$Days(WS > 7km/h)^2$ (t-2)	0.000	0.000	1.170	0.244	0.000	0.000
$Days(WS > 6km/h)^2$ (t-2)	0.000	0.000	-0.520	0.606	0.000	0.000
$Days(WS > 5km/h)^2$ (t-2)	0.000	0.000	-0.210	0.835	0.000	0.000
$Days(WS > 4km/h)^2$ (t-2)	0.000	0.000	-0.540	0.589	0.000	0.000
GDP	0.000	0.000	-0.370	0.709	0.000	0.000

Table A.14: OLS and 2SLS regression results: including only one pollutant at a time

	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS	(9) OLS	(10) 2SLS	(11) OLS	(12) 2SLS	(13) OLS	(14) 2SLS
PM_{10} (t-1)	-0.000	-0.005***												
	[0.001]	[0.001]												
PM_{10} (t-2)	0.000*	-0.002**												
	[0.000]	[0.001]												
PM_2 (t-1)			-0.000	-0.005***										
			[0.000]	[0.001]										
PM_2 (t-2)			0.000	-0.003**										
			[0.000]	[0.001]										
CO (t-1)					0.006	-0.027								
					[0.005]	[0.029]								
CO (t-2)					0.009*	-0.028								
					[0.005]	[0.024]								
SO_2 (t-1)							0.000	-0.011***						
							[0.001]	[0.004]						
SO_2 (t-2)							0.000	-0.009**						
							[0.000]	[0.004]						
O_3 (t-1)									0.000	-0.001*				
									[0.000]	[0.000]				
O_3 (t-2)									0.000*	0.001**				
									[0.000]	[0.000]				
NO_2 (t-1)											-0.001*	-0.005***		
											[0.000]	[0.001]		
NO_2 (t-2)											0.000	-0.000		
											[0.000]	[0.001]		
NOx (t-1)													0.000	-0.000
													[0.000]	[0.000]
NOx (t-2)													0.000**	0.002***
													[0.000]	[0.001]
Constant	-3.128***		-3.121***		-3.006***		-3.077***		-3.092***		-3.097***		-3.033***	
	[0.060]		[0.075]		[0.094]		[0.079]		[0.061]		[0.057]		[0.075]	
NUTS-3 FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
NUTS-3 linear trend	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	3,933	3,933	2,763	2,763	1,974	1,974	2,602	2,602	3,638	3,638	3,827	3,827	2,443	2,443
F test	3.80e-05	0	0.0337	0.000108	0.000932	0.00355	0.00276	0.000248	0.00134	0.00	3.32e-05	1.43e-08	0.000853	3.73e-05

*Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, clustered, robust standard errors are in brackets.*
In the OLS regressions we use Eq. 3, in the 2SLS regressions we use Eqs. 4 and 5. The dependent variable is log birth rate, and only one type of pollutant is included in the regressions at a time.

Table A.15: LASSO results with λ selection methods of cross-validation and BIC

	CV				CV SE Rule				BIC				BIC All Lambdas			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
PM_{10}	1	1	1	1										1		
PM_2	3	3	3	3					1	1			2	2	x	3
CO			x	x				1			2	1			x	2
NO_2	x	x	x	x												
NOx	2	2	2	x									1	3		
O_3	x		x	2			1				1				1	1
SO_2	x	x	x	x							x		3	x	2	
Pb	x	x	x	x												
C_6H_6	x	x	x	x												
NO	x	x														

Notes: The grid for λ is set to 10,000. Seed: 1234. Dependent variable: log birth rate. Independent variables: air pollutants (in years t-1 and t-2), year FE*, country FE*, country linear trend*. * Omitted from the Table. The rule used to select λ: CV - minimum of the CV function; CV SE Rule - "one-standard-error rule" (Hastie et al., 2015); BIC - minimum BIC function value; BIC All Lambdas - minimum of the BIC function, models are fit for all lambdas in the grid until the tolerance value is reached. Measures of air pollution: (1) ACS 125% (2), ACS 175%, (3) Decile 10, (4) Mean. Results: 1: included in the model with the highest importance (highest lasso coefficient in absolute value); 2: second; 3: third; x: included in the model with lower importance.

Table A.16: LASSO results with adaptive λ selection methods

	Adaptive Ridge				Adaptive Steps (100)				Adaptive Power 1.5			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
PM_{10}	1	1	1	1		1			1	1	1	2
PM_2	3	x	x	3					3	x	x	3
CO			x	x							x	
NO_2		x	x							3	x	
NOx	2	2	2	x					2	2	2	x
O_3			3	2				1			3	1
SO_2	x	x	x						x	x	x	
Pb			x								x	
C_6H_6			x	x							x	
NO												

Notes: The grid for λ is set to 10,000. Seed: 1234. Dependent variable: log birth rate. Independent variables: air pollutants (in years t-1 and t-2), year FE*, country FE*, country linear trend*. * Omitted from the Table. The rule used to select λ: Adaptive Ridge - adaptive lasso, using the ridge estimator to construct the initial weights in the first lasso; Adaptive Steps (100) - adaptive lasso with 100 lassos; adaptive Power 1.5 - adaptive lasso, weights are raised to the 1.5^{th} power. Measures of air pollution: (1) ACS 125%; (2) ACS 175%; (3) Decile 10; (4) Mean. Results: 1: included in the model with the highest importance (highest lasso coefficient in absolute value); 2: second; 3: third; x: included in the model with lower importance.

Table A.17: LASSO results summary

	Among the first 3 most important	Among the less important	Included in the model
PM_{10}	14	0	14
PM_2	13	5	18
CO	4	6	10
NO_2	1	7	8
NOx	11	3	14
O_3	10	2	12
SO_2	2	12	14
Pb	0	6	6
C_6H_6	0	7	7
NO	0	2	2

Notes: 3 most important: 1, 2, and 3 in Tables A.15 and A.16. Less important: x in Tables A.15 and A.16.

Table A.18: Instrumental variables estimates for various measures of ambient pollution

	(1) ACS: 125%	(2) ACS: 175%	(3) D10	(4) Mean
L.PM Factor	-0.051***	-0.025***	-0.052***	-0.035**
	(0.010)	(0.015)	(0.014)	(0.008)
L2.PM Factor	-0.059***	-0.032***	-0.057***	-0.047***
	(0.013)	(0.008)	(0.013)	(0.015)
L.NO Factor	-0.005	0.003	-0.019	-0.000
	(0.011)	(0.009)	(0.017)	(0.012)
L2.NO Factor	0.018*	0.008	0.016	0.007
	(0.010)	(0.009)	(0.016)	(0.014)
L.SO_2	0.299		0.009	-0.014
	(1.198)		(0.017)	(0.026)
L2.SO_2	1.683		-0.047**	-0.056*
	(1.686)		(0.023)	(0.032)
Observations	5320	5320	5320	5320
Prob > F	0.000	0.000	0.000	0.001
Clusters	889.000	889.000	889.000	889.000
Model	2SLS	2SLS	2SLS	2SLS
NUTS-3 FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
NUTS-3 linear trend	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes

Standard errors in parentheses

Notes: Dependent variable : log birth rate. ACS: European air quality standard.

L.: first lagged values; L2.: second lagged values.

PM Factor: PM_{10}, $PM_{2.5}$, CO; NO Factor: NO_2, NO_X, O_3 (Principal factor method).

Robust standard errors clustered at the NUTS-3 region level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.19: Instrumental variables estimates for various methods of factorizing

	(1) PF	(2) PCF	(3) IPF	(4) ML
L.PM Factor	-0.051***	-0.051***	-0.051***	-0.040***
	(0.010)	(0.010)	(0.010)	(0.011)
L2.PM Factor	-0.059***	-0.059***	-0.059***	-0.045***
	(0.013)	(0.013)	(0.013)	(0.012)
L.NO Factor	-0.005	-0.006	-0.006	-0.046**
	(0.011)	(0.011)	(0.011)	(0.022)
L2.NO Factor	0.018*	0.018*	0.018*	-0.039**
	(0.010)	(0.010)	(0.010)	(0.019)
L.SO_2	0.299	0.283	0.302	0.471
	(1.198)	(1.193)	(1.197)	(1.199)
L2.SO_2	1.683	1.660	1.680	1.504
	(1.686)	(1.673)	(1.686)	(1.548)
Observations	5320	5320	5320	5320
Prob > F	0.000	0.000	0.000	0.000
Clusters	889.000	889.000	889.000	889.000
Model	2SLS	2SLS	2SLS	2SLS
NUTS-3 FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
NUTS-3 linear trend	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes

Standard errors in parentheses

Notes: Dependent variable : log birth rate.

L.: first lagged values; L2.: second lagged values.

PM Factor: PM_{10}, $PM_{2.5}$, CO; NO Factor: NO_2, NO_X, O_3

Methods of variable reduction in the Factor variables(1) Principal factor method

(2) Principal-component factor method

(3) Iterated principal-factor method (4) Maximum-likelihood factor method

Air pollution measure: number of days when the concentrations exceeded the 125% of the air quality standards.

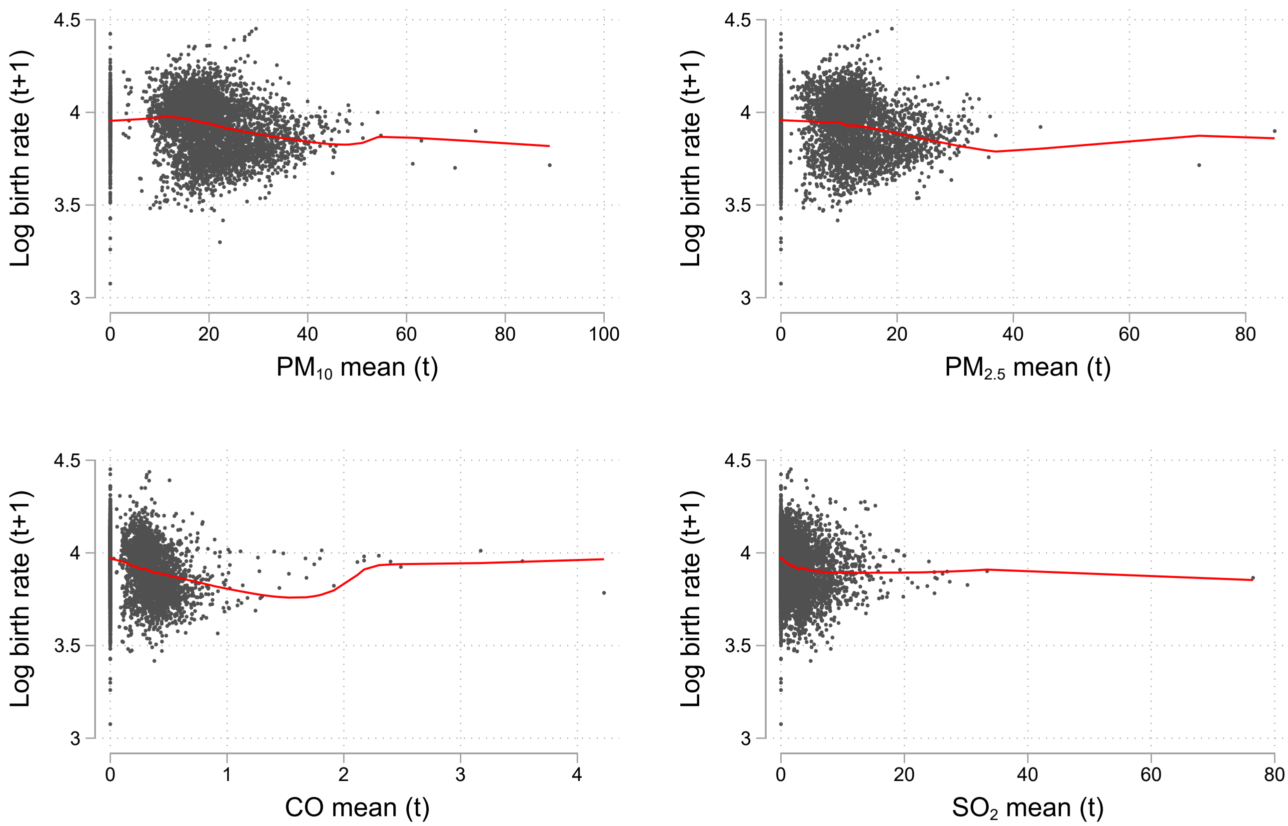
Robust standard errors clustered at the NUTS-3 region level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.20: Number of NUTS-3 - year observations by country in the regressions

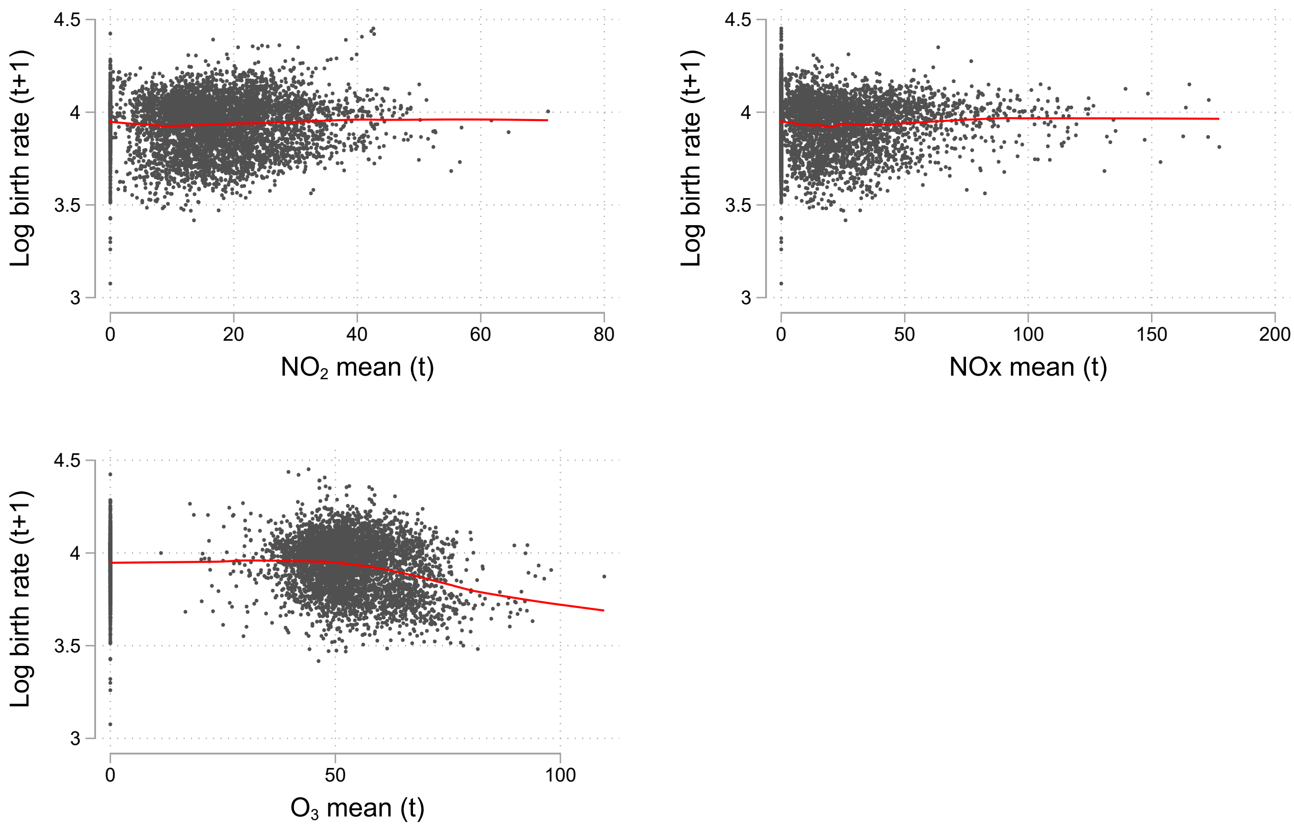
	Freq.	Percent	Cumulative
Austria	6	0.625	0.625
Belgium	5	0.521	1.146
Croatia	3	0.312	1.458
Cyprus	6	0.625	2.083
Czechia	54	5.625	7.708
Denmark	10	1.042	8.750
Estonia	12	1.250	10
Finland	5	0.521	10.52
France	10	1.042	11.56
Germany	175	18.23	29.79
Greece	224	23.33	53.12
Hungary	29	3.021	56.15
Ireland	11	1.146	57.29
Italy	88	9.167	66.46
Lithuania	18	1.875	68.33
Luxembourg	5	0.521	68.85
Netherlands	14	1.458	70.31
Poland	209	21.77	92.08
Portugal	22	2.292	94.38
Romania	5	0.521	94.90
Slovakia	37	3.854	98.75
Slovenia	6	0.625	99.38
Sweden	6	0.625	100
Total	960	100	

Figure A.4: Average pollution concentrations (in year t) and log birth rates (in year t+1)



Data source: EUROSTAT and European Environment Agency. Each point denotes a NUTS-3 region.

Figure A.5: Average pollution concentrations (in year t) and log birth rates (in year t+1)



Data source: EUROSTAT and European Environment Agency. Each point denotes a NUTS-3 region.

Figure A.6: Instruments and average pollution concentrations

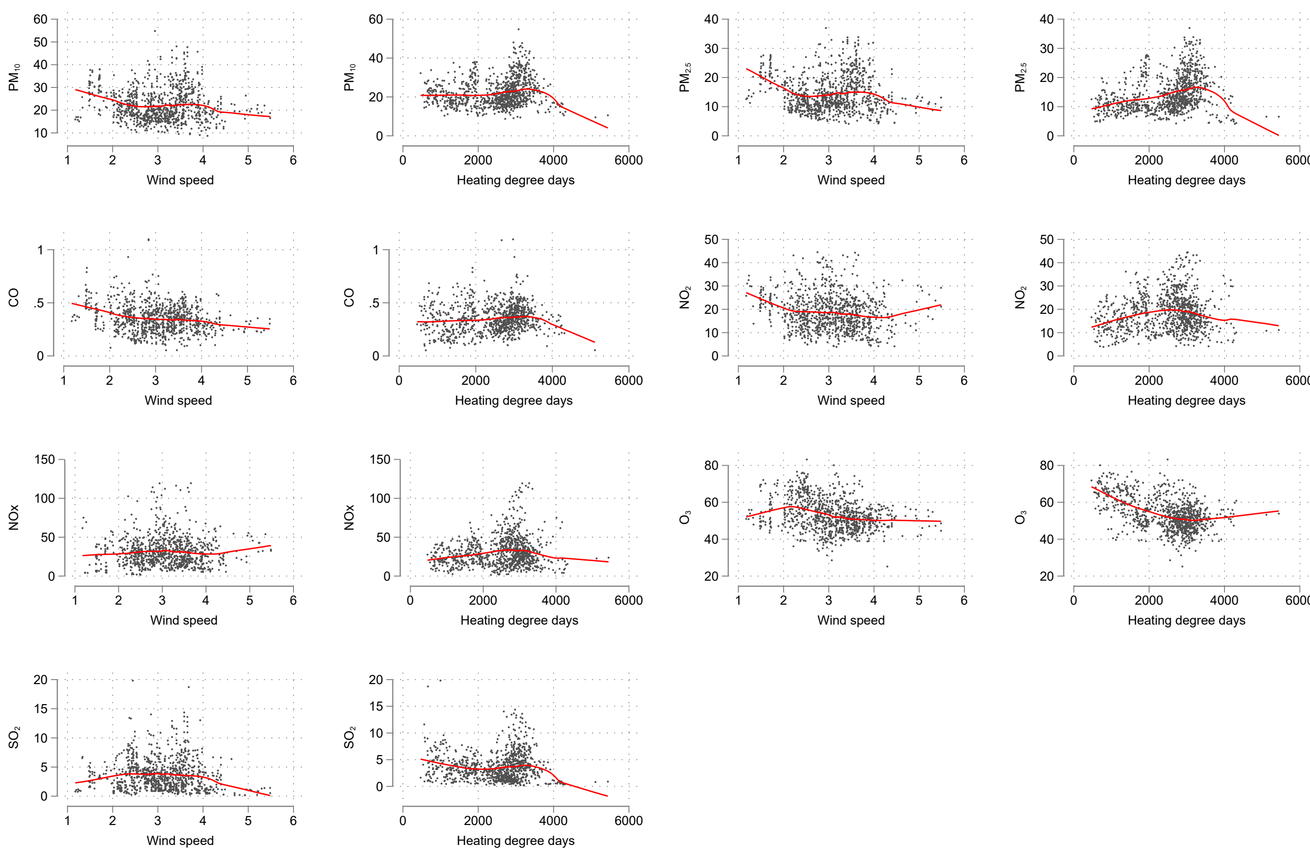


Figure A.7: Histograms of pollution concentrations with European health standard limits (2022)

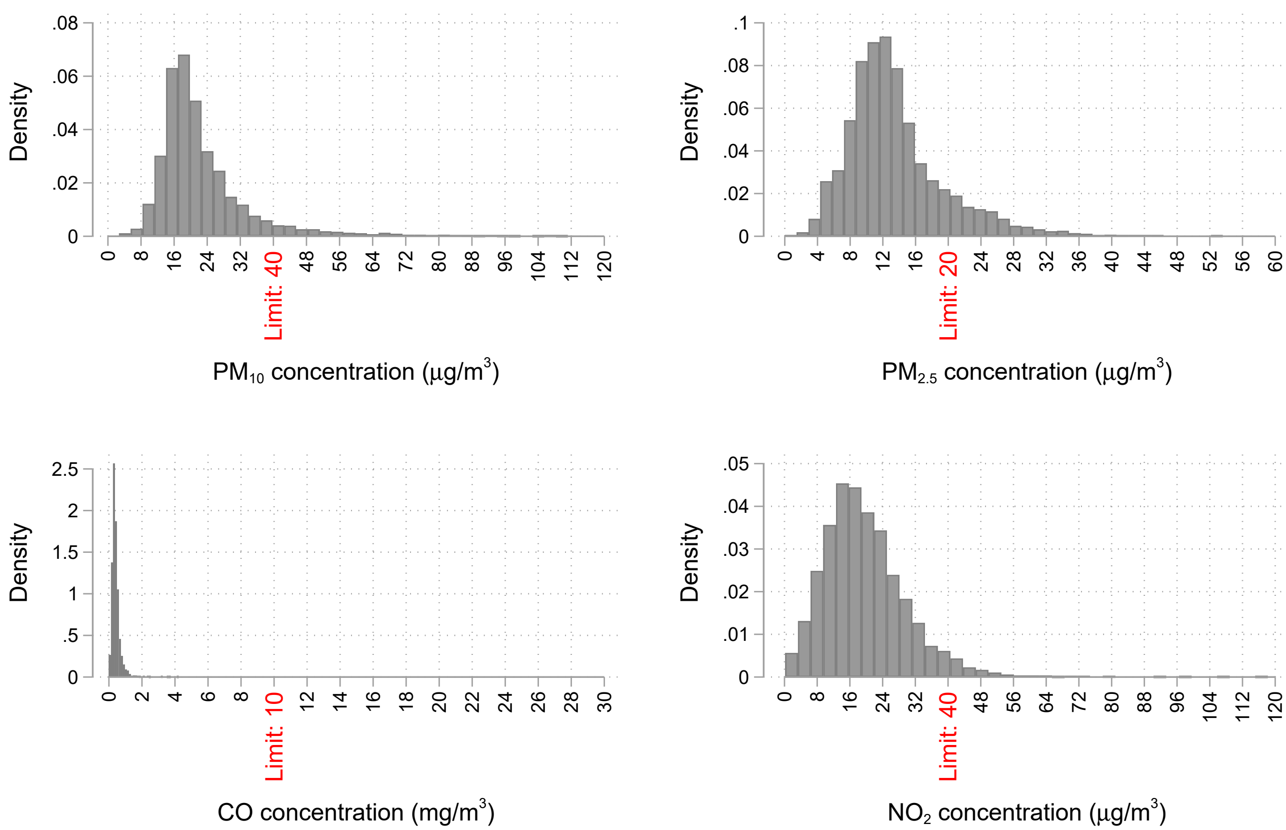


Figure A.8: Histograms of pollution concentrations with European health standard limits (2022)

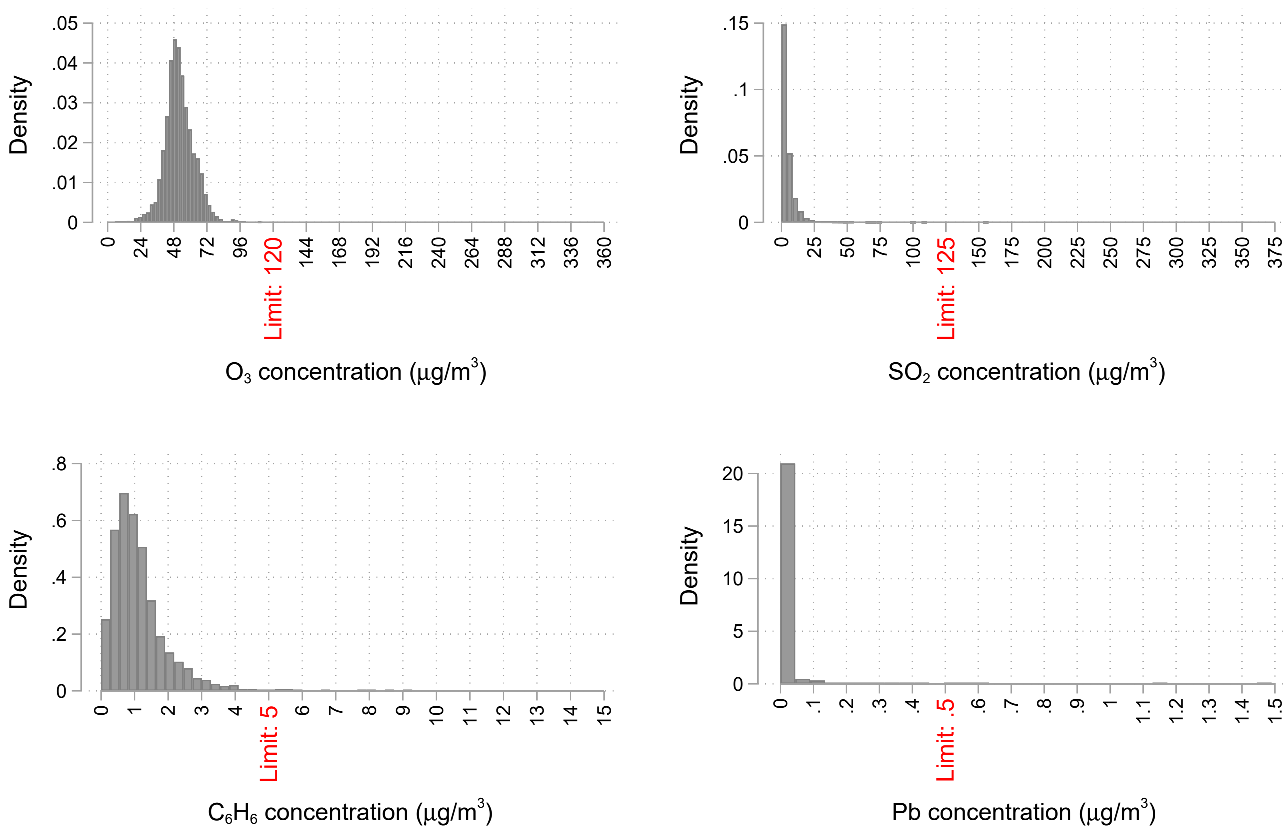


Figure A.9: The average $PM_{2.5}$ pollution in NUTS-3 regions

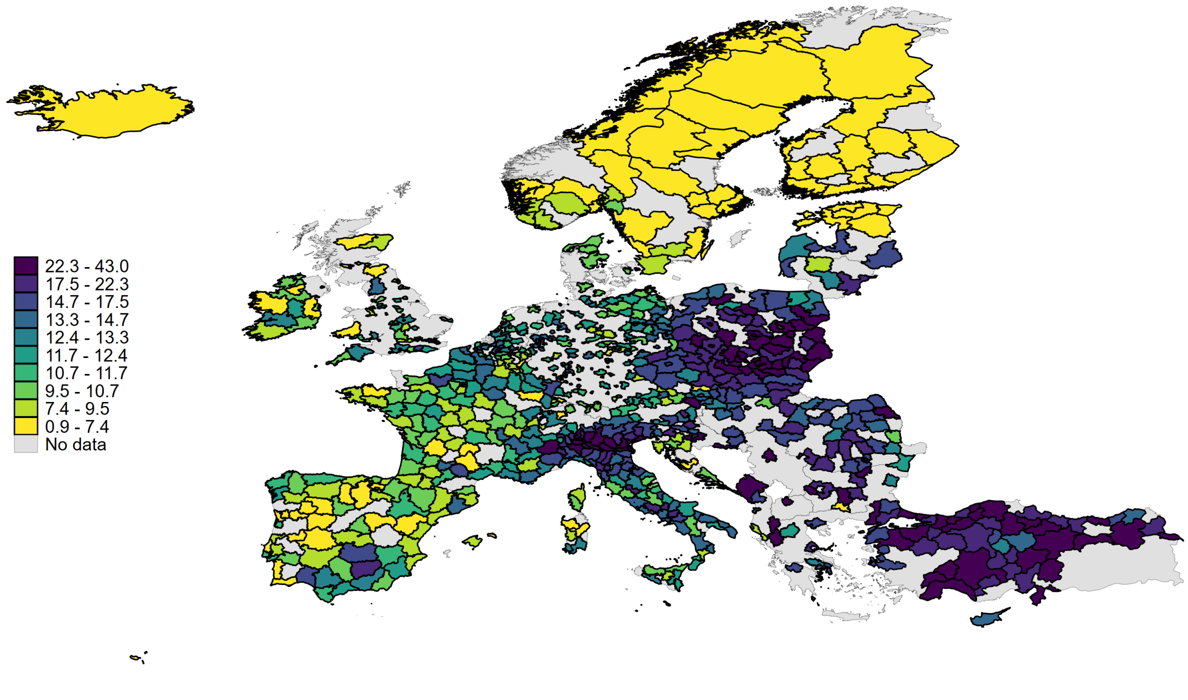


Figure A.10: The average CO pollution in NUTS-3 regions

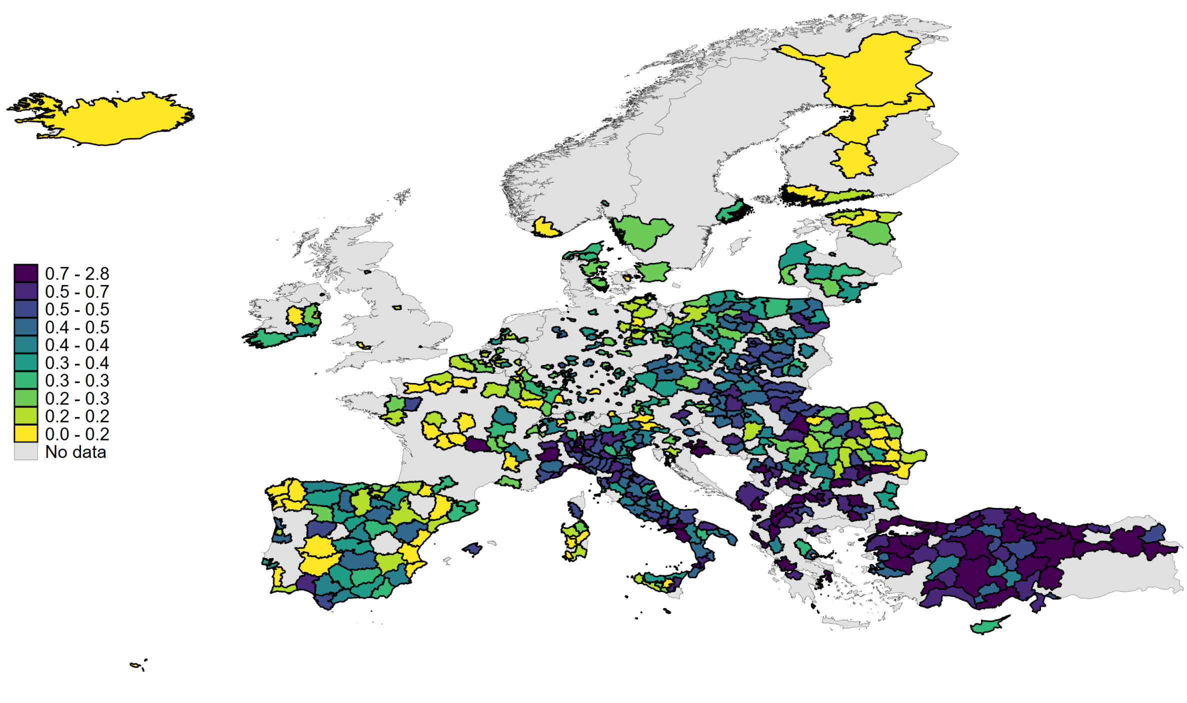


Figure A.11: The average NO_2 pollution in NUTS-3 regions

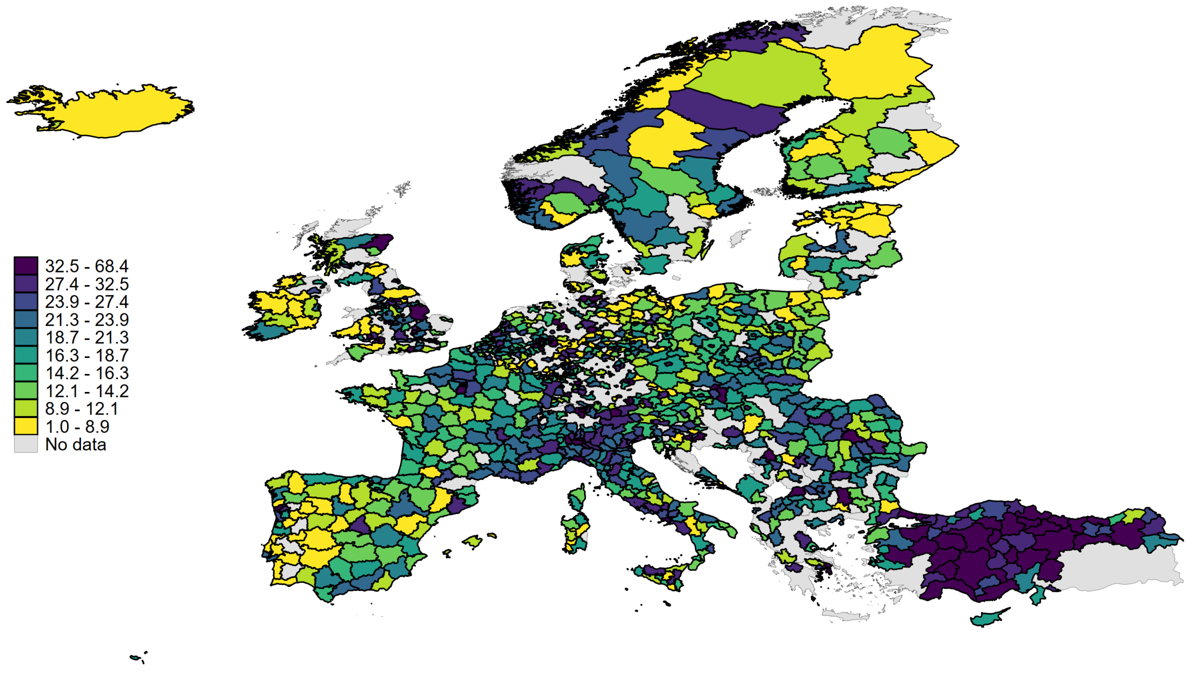


Figure A.12: The average NOx pollution in NUTS-3 regions

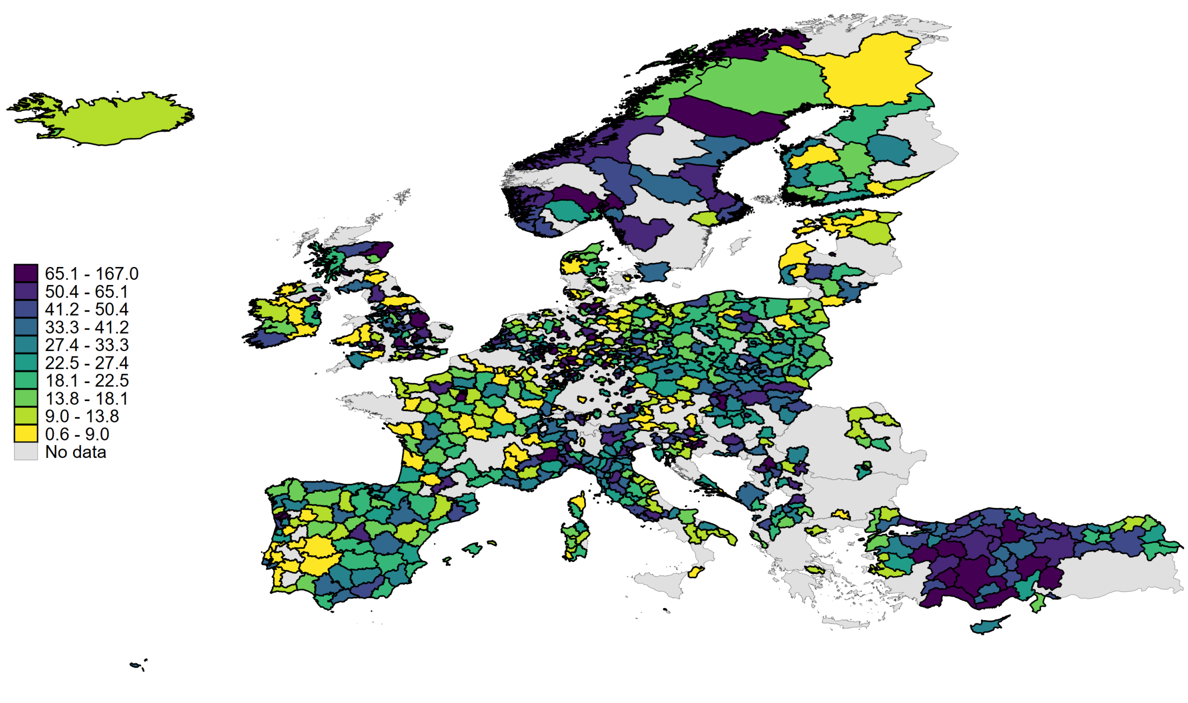


Figure A.13: The average O_3 pollution in NUTS-3 regions

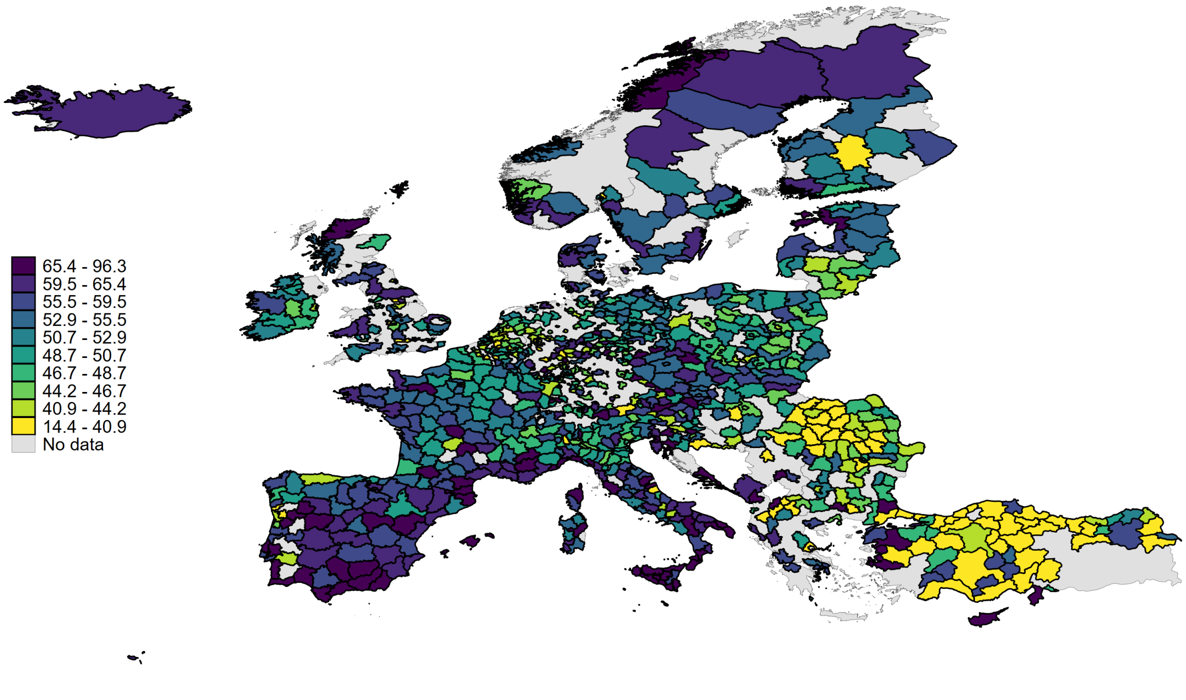


Figure A.14: The average SO_2 pollution in NUTS-3 regions

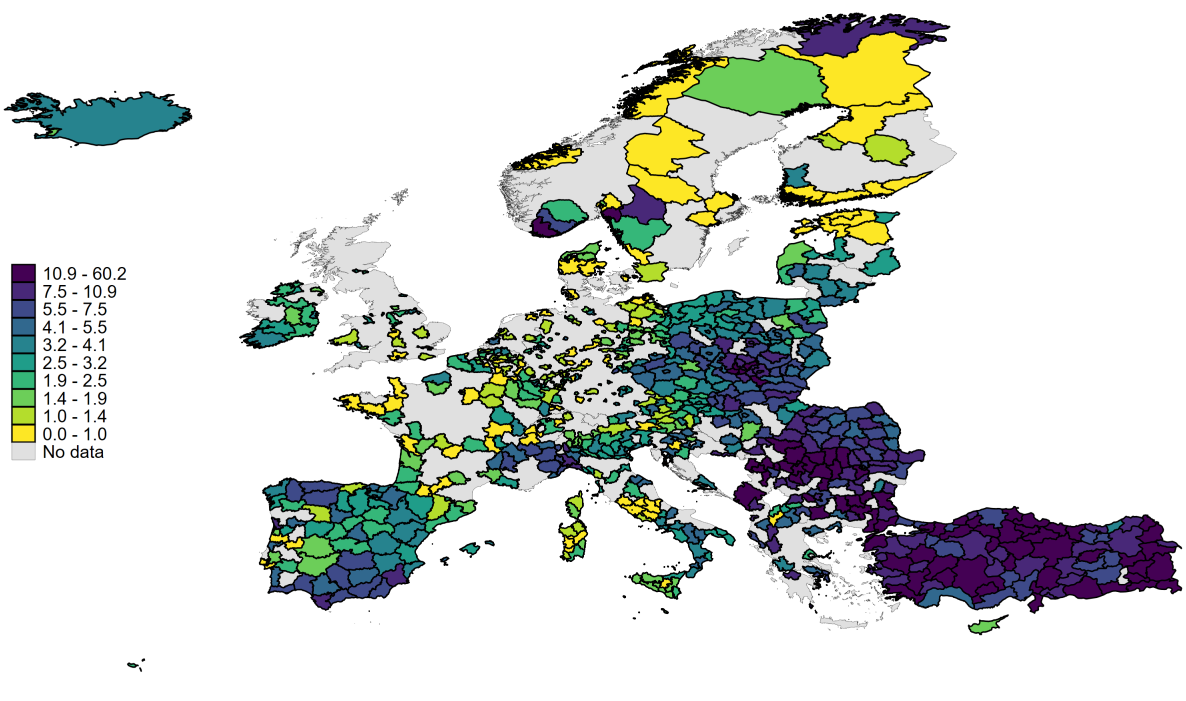


Figure A.15: The average Pb pollution in NUTS-3 regions

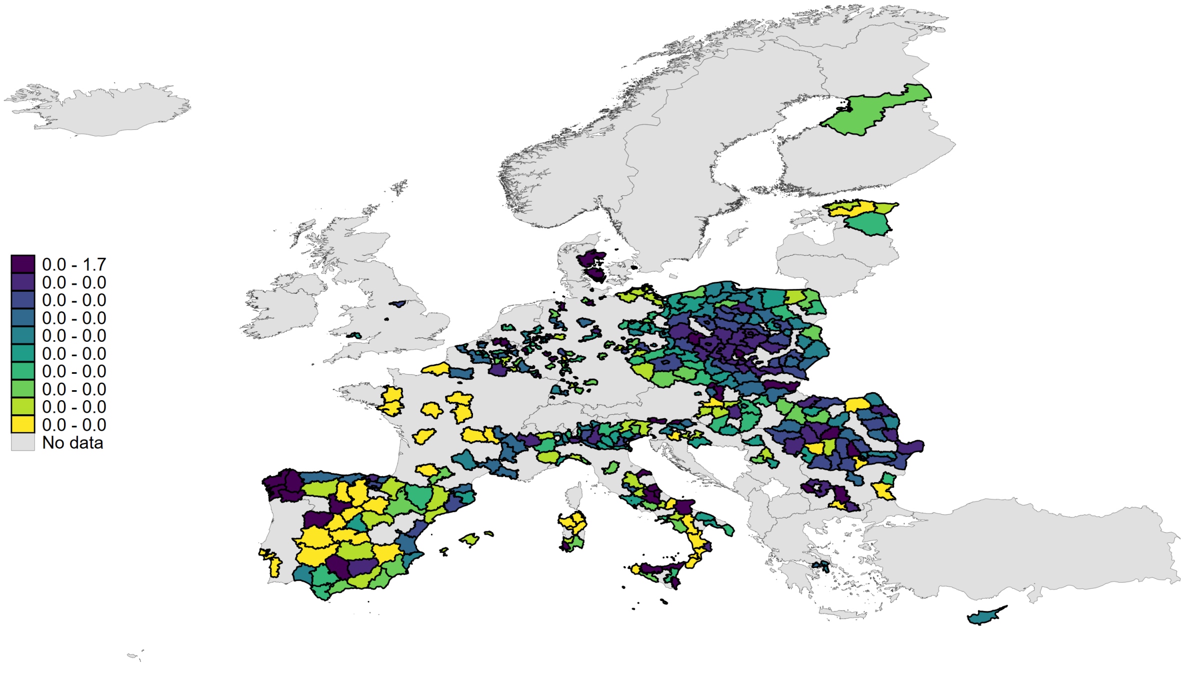


Figure A.16: The average C_6H_6 pollution in NUTS-3 regions

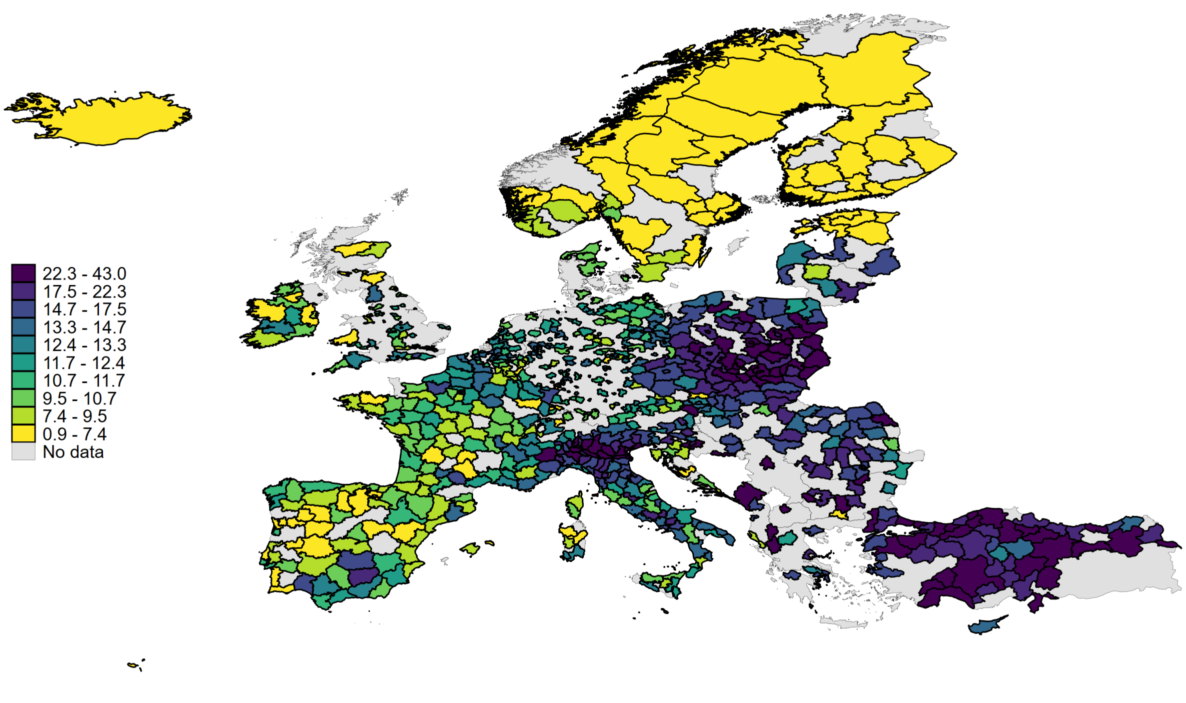


Figure A.17: The average NO pollution in NUTS-3 regions

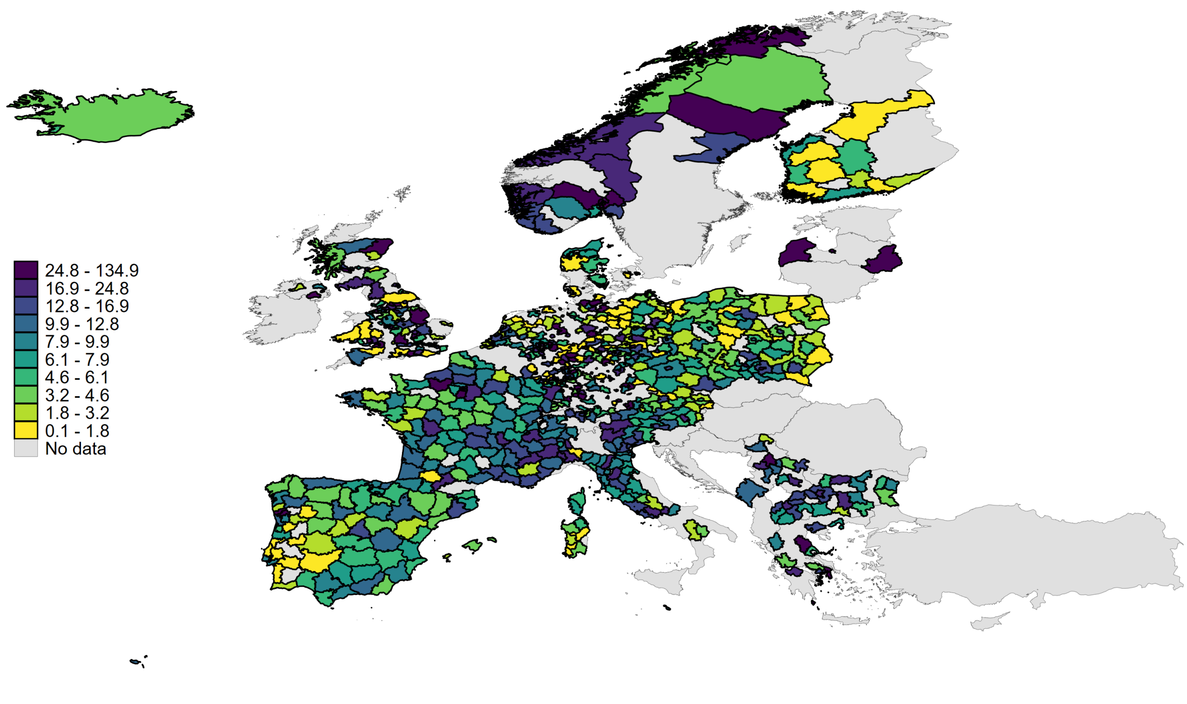


Figure A.18: NUTS-3 regions above and below median GDP