

Environmental justice: are socially vulnerable more likely to be exposed to air pollution? Evidence from Poland

Piotr M. Jóźwik (41443)

Abstract. According to previous literature, this is the first study that uses quantitative methods to analyse the correlation between air pollution and socioeconomic deprivation in Poland. Using ordinary least squares (OLS) model and cross-sectional data from 2018, gathered from the European Environment Agency and Statistics Poland, I arrive at two main results. Firstly, I observe that more impoverished populations are exposed to lower concentrations of nitrogen dioxide (NO_2) and oxides (NO_x), particulate matter with diameters smaller than $10\text{ }\mu\text{m}$ and $2.5\text{ }\mu\text{m}$ (PM_{10} and $\text{PM}_{2.5}$) and sulphur dioxide (SO_2). Under different specifications this correlation is statistically significant only for PM_{10} , $\text{PM}_{2.5}$ and SO_2 . Further quantile comparison suggests that more affluent and educated populations in Poland are exposed to lower concentrations of benzo(a)pyrene in PM_{10} (BenPM_{10}) than poorer and less educated populations. This relationship is reversed for other pollutants. Additionally, the relationship between socioeconomic deprivation and SO_2 concentration follows an inverse U-shape path, meaning that the middle part of the population is exposed to lower concentrations of SO_2 . Results of my analysis stay largely unchanged when using alternative methods of data aggregation. Finding that the more socioeconomically deprived ones are exposed to lower levels of pollution than the wealthier people is in contrast with most of previous environmental justice literature, which finds that on average impoverished populations are exposed to higher levels of pollution.

Keywords: environmental justice, social deprivation, air pollution, inequality

JEL: D63, I32, Q53, Q56

Supervisor: Pamela Campa
Date submitted: 18 May 2020
Date examined: 25 May 2020
Discussant: Daniel Heimgartner
Examiner: Elena Paltseva

Table of contents

List of Tables	3
List of Figures	3
1 Introduction	4
2 Background	6
2.1 Research motivation	6
2.2 Literature overview	7
3 Methodology	9
3.1 Data	10
3.2 Socioeconomic deprivation index	13
3.3 Model	14
3.4 Quantile comparison	15
4 Results	16
4.1 Ordinary least squares estimation	17
4.2 Quantile comparison	21
4.3 Robustness	24
4.3.1 Varying the aggregation radii	24
4.3.2 Exclusion of industrialized vovoidships	25
5 Caveats and areas for future research	28
5.1 Data collection process	29
5.2 Estimation of the damage function	29
5.3 Longitudinal analysis	30
5.4 External validity	30
6 Conclusions	31
7 References	32
A Appendix A	37
A.1 Pollutants description	37
A.2 Additional tables	38
A.3 Additional figures	43

List of Tables

1	Descriptive statistics (join-by-location)	12
2	Basic model (join-by-location)	18
3	Expanded model (join-by-location)	19
4	Expanded model excl. 4 largest cities (join-by-location)	20
5	Expanded model excl. 4 largest cities and high energy prod. vovoidships (join-by-location)	27
6	Expanded model for only high energy prod. vovoidship (join-by-location)	28
7	Descriptive statistics (2 km buffer)	38
8	Expanded model without SDI (join-by-location)	39
9	Basic model (2 km buffer)	40
10	Expanded model (2 km buffer)	41
11	Expanded model excl. 4 largest cities (2 km buffer)	42

List of Figures

1	Air pollution monitoring stations over Polish municipalities total population cartogram, darker colour indicates higher total population	11
2	Quantile graphs for socioeconomic deprivation and air pollution	22
3	Quantile graphs for relative gross salary and air pollution	23
4	Quantile graphs for gross enrolment ratio and air pollution	23
5	Quantile graphs for social benefits per 10k and air pollution	43
6	Quantile graphs for unemployment share and air pollution	44

1 Introduction

Environmental justice is usually referred to as a grass-root movement working towards a world *"[...] in which no population, due to policy or economic disempowerment, is forced to bear a disproportionate share of the negative human health or environmental impacts of pollution or environmental consequences resulting from industrial, municipal, and commercial operations or the execution of federal, state, local, and tribal programs and policies."* (Banzhaf et al., 2019a).

The United Nations Human Rights Council presents the right to health as a fundamental part of human rights and living in dignity. They have established healthy working and environmental conditions as a crucial part of the human right to health (UN, 2008). At the same time, World Health Organization reports that in 2016 24% of global deaths and 28% of deaths among children under five were due to modifiable environmental factors, with ischaemic heart disease, chronic respiratory diseases, cancers and unintentional injuries being the leading causes of death. This health burden is reported to be unequally distributed among world population, with lower income countries bearing its larger share (Prüss-Üstün et al., 2016). This aggravates the social burden for the population of those countries, since their population usually needs to handle both the burden of a dire economic situation of the country and the health burden of environmental damage.

Poland has one of the worst air quality indices in European Union (Ortiz, 2015). It has 33 out of 50 most polluted cities in the European Union (WHO, 2016). Additionally, from 2012 to 2017 it was one of the only European countries in which 100% of urban population have been exposed to pollutant concentrations (i.a. PM_{2.5} or benzo(a)pyrene) critically above EU standards (EEA, 2019). It is estimated that every year air pollution in Poland is causing around 43,000 premature deaths.

Economic literature reports impoverished groups of population to be exposed to disproportionately large shares of environmental pollution (Banzhaf et al., 2019a, 2019b; Mohai et al., 2009). Scholars list a variety of reasons supporting this claim, such as the inability of a destitute population to move out of increasingly polluted areas, prices of land being lower in areas with elevated levels of environmental pollution or companies targeting societies with low political power to ensure their operations will follow "path of least resistance", when sitting a new factory (Mohai & Bryant, 1991).

Polish income distribution across the last century followed a U-shaped evolution, with elevated levels of inequality before the Second World War, relatively low levels during the Polish People's Republic and rapidly rising levels after the transition from central planned economy to democratic capitalism in 1989. In just one generation system transformation managed to turn Poland from one of the most egalitarian European countries into one with the highest levels of income disparity, with top 10% of population raising its share of total national income from 20% in 1989 to almost 35% in 2015 (Bukowski & Novokmet, 2019).

The main research question of this paper is whether there is a correlation between socioeconomic deprivation index and levels of air pollution in Poland. To answer this question I exploit two data sets. For my dependant variables I use the European Environment Agency’s air quality data set which consists of 13,210 air pollution measurements across 270 unique monitoring stations in Poland. Independent variables used in my study are gathered from socioeconomic census data aggregated at Nomenclature of Territorial Units 3 (NUTS 3) level from the National Statistical Bureau of Poland (Statistics Poland). I use those variables to construct my synthetic deprivation index (SDI) and as controls for my ordinary least squares (OLS) model. I study the correlation between SDI and concentrations of benzo(a)pyrene in PM_{10} (Ben PM_{10}), nitrogen dioxide (NO_2) and oxides (NO_x), particulate matter with diameters smaller than 10 μm and 2.5 μm (PM_{10} and $PM_{2.5}$) and sulphur dioxide (SO_2).

The main results of my paper are twofold. Firstly, across all of my specifications, I observe a negative correlation between all of the studied pollutants and socioeconomic deprivation index in Poland, meaning that on average wealthier people are exposed to higher levels of hazardous air pollution than people from socioeconomically deprived groups. This relation is contrasting most of the environmental justice literature, which usually reports the inverse relation, in which poorer people are exposed to larger shares of air pollution. When controlling for total population, feminity ratio, municipality’s revenues, house prices and share of workers in agriculture, fishery and forestry only the correlations between SDI and PM_{10} , $PM_{2.5}$ and SO_2 are statistically significant.

Secondly, using quantile comparison method, I confirm that the relationships between socioeconomic deprivation index and air pollution are negative for all the pollutants except Ben PM_{10} . Further when I break up my synthetic deprivation index, I find that Ben PM_{10} exhibits the same relationships with unemployment and social benefits proxy as other pollutants, but an inverse relationships with income and education. Moreover, I find that SO_2 follows an inverse U-shape path, meaning that the middle part of the population is exposed to lower levels of sulphur dioxide concentration.

I vary my data aggregation methods by using a simple join-by-location method and a simplified distance-based method of data aggregation. First method matches the monitoring stations to the exact municipality they are situated in. The second one imposes a circular buffer around air quality monitoring stations, which I use to construct mean socioeconomic data of municipalities inside the buffer. For my main analysis I use buffers with 2 km radius. The results under different data aggregation method stay largely unchanged.

My paper contributes to environmental justice literature in three ways. According to my best knowledge, previous environmental justice literature only analysed or mentioned Poland in passing and this is the first study to quantitatively investigate the relationship between air pollution and

socioeconomic deprivation in Poland. More broadly, my research contributes to a growing, albeit still scarce, literature studying Eastern European and post-communist countries. Expanding this strain of literature is crucial since oftentimes extrapolating results of research from US or Western Europe can be misleading, given Eastern European countries can be socially and economically different from US or Western Europe (i.a. due to post-communist history).¹ Moreover, by exploiting homogeneous racial and ethnic structure of Poland, this paper contributes to growing amount of literature that tests the relationship between air pollution and socioeconomic variables that is not confounded by ethnicity or race.

I structure the remainder of this paper in a following way: section 2 presents the empirical motivation for the paper and a brief literature overview. In section 3 I describe the data and methodology used, while in section 4 I present the results of the analysis. I discuss caveats and limitations of my analysis in section 5 and I conclude in section 6.

2 Background

The central research question of this paper is whether there is a significant correlation between socioeconomic measures of deprivation and concentration of different air pollutants in Poland. Below I present the three empirical motivations that lead me to formulate this question: strong, negative impact of air pollution on health documented in previous literature, elevated levels of air pollution and high socioeconomic disparities in Poland. My additional motivation is the scarcity of environmental justice literature concerning Poland. I expand further on the last incentive in the subsection 2.2, which presents a brief literature overview.

2.1 Research motivation

The negative impact of air pollution on health is well documented, with literature reviews such as Šrám et al. (2005) pointing towards adverse effects of pollution on number of outcomes (i.a. considerably lower birth weights, shorter duration of gestation and higher reported infant mortality). Moreover, according to the fetal origin hypothesis, pollution can not only have short-term negative effects on an individual, but also long-term ones. Scholars report those effects to be usually mediated through early life and neonatal exposure to pollution. For example, a literature review by Currie et al. (2014) notes that in utero exposure to fallout can detrimentally impact math grades achieved by children in secondary education (Almond et al., 2009). Moreover, exposure to elevated levels of total suspended particles during pregnancy and early childhood are reported to have a significant negative

¹Environmental justice literature in this part of the world has been mostly focused on studying the Roma population in countries such as Hungary and Slovakia (Filcák & Škobla, 2019).

impact on annual salaries of adult workers (Isen et al., 2017). Further an increase in exposure to PM_{10} during pregnancy considerably increases the death rates of children under age of 15 (Janke et al., 2009).

Further Poland has one of the lowest levels of air quality among European Union countries, with air pollutants levels severely surpassing standards set by World Health Organization (Zieliński et al., 2018). Yet, social awareness of air pollution is limited, with almost 60% of citizens claiming that they are enjoying high level of air quality, while 80% of population is unable to name even one air pollutant. This lack of awareness can negatively impact the ability of population to handle the adverse effects of pollution (Smogowy, 2015). The National Centre for Emissions Management (KO-BiZE) reports that the largest contributor to elevated levels of air pollution in Poland is municipal or so called "low" emission, that is emission that is created in zone up to 40 m high. It is mainly attributed to household boilers and domestic furnaces. KOBiZE estimates that 61% of total carbon dioxide and almost 36% of the total suspended particles come from household boilers. This stems from lack of regulations regarding the quality, maintenance and use of said boilers (i.e. use of wood and trash as a fuel). It is estimated that 70% of those boilers are of low emissions standards quality (Kleczkowski, 2020).

Moreover, during Polish People Republic the income inequalities in Poland were amongst lowest in Europe, while shortly after the collapse of USSR and transition to democratic capitalism Poland has quickly became one of the most "unequal" European countries. Although real average national income per capita in Poland has doubled since the transformation, it did not benefit all social groups proportionally, due to capture of large parts of capital via privatization by top income percentiles of society (Bukowski & Novokmet, 2018). Income is not the only facet of inequality burdening Poland, as scholars report notable division along economic, social and environmental lines between different regions of Poland (Roszkowska & Karwowska, 2014). Literature points to three main roots of this situation. The first one, "starting" the main division, were the Partitions in 19th century. Afterwards, the centralism policy of command economy during Polish People Republic exacerbated the situation, which was solidified with rampant aggravation of inequalities during the system transformation in 90s (Czyż & Hauke, 2011).

In conclusion, this paper aims to investigate the relationship between the elevated levels of air pollution and significant income disparities among Polish citizens.

2.2 Literature overview

In recent years there have been a considerable growth of interest in economic inequality literature, which was largely facilitated by works such as Piketty and Saez (2003), which pioneered

the use of tax data over survey data in inequality modeling. The resulting growth in literature made it possible for Piketty's book *Capital in the Twenty-First Century* to become one of the greatest sales success ever of Harvard University Press. On the other hand, recent years brought more public attention to environmental issues, with climate protests in around 185 countries spurred by actions of activists such as Gretha Thundberg. This growth of interest in inequality economics coupled with growing concern for environment could probably be attributed as one of the driving factors behind the surge of interest in environmental justice.

Most of environmental justice literature focuses on injustices reported in United States. Although literature regarding other parts of the world has grown considerably in recent years, it is still scarce when compared to amount of studies conducted on data from United States (Laurent, 2011). This is especially evident if we focus on literature concerning Eastern Europe and post-communist countries such as Poland. While there have been only a few studies investigating environmental justice in these countries, most of them are descriptive case studies of instances of injustices. For example, studies portraying cases of Roma population exposure to high amounts of toxic emissions in Hungary and Slovakia (Filcák & Škobla, 2019; Harper et al., 2009; Varró et al., 2001). Scholars point to this lack of research as unfortunate, portraying this region of Europe as an excellent model for environmental justice studies. On the other hand, they suggest that sometimes publication of environmental justice studies concerning this region can be blocked, due to political nature of environmental issues (Varga et al., 2002).

The closest previously conducted study to my paper is Choi et al. (2016) studying environmental injustices in South Korea by analyzing the correlation between deprivation index and synthetic air quality index. Using ordinary least square modeling, they report a negative correlation between air quality index and several socioeconomic variables, both at individual and area level. In my thesis I also model the relationship of socioeconomic variables and pollution, although rather than using a synthetic air quality index I focus on modelling the relationship separately for each pollutant. I do so in order to account for the possibility of differing correlations between socioeconomic variables and pollution. Further because I analyse each pollutant separately the interpretation of my results is more direct.

According to my best knowledge the only previous studies that have analysed issues connected to environmental justice in the context of Poland have been Paloniemi et al. (2015), Giardullo et al. (2016) and Ferreira et al. (2013). First of the studies focuses on public participation in biodiversity governance after the designation of a Natura 2000 program in Poland, Finland, Greece and United Kingdom. The second one investigates social constraints influencing air quality governance in Warsaw, Milan, Antwerp and Malmö. Third one links SO₂ concentrations with subjective well-being in Europe, and using hybrid subjective well-being function finds a significant negative effect of SO₂ on

the self-reported well-being of an individual.² Apart from the studies by Giardullo et al. (2016) and Paloniemi et al. (2015), the study by Ferreira et al. (2013) only uses data from Poland, but never explicitly analyses the country's environmental injustices. My paper contributes to this strand of literature in two ways. Firstly, I expand the scope of analysed environmental injustices by including socioeconomic variables on NUTS 3 level, which enables me to observe the relationships at a more granular level. Secondly, I am the first one to explore correlation of air quality and social deprivation in Poland with the use of quantitative methods.

Typically initial studies of environmental justice have been focused on treatment, storage and disposal facilities (TSDFs) or toxic release inventories (TRIs). For example, Anderton et al. (1994) study reports TSDFs in US are more likely to be located in census tracts with Hispanic groups, especially in ones with a large share of Hispanics. Further Ringquist (1997) reports minority communities in US to be disproportionately exposed both to TRIs and pollution from them. More broadly, this paper contributes to the growing part of environmental justice literature that studies air pollution and its relation to race and income, such as Wang and Fang (2016) who observe a positive correlation between concentration of PM_{2.5} with urbanization and a negative one with GDP per capita in China. Another example could be Mikati et al. (2018) study which finds, even after sensitivity analysis, both people in poverty and non-Whites to bear significantly larger burden of particulate matter emissions in US.

In environmental justice literature the first-order relation between air pollution, race and income of individuals is widely agreed upon, although there is no definite consensus on that correlation if we control for either race or income (Banzhaf et al., 2019a; Mohai et al., 2009; Mohai & Saha, 2015). My paper contributes to this part of literature by exploiting the unique racial and ethnic composition of Poland, which makes it possible for me to observe the relation between social deprivation and air pollution that is not confounded by racial or ethnic population structure.

Concluding, my paper contributes to previous literature in following ways: it expands the existing environmental justice literature in context of Eastern Europe, particularly Poland and it contributes to growing amount of research on relation between social deprivation and air pollution, more specifically to one not confounded by race or ethnicity.

3 Methodology

In this section I first present my data sources and data aggregation methods I used. Later, I present the construction of my socioeconomic deprivation index (SDI) and describe the model used in the analysis. In the last subsection I justify the use of quantile comparison in my analysis.

²It is the only study I found that uses the same data set as I do.

3.1 Data

I collect data on air pollution from the air quality database of European Environment Agency (EEA). It features 13,210 air pollution measurements across 270 unique monitoring stations in Poland from 2018. All of the information in the database are reported by Chief Inspectorate of Environmental Protection in Poland and verified by EEA. In my analysis I only use records that have been classified as "valid" by EEA. The agency defines "valid" records as the ones for which measurements cover at least 75% of the analysed time period. I further restrict this sample to only include data on annual mean concentrations of six pollutants of interest. My main criterion for this restriction was to achieve data samples for each pollutant of at least 90 observations. The pollutants analysed are: benzo(a)pyrene in PM₁₀ (BenPM₁₀), nitrogen dioxide (NO₂) and oxides (NO_x), particulate matter with diameters smaller than 10 μm and 2.5 μm (PM₁₀ and PM_{2.5}) and sulphur dioxide (SO₂).³

Each of the records in the EEA database is tied to one of the monitoring stations. Each of the monitoring stations is described with precise longitude and latitude spatial data. Moreover, I obtain the official shapefiles (data set storing geometric location of given polygons) of all Polish administrative regions from Polish Head Office of Geodesy and Cartography (GUGiK).

I gather the socioeconomic variables from National Statistical Bureau of Poland (Statistics Poland), which is the main provider of statistical data in the country. Except for the energy production variable, all of the data is aggregated at Nomenclature of Territorial Units 3 (NUTS 3) level and references whole 2018. In Poland NUTS 3 level is associated with the penultimate level of administration and consists of 380 municipalities called powiaty, of which 66 are cities with municipality rights. The energy generation data is aggregated at the NUTS 2 level, which in Poland represents 16 administrative regions called województwa or vovoidships.

In Figure 1 I present monitoring stations over a cartogram of total population for Polish NUTS 3 level municipalities. The darker color of municipality indicates higher total population. Although the dispersion of the monitoring stations across the country is unequal, the most populated regions of the country are included in the analysis, while the municipalities that are left out are the ones with the lowest population.

To combine those data sets I use two methods of data aggregation. First of them is a simple join-by-location method, while the second is a distance-based areal containment method based on Mohai and Saha (2015) and Anderton et al. (1994).

In the first method I match each of the monitoring station in the database with the exact administrative regions in which they are located, which allows me to tie them to the socioeconomic data for respective municipalities. In the second method I create a buffer of approximately⁴ 2 km around

³I briefly describe each of those pollutants in Appendix A.1. Moreover, I discuss data limitations in subsection 5.1.

⁴I use decimal degrees in order to create the buffers, meaning that depending on the longitude and latitude of the

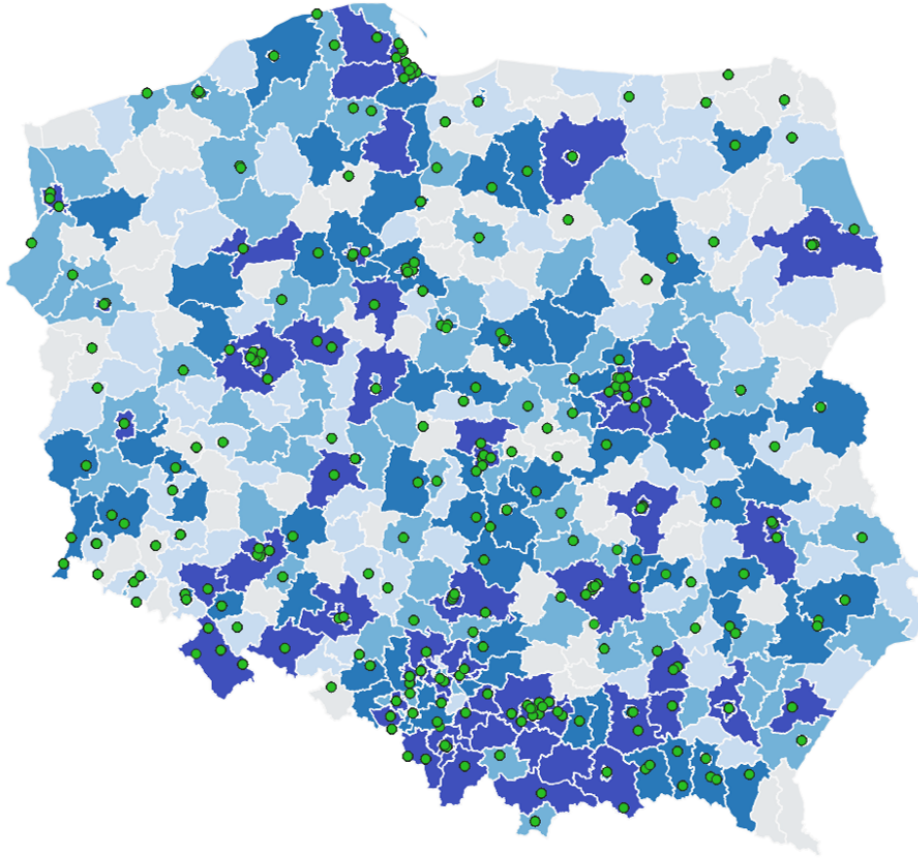


Figure 1: Air pollution monitoring stations over Polish municipalities total population cartogram, darker colour indicates higher total population

each of the monitoring stations and mark each of the administrative regions included inside of the buffer. Afterwards, I average out the socioeconomic variables across all of the marked municipalities to obtain the simplified mean socioeconomic data of population around the monitoring station. I employ second method of aggregation in order to account for the possibility of the measurement station being located far from the centroid of the respective administrative region. For example, some of the stations are placed in a close proximity to borders of multiple municipalities, meaning that the air pollution they register transgresses the municipal borders and affects different populations. In this situation using data aggregated with the first method could induce a measurement error in my analysis due to selection miss match. For both of those methods I use QGIS 3.12.⁵

Below I present descriptive statistics with respect to each pollutant. I choose to present the measurement station the buffers sizes can vary. Given relatively "small" geographical extent of Poland this variance is small.

⁵I discuss varying the size of buffer radii in subsection 4.3.1.

statistics for each of the pollutants separately, since I also separate the pollutants when calculating my socioeconomic deprivation index. This data has been collected with the use of simple join-by-location method.⁶ I report means and standard deviations in parentheses.

The main takeaway from the table is the sizeable standard deviation for total population, which is caused by high difference in the sizes of municipalities. In Poland, large cities can be given the rights of municipality, which effectively means that the largest cities in Poland like Warsaw and Kraków are having a disproportional impact on the sample means.⁷ I address this characteristic of my sample in section 4.1.

Table 1: Descriptive statistics (join-by-location)

	BenPM ₁₀	NO ₂	NO _x	PM ₁₀	PM _{2.5}	SO ₂
Concentration	4.32 (2.76)	17.95 (9.20)	30.06 (26.93)	32.29 (6.87)	23.57 (5.07)	5.21 (2.74)
Rel. gross sal.	88.56 (13.29)	93.19 (13.89)	93.26 (13.90)	90.55 (13.56)	94.29 (14.24)	92.11 (13.50)
Gross enrol. ratio	105.00 (12.02)	105.72 (11.72)	105.72 (11.78)	105.29 (11.78)	109.46 (12.00)	105.06 (11.86)
Soc. ben. per 10k	525.40 (221.13)	475.57 (233.06)	475.16 (233.03)	499.38 (226.86)	448.63 (210.84)	479.15 (229.16)
Unemp. share	4.29 (1.90)	3.99 (1.89)	4.01 (1.91)	4.09 (1.90)	4.03 (1.91)	4.01 (1.89)
Total pop.	190.60 (273.25)	231.47 (293.75)	233.43 (294.30)	214.17 (295.24)	275.58 (369.53)	204.46 (252.22)
Feminity ratio	107.60 (4.79)	108.61 (4.74)	108.65 (4.74)	108.00 (4.81)	110.18 (4.56)	108.26 (4.67)
Mun. rev. pc	5457.36 (1145.96)	5659.93 (1222.52)	5663.93 (1225.59)	5500.29 (1197.90)	5971.80 (1270.43)	5596.57 (1188.86)
House price	3692.11 (1339.55)	3958.34 (1601.81)	3957.43 (1610.66)	3787.33 (1498.46)	4090.33 (1511.20)	3842.98 (1564.28)
Agri. share	0.19 (0.18)	0.14 (0.17)	0.14 (0.17)	0.17 (0.18)	0.10 (0.14)	0.15 (0.17)
Observations	118	138	137	221	95	129

Standard deviations reported in parentheses. Rel. gross sal. is relative gross salary in relation to national average. Gross enrolment ratio is a percentage of population enrolled in the middle school education in the age group corresponding to this level of education, so given adult education it can be over 100. Soc. ben. per 10k is number of individuals receiving social benefits per 10,000. Mun. rev. pc is municipality revenue per capita. House price is median for 1 m² of residential premises. Agriculture share is share of workers in agriculture, fishery and forestry in total working population. All the pollutants except BenPM₁₀ are reported in µg in m³, while BenPM₁₀ is reported in ng in m³.

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

⁶The descriptive statistics for 2 km buffer method are presented in the Appendix A.2.

⁷For example Warsaw total population exceeds the sample mean by 5 standard deviations.

3.2 Socioeconomic deprivation index

I construct my SDI with principal component analysis (PCA) of four variables of interest. That way I am able to compact the impact of the variables of choice without introducing co-linearity in my analysis. I base my method and selection of variables on the previous works of Li et al. (2018) and Choi et al. (2016), with the main idea behind the selection of variables for the SDI being that they should be good proxies of health risks associated with hazardous air pollution. This is to ensure that SDI is not only going to indicate impoverished population but also one that is going to potentially be less able to cope with the side effects of pollution (i.a. due to lower base health stock or knowledge about means of protection). For the final components used in my model I follow the approach of Li et al. (2018) and use components with eigenvalues larger than 1.

I choose variables representing two facets of socioeconomic deprivation: economic and educational one. For economic variables I pick average gross salary in relation to national average, share of unemployed in total working age population and number of individuals receiving social benefits per 10,000. For educational proxy I choose gross enrolment ratio for middle school education.

Another variable commonly included in creation of socioeconomic indices is a variable representing racial or ethnic makeup of an analysed community (Mohai & Bryant, 1991). I do not include such variable in my model, since I argue for negligible impact of racial and ethnic structure on my analysis. According to 2011 Polish national census 95% of population declares homogeneous Polish nationality and only 1.5% of population declares a non-Polish nationality (Barwiński, 2014). This makes Poland one of the most homogeneous countries in Europe (Vasileva, 2012). Thus, I assume that race and ethnicity should not have a significant impact on the results of my study (at least at the level of municipal analysis).

I model my index so that that the weights associated with average salary and gross enrolment ratio have the same sign, given high positive correlation between those variables. Similarly, I model the weights attributed to share of unemployed population and amount of individuals receiving social benefits to share the same sign. This facilitates the interpretation of my SDI: with the rise in assumed deprivation of the municipality I observe a rise in SDI score. I use separate principal component analysis for each of six pollutants of choice, since I want to account for the possibility of each pollutant exhibiting a different relationships with SDI and my sample does not have an equal amount of monitoring stations for each of the pollutants.

To create the final index I chose components with eigenvalues larger than 1, which leave me with one independently estimated SDI for each pollutant. Each of those indices has Kaiser-Meyer-Olkin measure of sampling adequacy between 0.64 and 0.72, which according to (Kaiser, 1974) is a sign of “mediocre/middling” indices, which justifies using the indices in analysis, as opposed to

including each of the variables separately into the model.

3.3 Model

To estimate the correlation between socioeconomic deprivation and air pollution in Poland I employ the following equation:

$$Y_{ij} = \beta_0 + \beta_1 SDI_i + \epsilon_i \quad (1)$$

In which Y_{ij} is the annual mean concentration of pollutant i measured in station j , SDI_i is the socioeconomic deprivation index for pollutant i and ϵ_i is the standard error term.

I estimate all models with ordinary least squares model (OLS). In all specifications I use heteroskedastic-consistent (robust) standard errors. I do not employ standard error clustering since clustering the errors at NUTS 3 levels results in creating almost exclusively 1 observation clusters. Clustering at NUTS 2 level results in only 16 clusters, which is too few to justify the use of clustering (Angrist & Pischke, 2008).

In econometric modeling there is usually a possibility of omitted variable bias (OVB) affecting the regressions in question. Oftentimes, in order to deal with OVB the models are expanded with additional variables. Nevertheless, in my model there is a high probability of including “bad” controls, meaning controls which levels can be affected by SDI. This stems from the fact that usually in environmental justice research it is difficult to observe a completely exogenous variation (Hsiang et al., 2019). Thus, instead of focusing on controls, I expand my model with potential channels through which I expect SDI and pollution to interact with each other.⁸ I expand the model with five different variables, all of which are aggregated at NUTS 3 level. Those are: total population, feminity ratio, municipality’s revenue per capita, median price for 1 m² of residential premises and share of population working in agriculture, fishery and forestry in total working population.

Although all of the variables used to create the SDI are normalized I add total population to the equation (1) in order to account for high variance of population between the municipalities. Further I include feminity ratio in the model too observe the possible impact of heterogeneous gender structure of municipalities. Scholars report the migration in Poland to be largely “feminized”, meaning that women tend to migrate more often and stay for longer periods of time (Grabowska-Lusińska & Jaźwińska-Motyłska, 2013; Pawlak, 2016). In line with the theory, feminity ratio displays high correlation with relative average salary and air pollution values (the sign of the correlation signals that rural regions on average tend to have lower levels of feminity ratio than the urban ones).

⁸This is a standard procedure in environmental justice literature. Technically I am still expanding the equations with “control” variables, but my interpretation of those variables is that, given their significance/insignificance I explore potential channels of correlation for pollution, instead of a “causal” effect.

I add municipality's revenue per capita to the model to observe the impact of the affluence of municipalities. I expect that this affluence is not fully captured by relative average salary due to people working and living in different municipalities and municipalities having various sources of funding. I presume municipality's revenues to be impacted by SDI through taxes, but also to impact the SDI (i.a. through educational spending). At the same time those revenues could affect the air pollution measurements through municipal spending on activities connected to environmental protection.

I use median price for 1 m² of residential premises to observe the possibility of house prices affecting the sitting decisions of companies and individuals. The effect on individual is the straightforward impact on the price of living in the municipality. For the companies, I assume residential premises prices to be positively correlated with general land prices, which I expect to have an impact on the company decision process when sitting a new plant or factory, which have a direct impact on air pollution levels recorded by the monitoring stations in the municipality.

In order to account for the impact of agriculture on pollution and education, I include in my model the share of workers in agriculture, fishery and forestry in total working population (referred to as agri. share). Poczta and Bartkowiak (2012) reports a high differentiation of agricultural development in Poland, which depending on the quality of the methods used could have a varying impact on the air pollution (mainly through methods of soil management, which is the largest anthropogenic source of nitrogen oxide pollution). At the same time, this variable could impact the amount of people enrolled in middle education, by requiring children to support their parents during the harvests.

These potential channels of interaction between the SDI and pollution should expand the analysis by further explaining what can be the driving factors behind the observed correlations. Moreover, it is worth to note that some of those channels can showcase reverse causality. This can be true especially in the case of median price for 1 m² of residential premises. For example, literature reports that US residential prices can be negatively affected by visual disamenities connected to shale gas extraction (Muehlenbachs et al., 2015).

3.4 Quantile comparison

Occurrence of spatial auto-correlation affects the distribution and thus the independence of the observations, which can mean that the residuals of the estimation are spatially auto-correlated, which biases the standard errors of models such as OLS (Chun et al., 2012).

Spatial auto-correlation is a concept referring to a situation when geographically referenced observations are statistically correlated with other observations in their close proximity. Usually

positive spatial auto-correlation is interpreted as the tendency of high values and low values to cluster with values similar to themselves (so high-high and low-low clusters). The negative spatial auto-correlation is the reverse scenario, in which dissimilar values tend to cluster together (low-high and high-low clusters).

One of the measures used to test for spatial auto-correlation of data is Moran's Global I, originally developed by Moran (1948). Statistically significant positive values of Moran's Global I indicate a positive spatial auto-correlation of given spatial variables, while negative values indicate a negative spatial auto-correlation.

Given I am analysing data on pollution there is a high chance of positive auto-correlation affecting my sample. Conducting Moran's Global I test I confirm those expectations with high significance levels. Nevertheless, Shapiro-Wilk tests for normality of residuals distribution of my models suggest some of the residuals not to follow the normal distribution, which could bias the estimates of significance for Moran's Global I (Chun et al., 2012; Cliff & Ord, 1981). Moreover, the specification of Moran's test is highly subjective, with the need to properly conceptualize the spatial correlation of variables in question.⁹ This is best seen when I use two different weight matrices to conduct the test, exponential and binary. Both of the approaches report statistically significant, yet contradictory conclusions about the existence of spatial auto-correlation.

Acknowledging this problem I decide to expand my analysis without making additional assumptions regarding the unobserved spatial relationships of my variables. To do so I use quantile comparison method. It is a relatively simple and robust method used in environmental justice literature in situations when standard modeling of the relationships is not efficient enough (Hsiang et al., 2019; Li et al., 2018; Mitchell & Dorling, 2003; Pearce et al., 2006).¹⁰ At the same time it enables me to observe whether the relationships I observe follow a linear trend. The method consists of graphical comparison of the quantile distribution of dependent and independent variables of interest.

4 Results

In this section I present results of my empirical analysis. First, I go over the results from estimating the relationship between air pollution and socioeconomic deprivation index with the use of OLS modeling. Afterwards, I present results of the quantile comparison method and summarize the results.

⁹This affects the choice of weight matrix and distance bands used in the test.

¹⁰For example due to low explanatory power of OLS model.

4.1 Ordinary least squares estimation

In this subsection I present the OLS results using data aggregated with simple join-by-location method. I present the corresponding 2 km buffer results in the Appendix A.2. The main coefficient of interest in each of the tables in this subsection estimates the impact of the PCA weighted socioeconomic deprivation index calculated separately for each pollutant on the air pollution. Negative coefficients indicate that wealthier groups of population are exposed to higher shares of air pollution, while positive coefficient indicates the inverse.

When estimating my OLS models using separate variables which construct SDI instead of the SDI in most cases I do not find any of the variables except relative average salary to be significant. This hints at income being the driving force behind the observed correlations, while other variables having an auxiliary effect.¹¹

In Table 2 I present the results from the basic regression using data gathered with the simple join-by-location aggregation method. The main takeaway is the negative correlation between all pollutants and SDI except for BenPM₁₀. The coefficients are significant only for NO₂, NO_x and PM₁₀. The interpretation of this relation is that the wealthier parts of the population are exposed to higher levels of hazardous air pollution with the strongest relationship being displayed for NO_x and relation for PM₁₀ being rather small (-0.821 is about 11% of PM₁₀ standard deviation). The insignificant estimate for BenPM₁₀ precludes me from drawing any conclusions about the relationship between SDI and concentration of BenPM₁₀ at this point.

In my analysis I am investigating the correlation between the variables rather than a causal effect, thus I am not focusing my analysis on the size of each of the coefficients.¹² Nevertheless, the SDI values for all pollutants are normally distributed in range from -4 to 5. Simple interpretation for the coefficient for NO_x would be that on average one point increase in SDI of the municipality results in 7.367 decrease in NO_x concentration observed at the monitoring station. In terms of standard deviations it means that a 0.7 standard deviation increase in SDI results in 0.8 standard deviation increase in NO_x concentration.

The results of this model when using data aggregated with 2 km buffer method are largely the same. They report larger coefficients with smaller standard errors, but there is no change in terms of significance levels. For both aggregation methods all of the estimations suffer from relatively low adj. R² scores, with BenPM₁₀ equation displaying a negative score. Bearing in mind the simplicity of

¹¹I present the results for expanded model without SDI in Appendix A.2.

¹²Usually, environmental justice literature focuses on correlations. I follow this general trend for two reasons. Firstly, limitations of available Polish socioeconomic data make it impossible for me to come up with an exogenous variation in my equations. Secondly, as my paper is the first one to quantitatively explore relation of pollution and socioeconomic deprivation in Poland I focus on more "descriptive" modeling, which in future years, with additional data, could be expanded into the casual investigations.

Table 2: Basic model (join-by-location)

	BenPM ₁₀	NO ₂	NO _x	PM ₁₀	PM _{2.5}	SO ₂
	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE
SDI BenPM ₁₀	0.011 (0.17) [0.14]					
SDI NO ₂		-3.320*** (0.42)*** [0.48]***				
SDI NO _x			-7.367*** (1.35)*** [1.61]***			
SDI PM ₁₀				-0.821** (0.30)** [0.31]**		
SDI PM _{2.5}					-0.514 (0.35) [0.33]	
SDI SO ₂						-0.204 (0.16) [0.14]
Constant	4.325*** (0.25)*** [0.25]***	17.948*** (0.65)*** [0.65]***	30.062*** (2.09)*** [2.09]***	32.287*** (0.46)*** [0.46]***	23.567*** (0.52)*** [0.52]***	5.207*** (0.24)*** [0.24]***
Observations	118	138	137	221	95	129
Adjusted R^2	-0.009	0.307	0.175	0.029	0.012	0.005

Standard errors reported in parenthesis and robust standard errors reported in brackets.

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

the model I expand the model with a set of variables in order to further investigate the relationship between socioeconomic deprivation index and air pollution.

Table 3 presents the results from estimating the basic model expanded with controls for total population, feminity ratio, municipality's revenue per capita, median price for 1 m² of residential premises and share of population working in agriculture, fishery and forestry in total working population.

The main takeaway after expanding the model is the dampening and loss of significance for NO₂ and NO_x coefficients. In contrast, I observe a significant coefficient for previously insignificant SO₂ and a change in sign for BenPM₁₀ coupled with positive adj. R^2 for this pollutant. Most of the controls in the model are not significant and in the case they are significant they have a negligible impact on the dependent variable. The only significant control with large coefficient is share of agricultural workers for NO₂ and NO_x. This control also displays large, but not significant coefficients for all pollutants.

I expect the coefficient for agri. share to be negative assuming it is a good proxy of rural regions.¹³ In Poland approximately 78% of total emissions of nitrogen oxide is agricultural soil

¹³Agri. share has a high negative correlation with feminity ratio, municipality revenues and town status meaning that high values of it should represent more "rural" municipalities.

Table 3: Expanded model (join-by-location)

	BenPM ₁₀	NO ₂	NO _x	PM ₁₀	PM _{2.5}	SO ₂
	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE
SDI BenPM ₁₀	-0.508 (0.29) [0.28]					
SDI NO ₂		-0.957 (0.68) [0.58]				
SDI NO _x			-1.081 (2.31) [1.81]			
SDI PM ₁₀				-1.695*** (0.50)*** [0.50]***		
SDI PM _{2.5}					-1.052 (0.61) [0.58]	
SDI SO ₂						-0.861** (0.26)** [0.27]**
Total pop.	0.000 (0.00) [0.00]	0.000*** (0.00)*** [0.00]**	0.000** (0.00)** [0.00]*	0.000*** (0.00)*** [0.00]***	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]
Feminity ratio	-0.167 (0.09) [0.08]	0.272 (0.23) [0.19]	0.452 (0.76) [0.56]	-0.034 (0.16) [0.14]	-0.038 (0.21) [0.18]	-0.084 (0.09) [0.08]
Mun. rev. pc	-0.001** (0.00)** [0.00]**	-0.000 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.002* (0.00)* [0.00]**	-0.001 (0.00) [0.00]	-0.000 (0.00) [0.00]
House price	0.000 (0.00) [0.00]	-0.000 (0.00) [0.00]	-0.000 (0.00) [0.00]	-0.002** (0.00)** [0.00]**	-0.001 (0.00) [0.00]	-0.001** (0.00)** [0.00]
Agri. share	-3.206 (2.53) [2.61]	-14.342* (6.34)* [3.72]***	-34.793 (21.68) [9.02]***	-3.417 (4.31) [3.87]	-3.316 (6.78) [4.92]	-2.007 (2.41) [2.53]
Constant	28.203** (9.89)** [9.23]**	-8.373 (24.20) [19.09]	-16.163 (82.12) [55.73]	49.788** (16.91)** [15.19]**	35.058 (22.89) [20.21]	19.454* (9.37)* [9.24]*
Observations	118	138	137	221	95	129
Adjusted R^2	0.081	0.426	0.251	0.119	0.003	0.128

Standard errors reported in parenthesis and robust standard errors reported in brackets. Mun. rev. pc is municipality revenue per capita. House price is median for 1 m² of residential premises. Agri. share is share of workers in agriculture, fishery and forestry in total working population.

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

management, but it accounts for only about 13% of NO_x emissions (Debski et al., 2018). The main sources of emission for NO_x and NO₂ in Poland are traffic and combustion of fossil fuels. Thus, with urbanized regions being the biggest emission sources for NO₂ and NO_x the large coefficients for both of them are largely in line with the theory.

Large coefficients of agri. share for those two pollutants can be connected to the variation in my sample, meaning that the largest cities, with low agricultural workers shares are having a considerable influence over the estimates. Nevertheless, there is possibility of agri. share being a

poor proxy of agricultural regions, since it also includes fishery and forestry workers. According to Sikora (2012) fishery and forestry are not as popular in Poland as agriculture, thus I do not expect them to constitute a large part of this variable and even if that is the case, I still assume fishery and forestry employment to at least be good predictors of more rural regions.

Table 4: Expanded model excl. 4 largest cities (join-by-location)

	BenPM ₁₀	NO ₂	NO _x	PM ₁₀	PM _{2.5}	SO ₂
	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE
SDI BenPM ₁₀	-0.552 (0.32) [0.29]					
SDI NO ₂		-0.895 (0.61) [0.56]				
SDI NO _x			-0.793 (1.81) [1.62]			
SDI PM ₁₀				-2.021*** (0.51)*** [0.52]***		
SDI PM _{2.5}					-1.327* (0.64)* [0.61]*	
SDI SO ₂						-0.930** (0.28)** [0.28]**
Total pop.	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]	-0.000 (0.00) [0.00]	-0.000 (0.00) [0.00]	-0.000 (0.00) [0.00]
Feminity ratio	-0.194 (0.10) [0.10]*	0.258 (0.21) [0.19]	0.417 (0.62) [0.52]	-0.128 (0.17) [0.16]	-0.078 (0.23) [0.22]	-0.096 (0.10) [0.09]
Mun. rev. pc	-0.001* (0.00)* [0.00]*	0.000 (0.00) [0.00]	0.001 (0.00) [0.00]	-0.001* (0.00)* [0.00]*	-0.001 (0.00) [0.00]	-0.000 (0.00) [0.00]
House price	-0.000 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.002 (0.00) [0.00]	-0.002** (0.00)** [0.00]**	-0.001 (0.00) [0.00]	-0.001** (0.00)** [0.00]
Agri. share	-2.926 (2.65) [2.67]	-13.395* (5.34)* [3.52]***	-31.636 (16.05) [7.36]***	-3.692 (4.21) [3.87]	-3.726 (6.78) [4.67]	-2.133 (2.48) [2.57]
Constant	30.494** (10.82)** [10.06]**	-8.069 (21.93) [19.64]	-16.428 (65.70) [52.00]	60.121*** (17.63)*** [16.32]***	39.811 (24.65) [22.65]	21.004* (10.20)* [10.03]*
Observations	107	125	124	202	82	122
Adjusted R^2	0.066	0.353	0.188	0.129	0.030	0.128

Standard errors reported in parenthesis and robust standard errors reported in brackets. Mun. rev. pc is municipality revenue per capita. House price is median for 1 m² of residential premises. Agri. share is share of workers in agriculture, fishery and forestry in total working population.

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

In order to limit the influence of largest cities with municipality rights on the estimates I employ Cook's distance proposed by Cook (1977). This measure displays the influence of each observation on the regression in question. I exclude observations for which Cook's distance value is

three times higher than the mean of Cook’s distance value (this approach is used i.a. by Stemmler et al. (2015) and Makhavikova (2018)). Thus, I exclude from my model four largest cities with municipality rights in Poland - Warsaw, Kraków, Łódź and Wrocław.

The expanded model using the limited sample is presented in Table 4. In line with discussion from previous table I observe a dampening of coefficients for NO_2 and NO_x , due to the largest cities not having as strong influence over the estimates as previously. Nevertheless, the coefficient for agri. share is still significant, which can be explained by still having considerably large cities with municipality rights in sample, such as Poznań and Katowice. Except for those two pollutants and BenPM_{10} I observe an improvement of adj. R^2 and significant correlation between SDI and $\text{PM}_{2.5}$.

The samples used in this analysis are rather small, which precludes any strong conclusions. Nevertheless, the results hint at a negative correlation between socioeconomic deprivation index and air pollution in Poland, which is statistically significant for PM_{10} , $\text{PM}_{2.5}$ and SO_2 . Those results differ from the large consensus in environmental economics literature about an inverse relation, in which poorer parts of population are exposed to elevated levels of hazardous air pollution (Banzhaf et al., 2019a, 2019b; Hsiang et al., 2019). To investigate whether those relations follow a linear path in the next subsection I present the results from quantile comparison analysis.

4.2 Quantile comparison

To account for the potential low explanatory power of the presented models I use quantile comparison as a simple and robust method of establishing the correlations between variables of interest (Li et al., 2018). Every figure in this section presents the relationships between a variable of choice with each pollutant’s mean concentrations. Moreover, each graph plots the relation both for the simple join-by-location method (solid line) and for 2 km buffer aggregation method (dashed line). I first present the relation between the SDI and pollution and afterwards look at the relations between pollution and variables composing the SDI.

In Figure 2 I plot the relationship between SDI and air pollution. As in previous subsection the straightforward interpretation of SDI quantiles is that the lower the quantile the wealthier is the municipality. When analysing top and bottom quantiles for all the pollutants I observe that the wealthier quantiles, except for BenPM_{10} , are always exposed to higher levels of pollution, confirming the negative correlation observed in OLS model estimation. At the same time I observe that the graphs for BenPM_{10} and SO_2 display inverted U-shape relation, which hints at "middle" municipalities being exposed to the highest levels of pollution. In case of BenPM_{10} the graph is also less "steep" for simple method, which can explain the lack of significance for all OLS estimations regarding BenPM_{10} . Moreover, the fact that for BenPM_{10} the most socioeconomically deprived

quantile is exposed to higher pollutant concentration than the wealthiest one can explain the positive coefficient for BenPM₁₀ in Table 2.

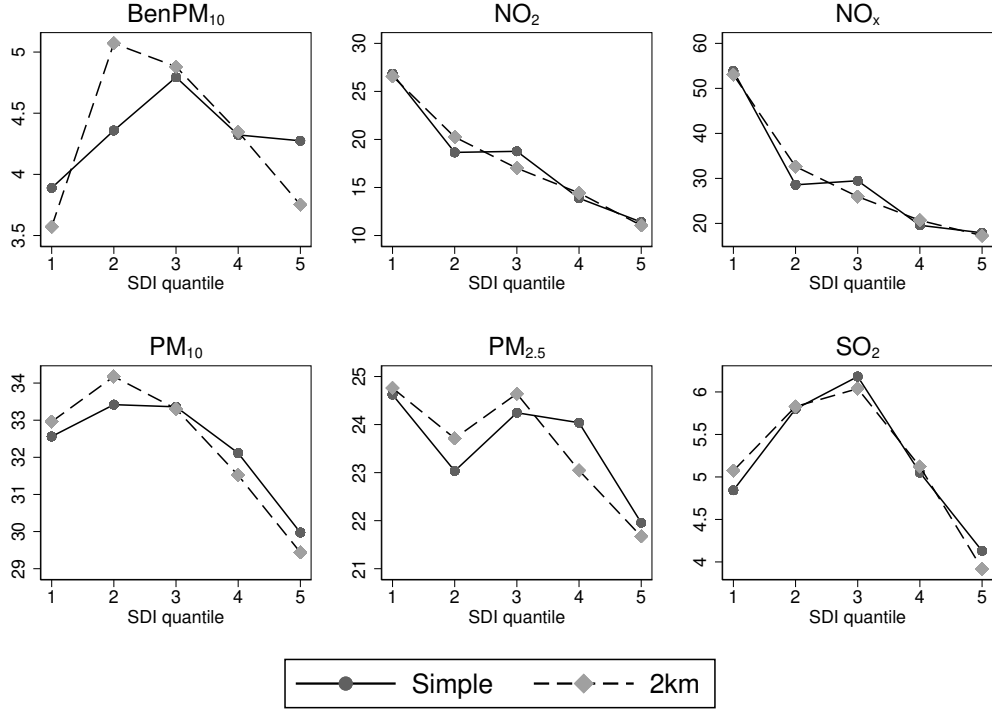


Figure 2: Quantile graphs for socioeconomic deprivation and air pollution

In Figure 3 I plot the quantiles for air pollution and average gross salary relative to national average. Due to negative correlation between income and SDI I expect those graphs to exhibit positive relationships. As in Figure 2, I observe that the highest quantiles of population are exposed to highest levels of pollution for all pollutants except BenPM₁₀. Moreover, SO₂ showcases the inverted U-shape relation, while BenPM₁₀ is the only pollutant to display a negative relation. The differing result for BenPM₁₀ and other pollutants are in line with the differing estimates of OLS and the initial assumption of different air pollutants displaying different relationships with socioeconomic deprivation index.

Given strong correlation of average salary and enrolment ratio I expect Figure 4 to display similar relationships as Figure 3. Nevertheless, the graphs for PM₁₀ and PM_{2.5} do not follow this expectation. For PM₁₀ I observe the same inverse U-shaped relation I observed in previous graphs for SO₂, while PM_{2.5} is not following any particular pattern, especially when comparing it between two aggregation methods. This figure further confirms that BenPM₁₀ is following a different pattern than all the other pollutants and displays a negative relation between proxy for education and concentration of BenPM₁₀.

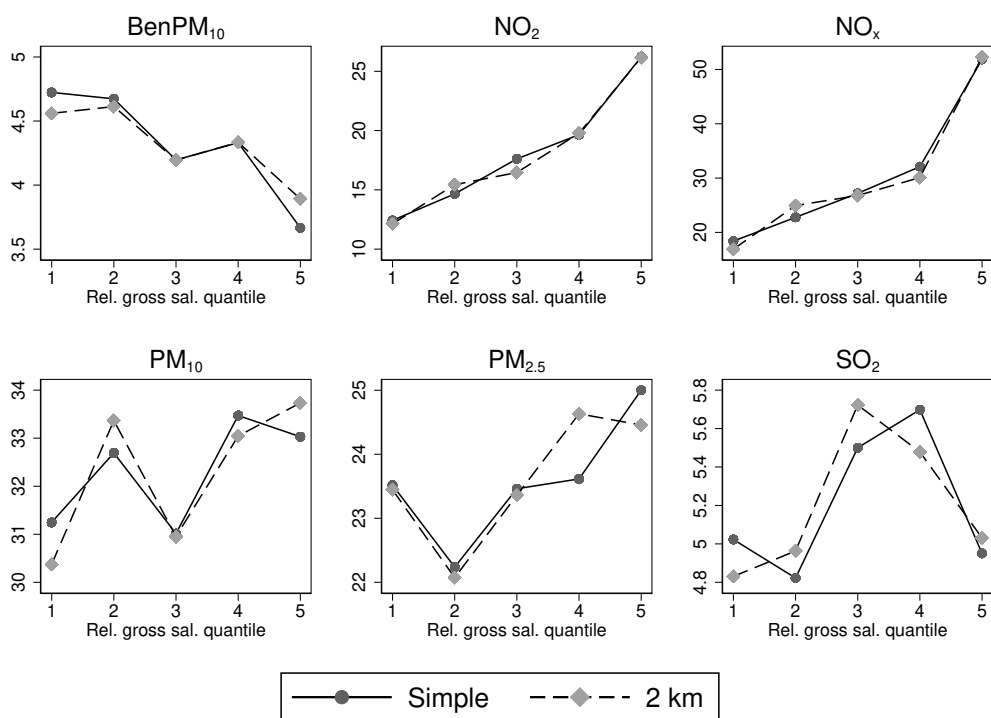


Figure 3: Quantile graphs for relative gross salary and air pollution

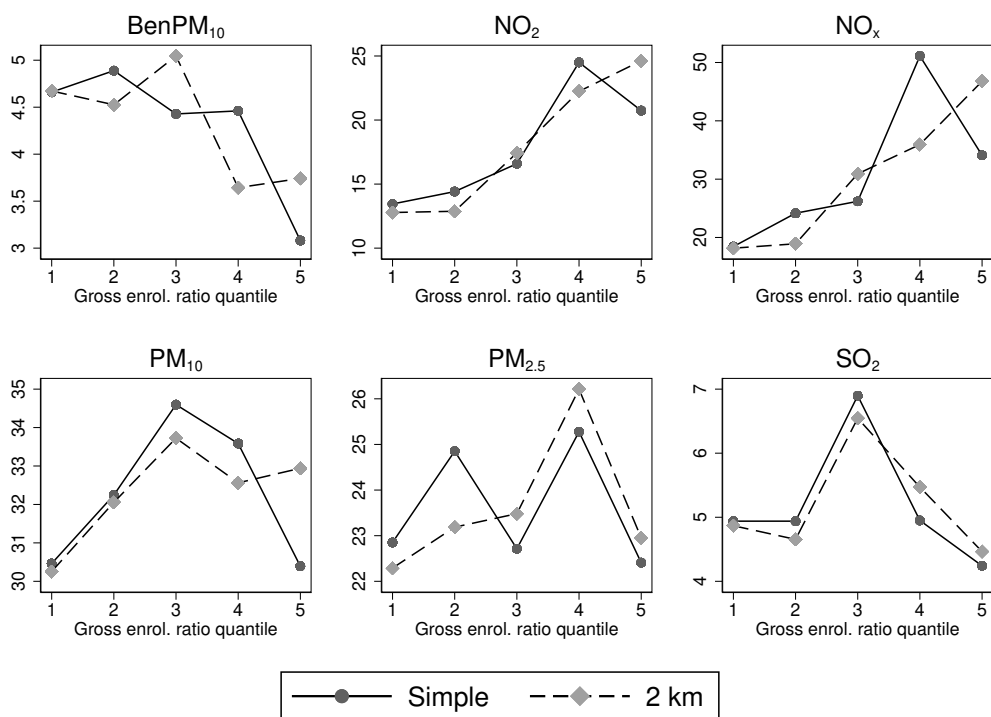


Figure 4: Quantile graphs for gross enrolment ratio and air pollution

I present the graphs plotting the relationship between pollution and social benefits per 10,000 and share of unemployment in Appendix A.3. Generally graphs for those relations follow the negative correlation path which is expected when looking at Figure 2 and Table 4. Moreover, those graphs further showcase the inverted U-shape relations from previous graphs when looking at relations for BenPM₁₀ and SO₂. Together with Figure 3, which described relation of air pollution and income, the graphs for social benefits and unemployment showcase almost a linear negative correlation for NO₂ and NO_x, which are in line with those pollutants exhibiting the highest coefficients in the basic model presented in Table 2. Contrary to previous figures, BenPM₁₀ displays same relation with social benefits and unemployment share as other pollutants.

Summing up the results from OLS modeling and quantile comparison, the main takeaway from the analysis is the negative correlation for all the analyzed pollutants, which is statistically significant for PM₁₀, PM_{2.5} and SO₂. Moreover, the pollutants do not exhibit the same pattern of relationships with socioeconomic deprivation, as some of them take form close to linear (as NO₂ and NO_x), while others exhibit a relation closer to inverse U-shape (SO₂), which hints at the air pollution distribution being allocated disproportionately to both the wealthiest and poorest parts of the population. Further the analysis hints at BenPM₁₀ displaying a different relationship with socioeconomic deprivation when compared to other pollutants, which can be seen in Table 2 and Figure 3 and 4. Further my results are in contrast with most of the environmental justice literature, which generally reports positive relation between socioeconomic deprivation and air pollution, meaning that poorer populations are exposed to elevated levels of air pollution.

4.3 Robustness

Below I discuss the use of areal containment aggregation method and investigate the impact of further sample limitations in OLS modeling.

4.3.1 Varying the aggregation radii

In previous sections I compared my results using data from simple join-by-location method with the results using the 2 km buffer method. I set my buffer distance to 2 km for it to be between the size of the buffer used by Currie et al. (2015)¹⁴ who report hazardous air pollution to generally be detectable within one mile of a plant producing the pollutants and 3 km buffer of Mohai and Saha (2015), which is based on the previous literature.

In addition to 2 km buffer estimation, I estimate all the OLS models with data aggregated with

¹⁴I am aware of the difference between measuring station and a polluting plant, although for simplicity I assume air pollutants in my study to follow similar dispersion from the measuring station as the ones in Currie et al. (2015).

1, 5 and 10 km buffers. The results for 1 km buffer are almost identical to the normal aggregation method and most of the literature is not using such a small buffer, thus I decided not to use it as a comparison point for my analysis. When using 5 km and 10 km buffers all of the coefficient of interest tend to be larger, while the standard errors smaller which leads to almost all of the relationships between pollution and SDI becoming statistically significant. Except for early environmental justice literature such as Anderton et al. (1994), I did not find studies employing buffers of this size, which is why I refrain from using them in my analysis.

Moreover, given that my method of aggregation is simplified and is only using an average of socioeconomic data (rather than assigning each of municipalities an individual weight) of all municipalities within the buffer with the increase of the buffer coverage to 5 km and 10 km I significantly increase risk of committing an ecological fallacy, which indicates that my aggregated results can significantly differ from the results of same analysis at a micro scale. In simple terms for the ecological fallacy not to happen I would need to make an extreme assumption that in my sample there are no group-level effects correlated with pollution (Banzhaf et al., 2019b). This occurrence could further preclude me from any strong conclusions. Further from goodness-of-fit perspective the models with larger buffers tend to have smaller adj. R^2 scores.¹⁵

4.3.2 Exclusion of industrialized voivodships

Previous OLS tables report negative correlation between SDI and air pollution, which is different than the results for similar studies in US or Korea (Banzhaf et al., 2019a; Choi et al., 2016). One possible explanation of this phenomena could be the "planned industrialization" hypothesis.

Even though, in previous specifications I excluded largest municipalities from the estimation, this hypothesis would suggest that my results can be affected by "planned industrialization" of Poland during 40 years of communist rule. During the communist rule the economic development of Poland was directed by political powers in Poland and USSR. This resulted in building of power plants and heavy industry without an economical reasoning behind it. This would skew the results of this analysis, because the polluting plants and factories were located based on economical incentives¹⁶ but on whims of ruling party (Dwilewicz, 2016; Kasztelewicz & Zajackowski, 2009). This could potentially lead to wealthy population in Poland to on average live in highly polluted areas, such as Katowice agglomeration (the largest agglomeration in the country) or Silesian conurbation, due to the party usually locating the polluting facilities in close proximity to large cities.¹⁷

¹⁵The improvement in adj. R^2 is only visible for BenPM₁₀, but the refinement is relatively small considering the loss of goodness-of-fit for models with other pollutants.

¹⁶Those economical incentives could i.a. lead to factories locating in poorer areas, where population has lower political capital.

¹⁷Those regions have the highest total population and population density, even when compared to the capital city

To examine this claim I split my data set into two groups, vovoidships with high energy production and vovoidships with low energy production (the high producing group consists of three vovoidships, which together produce approximately 35% of total national output). My justification for such selection is that the main source of air pollution is not mining of fossil fuels but their combustion, which can also take place in regions not typically associated with mining (i.e. the largest power plant in Poland, Bełchatów, is located in a "non mining and not industrial" vovoidship in central Poland).¹⁸ This sample restriction is not accounting for possibility of factories affecting the pollution, but given that the combustion of fossil fuels in plants is deemed as one of the main sources of pollution in Poland I focus on energy production.

I present the results of estimating my model on the low energy production sample in Table 5. The main takeaway from this table is the lack of significance for all the pollutants of interest and a change in coefficient sign for NO_2 and NO_x . This could imply that the high energy producing vovoidships are indeed driving the estimation results in the previous specifications, which implies that there could be a large heterogeneity of relations between air pollution and socioeconomic deprivation across different regions of Poland. The dampening of coefficients and the change in signs also implies that the less industrialized parts of Poland could display an opposite relationship between pollution and socioeconomic deprivation compared to ones "highly industrialized" during the communist rule.

I focus my analysis on observing the correlations and not a causal effect. Nevertheless, I observe a considerable drop in adj. R^2 scores in all equations, except the one for NO_x , which hints at a low explanatory power of the estimation. This leaves the possibility of heterogeneity of relationships across industrialized and non industrialized regions of Poland being a mere speculation.

In Table 6 I present the results of estimation of the expanded model using data limited to only high energy production vovoidships. Interestingly, I observe a considerable improvement in the adj. R^2 scores except for NO_x . When looking at estimates with heteroskedastic-consistent errors I observe significant correlation between socioeconomic deprivation and four pollutants: BenPM_{10} ¹⁹, PM_{10} , $\text{PM}_{2.5}$ and SO_2 . This supports the hypothesis of varying SDI and air pollution relationships and high energy producing vovoidships driving the significance of previous results.

More importantly, when limiting the sample further to include only the largest municipalities in high energy production vovoidships I observe a significant change in agri. share coefficient. It changes its sign and loses significance, signaling that in "largest and most industrialized" municipalities there is a positive relationships between agri. share and pollution, which further hints at the possibility of large heterogeneity of relations between air pollution and SDI across different regions

of Warsaw.

¹⁸Moreover, the largest power plants in Poland were built during the communist rule.

¹⁹It is important to note that BenPM_{10} has not been significant in any of the previous specifications and displayed an inverse relation in quantile comparison.

Table 5: Expanded model excl. 4 largest cities and high energy prod. vovoidships (join-by-location)

	BenPM ₁₀	NO ₂	NO _x	PM ₁₀	PM _{2.5}	SO ₂
	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE
SDI BenPM ₁₀	-0.579 (0.42) [0.41]					
SDI NO ₂		0.023 (0.58) [0.61]				
SDI NO _x			1.613 (1.40) [1.84]			
SDI PM ₁₀				-0.870 (0.55) [0.56]		
SDI PM _{2.5}					-0.143 (0.65) [0.56]	
SDI SO ₂						-0.476 (0.31) [0.30]
Total pop.	-0.000 (0.00) [0.00]	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]	-0.000 (0.00) [0.00]	-0.000 (0.00) [0.00]
Feminity ratio	-0.117 (0.13) [0.09]	0.375 (0.19) [0.21]	0.716 (0.47) [0.49]	-0.019 (0.18) [0.15]	0.188 (0.23) [0.22]	0.078 (0.10) [0.09]
Mun. rev. pc	-0.001 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.001 (0.00) [0.00]	0.000 (0.00) [0.00]	-0.001 (0.00) [0.00]
House price	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]	0.001 (0.00) [0.00]	-0.001 (0.00) [0.00]	0.000 (0.00) [0.00]	-0.000 (0.00) [0.00]
Agri. share	0.186 (3.67) [4.24]	-11.887* (4.98)* [4.49]**	-29.012* (11.92)* [9.07]**	1.488 (4.50) [4.73]	4.520 (6.46) [4.37]	2.897 (2.56) [3.10]
Constant	21.046 (13.35) [8.52]*	-21.641 (20.20) [20.68]	-50.386 (48.30) [49.18]	39.633* (18.41)* [15.08]**	1.505 (24.27) [20.54]	-0.026 (10.64) [9.40]
Observations	77	93	92	149	60	89
Adjusted R^2	0.028	0.384	0.270	0.025	-0.041	0.022

Standard errors reported in parenthesis and robust standard errors reported in brackets. Mun. rev. pc is municipality revenue per capita. House price is median for 1 m² of residential premises. Agri. share is share of workers in agriculture, fishery and forestry in total working population.

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

of Poland. However, as with Table 5 the sample size used in this estimation is small, thus the above interpretation can at best be speculative.²⁰

²⁰Given the mentioned limitations of this part of the study I decide to not include it in the main part of my paper, as it is speculative in nature. Nevertheless, this is one of the potential explanations for why Polish environmental justice results differ from the general US consensus.

Table 6: Expanded model for only high energy prod. vovoidship (join-by-location)

	BenPM ₁₀	NO ₂	NO _x	PM ₁₀	PM _{2.5}	SO ₂
	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE
SDI BenPM ₁₀	-0.764 (0.44) [0.37]*					
SDI NO ₂		-2.284 (1.74) [1.12]				
SDI NO _x			-5.729 (6.48) [4.04]			
SDI PM ₁₀				-2.560* (1.01)* [0.83]**		
SDI PM _{2.5}					-2.775** (0.89)** [0.76]**	
SDI SO ₂						-0.984* (0.40)* [0.36]*
Total pop.	0.000 (0.00) [0.00]*	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]*	0.000 (0.00) [0.00]	0.000** (0.00)** [0.00]**
Feminity ratio	-0.258 (0.14) [0.19]	0.111 (0.51) [0.34]	0.138 (1.90) [1.16]	-0.261 (0.27) [0.23]	-0.176 (0.26) [0.27]	-0.387** (0.12)** [0.10]**
Mun. rev. pc	-0.001** (0.00)** [0.00]**	-0.000 (0.00) [0.00]	-0.000 (0.01) [0.00]	-0.003** (0.00)** [0.00]**	-0.002** (0.00)** [0.00]**	-0.001* (0.00)* [0.00]**
House price	-0.001* (0.00)* [0.00]	-0.002 (0.00) [0.00]*	-0.005 (0.01) [0.00]	-0.002* (0.00)* [0.00]	-0.003** (0.00)** [0.00]**	-0.002*** (0.00)** [0.00]**
Agri. share	-8.667* (3.52)* [4.32]	-24.314 (14.52) [6.69]**	-52.074 (53.90) [20.78]*	-17.934* (7.29)* [6.05]**	-26.942* (10.19)* [13.07]	-16.652*** (3.34)** [3.48]**
Constant	43.784** (15.75)** [22.19]	19.597 (57.22) [37.28]	41.641 (212.39) [126.77]	88.512** (29.35)** [24.85]**	67.954* (28.33)* [29.51]*	61.830*** (13.46)** [11.46]**
Observations	34	38	38	63	29	37
Adjusted R^2	0.345	0.376	0.106	0.348	0.593	0.676

Standard errors reported in parenthesis and robust standard errors reported in brackets. Mun. rev. pc is municipality revenue per capita. House price is median for 1 m² of residential premises. Agri. share is share of workers in agriculture, fishery and forestry in total working population.

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

5 Caveats and areas for future research

In this section I discuss the possible extensions to my analysis, which could be potential avenues for future research of environmental justice in Poland. Moreover, I briefly discuss the external validity of my paper.

5.1 Data collection process

Given that across all the pollutants maximum number of observations used in analysis is 221 there is a possibility of greatly enriching the analysis given a richer data set. In absence of such data due to sparsity of monitoring stations in Poland it is still possible to artificially create such data. Some papers gather data on air pollution by using sophisticated pollution estimation methods. For example, using data on urban dynamics such as pressure, humidity, wind directions and road map Li et al. (2018) estimate the data on air pollution in Hong Kong in areas that are not covered by the initial monitoring stations in the data. Using this method Li et al. (2018) achieve 82% accuracy at 100 m x 100 m spatial resolution.

Moreover, a possible avenue for further research would be to obtain further socioeconomic data both on smaller scale and new ones on the present one. Smaller scale data would make committing an ecological fallacy harder, while broader data could facilitate deeper analysis. For example, data on car ownership, energy production or industrial workforce of particular municipalities would facilitate exploring additional channels of correlation. Nevertheless, right now such data are not available. This limits the scope of potential channels to explore, even if there was a possibility of obtaining data on smaller administrative levels than NUTS 3.

5.2 Estimation of the damage function

Even if I assume perfect analysis of relationship between SDI and air pollution the results will not inform me about the damage caused by this pollution. This is a result of heterogeneity of the damage function across the population. For example, even if the wealthier parts of the society are exposed to considerably higher levels of hazardous air pollution in comparison to poorer parts of the society, wealthy individuals have a considerably better access to measures protecting them from the damaging effects of such pollution (Hsiang et al., 2019). Thus, even with wealthier parts of the society being exposed to higher shares of pollution the damage function for both populations can be different and the actual "environmental injustice" can negatively affect the poorer parts of population due to higher "real" health and economic burden of pollution.

This affects the possible policy implications of this work, since without knowing the actual costs of the pollution for the groups exposed to it, it is impossible for the social planner to devise policies that would be efficient in mitigating the damages from pollution (or "equalizing" the environmental injustices).

5.3 Longitudinal analysis

Municipal emission is deemed as one of the largest sources of air pollution in Poland. Thus, in 2018 and 2019 dozens of municipalities mostly affected by the pollution (such as Kraków) have passed set of laws, which are implemented to curb the emissions of hazardous air pollution (Kleczkowski, 2020). Given the unequal distribution of air pollution in Poland such programmes can have unfair distribution of costs of transition from "polluted economy" to a "clean" one. Setting this aside and assuming efficiency of those measures, I expect that in the upcoming years the air pollution concentrations across Poland could change considerably, especially for emissions of particulate matter. Thus, extending the analysis to further years and investigating the marginal changes to relationships after the implementation of said programmes could showcase another facet of environmental injustice in Poland.

5.4 External validity

The European countries are highly heterogeneous when compared with each other, i.a. when accounting for energy generation, population or level of industrialization (Fortuński et al., 2016). Government of Poland is right now constructing Ostrołęka C, which is planned to be one of the largest coal powered power plants in Poland, while on the other hand a lot of cities are passing laws regulating municipal emission (Wilczyńska-Michalik et al., 2016). Given those two factors the external validity of this study is low, since Poland is likely not a good comparison unit for other European countries and at the same time the situation concerning air pollution in the country can largely change in the upcoming years.

6 Conclusions

This paper investigates the relation between socioeconomic deprivation and air pollution in Poland in 2018. This is the first paper to conduct an environmental justice analysis of air pollution in Poland. Environmental justice issues in Eastern Europe have been largely neglected by the modern environmental justice literature, particularly in case of Poland. This is troublesome, because extrapolating environmental justice literature results from US or Western European countries can oftentimes be misleading given the differences between those regions and Eastern Europe.

In my analysis I use ordinary least squares model. I expand the model with potential channels of interaction between my main variables and limit the sample size used in analysis in order to account for the high variation among observations in my data set. Moreover, I use quantile comparison to account for the possibility of spatial auto-correlation.

The main results from my analysis are twofold. Firstly, my analysis establishes that there exist a negative correlation between socioeconomic deprivation and air pollution in Poland for nitrogen dioxide (NO_2) and oxides (NO_x), particulate matter with diameters smaller than $10\text{ }\mu\text{m}$ and $2.5\text{ }\mu\text{m}$ (PM_{10} and $\text{PM}_{2.5}$) and sulphur dioxide (SO_2). This means that wealthier parts of population in Poland are exposed to disproportionately larger shares of hazardous air pollution than the poorer parts. Only the correlations for PM_{10} , $\text{PM}_{2.5}$ and SO_2 are statistically significant when including total population, feminity ratio, municipality's revenues, house prices and share of workers in agriculture, fishery and forestry into the model.

Secondly, the results of quantile comparison analysis point to a possibility of benzo(a)pyrene in PM_{10} (BenPM_{10}) exhibiting an inverse relation with income and education, when compared with other pollutants. This supports the hypothesis of air pollutants in Poland displaying a heterogeneous relationship with socioeconomic deprivation. When using simplified areal containment method of data aggregation results of my analysis do not change significantly and still report negative correlation for all pollutants.

My analysis concerns only the relationship between socioeconomic deprivation and air pollution and is not focused on estimating the welfare damages connected with hazardous air pollution. Due to possible heterogeneity of the pollution damage function for different populations my work does not have a direct policy implication.

7 References

- Almond, D., Edlund, L., & Palme, M. (2009). Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in Sweden. *The Quarterly Journal of Economics*, 124(4), 1729–1772.
- Anderton, D. L., Anderson, A. B., Oakes, J. M., & Fraser, M. R. (1994). Environmental equity: The demographics of dumping. *Demography*, 31(2), 229–248.
- Angrist, J. D., & Pischke, J. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- Banzhaf, S., Ma, L., & Timmins, C. (2019a). Environmental justice: Establishing causal relationships. *Annual Review of Resource Economics*, 11, 377–398.
- Banzhaf, S., Ma, L., & Timmins, C. (2019b). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives*, 33(1), 185–208.
- Barwiński, M. (2014). Struktura narodowościowa Polski w świetle wyników spisu powszechnego z 2011 roku. *Przegląd Geograficzny*, 86, 217–241.
- Bukowski, P., & Novokmet, F. (2018). Inequality in Poland: Estimating the whole distribution by g-percentile 1983-2015. *LIS Working Paper Series*.
- Bukowski, P., & Novokmet, F. (2019). Between communism and capitalism: Long-term inequality in Poland, 1892-2015. *International Inequalities Institute Working Paper*.
- Choi, G., Heo, S., & Lee, J.-T. (2016). Assessment of environmental injustice in Korea using synthetic air quality index and multiple indicators of socioeconomic status: A cross-sectional study. *Journal of the Air & Waste Management Association*, 66(1), 28–37.
- Chun, Y., Kim, Y., & Campbell, H. (2012). Using bayesian methods to control for spatial autocorrelation in environmental justice research: An illustration using toxics release inventory data for a Sunbelt county. *Journal of Urban Affairs*, 34(4), 419–439.
- Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: models & applications*. Taylor & Francis.
- Collart, P., Dubourg, D., Levêque, A., Sierra, N. B., & Coppieters, Y. (2018). Data on short-term effect of nitrogen dioxide on cardiovascular health in Wallonia, Belgium. *Data in brief*, 17, 172–179.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15–18.
- 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., Zivin, J. G., Mullins, J., & Neidell, M. (2014). What do we know about short-and long-term effects of early-life exposure to pollution? *Annual Review of Resource Economics*, 6(1), 217–247.

- Czyż, T., & Hauke, J. (2011). Evolution of regional disparities in Poland. *Quaestiones Geographicae*, 30(2), 35–48.
- Debski, B., Olecka, A., Bebkiewicz, K., Chłopek, Z., Kargulewicz, I., Rutkowski, J., Waśniewska, S., Zasina, D., Zimakowska-Laskowska, M., & Żaczek, M. (2018). Krajowy bilans emisji SO₂, NO_x, CO, NH₃, NMLZO, pyłów, metali ciężkich i TZO za lata 2015-2016 w układzie klasyfikacji SNAP. *Krajowy Ośrodek Bilansowania i Zarządzania Emisjami, Instytut Ochrony Środowiska-Państwowy Instytut Badawczy, Warszawa*.
- Dwilewicz, L. (2016). *Polityka energetyczna PRL*. Oficyna wydawnicza Szkoła Główna Handlowa w Warszawie.
- Edalatmanesh, M. A., Ketabchi, F., & Malekzadeh, S. (2018). The effect of air-polluting particles on cardiovascular disease: A review. *Report of Health Care*, 4(2), 60–67.
- EEA. (2019). Air pollution fact sheet 2013 Poland. Retrieved from: <https://www.eea.europa.eu/themes/air/country-fact-sheets/2019-country-fact-sheets/poland> (Date accessed: 10.05.2020).
- Ferreira, S., Akay, A., Brereton, F., Cuñado, J., Martinsson, P., Moro, M., & Ningal, T. F. (2013). Life satisfaction and air quality in Europe. *Ecological Economics*, 88, 1–10.
- Filčák, R., & Škobla, D. (2019). No pollution and no Roma in my backyard: Class and race in framing local activism in Laborov, eastern Slovakia. *Environmental Justice, Popular Struggle and Community Devt*, 53.
- Fortuński, B. Et al. (2016). Polityka energetyczna Unii Europejskiej–3× 20. Diagnoza i perspektywy w kontekście zrównoważonego rozwoju. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 453, 179–189.
- Giardullo, P., Sergi, V., Carton, W., Kenis, A., Kesteloot, C., Kazepov, Y., Kobus, D., Maione, M., Skotak, K., Fuzzi, S., Et al. (2016). Air quality from a social perspective in four European metropolitan areas: Research hypothesis and evidence from the SEFIRA project. *Environmental Science & Policy*, 65, 58–64.
- Grabowska-Lusińska, I., & Jaźwińska-Motyłska, E. (2013). Znaczenie migracji w życiu zawodowym kobiet i mężczyzn. *Kultura i Społeczeństwo*, 3, 85–108.
- Harper, K., Steger, T., & Filčák, R. (2009). Environmental justice and Roma communities in Central and Eastern Europe. *Environmental Policy and Governance*, 19(4), 251–268.
- Hou, Q., An, X., Wang, Y., Tao, Y., & Sun, Z. (2012). An assessment of China's PM₁₀-related health economic losses in 2009. *Science of the Total Environment*, 435, 61–65.
- Hsiang, S., Oliva, P., & Walker, R. (2019). The distribution of environmental damages. *Review of Environmental Economics and Policy*, 13(1), 83–103.
- Isen, A., Rossin-Slater, M., & Walker, W. R. (2017). Every breath you take—every dollar you'll make: The long-term consequences of the clean air act of 1970. *Journal of Political Economy*, 125(3), 848–902.

- Janke, K., Propper, C., & Henderson, J. (2009). Do current levels of air pollution kill? The impact of air pollution on population mortality in England. *Health Economics*, 18(9), 1031–1055.
- Kaiser, H. (1974). An index of factor simplicity. *Psychometrika*, 39, 31–36.
- Kasztelawicz, Z., & Zajackowski, M. (2009). Zarys historii i przyszłości węgla brunatnego w Polsce. *Przegląd Górniczy*, 65(5-6), 62–71.
- Kleczkowski, P. (2020). Dlaczego w Polsce jest smog? *Aura*, (2), 3–7.
- Laurent, E. (2011). Issues in environmental justice within the European Union. *Ecological Economics*, 70(11), 1846–1853.
- Li, V. O., Han, Y., Lam, J. C., Zhu, Y., & Bacon-Shone, J. (2018). Air pollution and environmental injustice: Are the socially deprived exposed to more PM2.5 pollution in Hong Kong? *Environmental Science & Policy*, 80, 53–61.
- Makhavikova, H. (2018). *Determinants of FDI in Central and Eastern Europe*. Springer.
- Mikati, I., Benson, A. F., Luben, T. J., Sacks, J. D., & Richmond-Bryant, J. (2018). Disparities in distribution of particulate matter emission sources by race and poverty status. *American Journal of Public Health*, 108(4), 480–485.
- Mitchell, G., & Dorling, D. (2003). An environmental justice analysis of British air quality. *Environment and Planning A*, 35(5), 909–929.
- Mohai, P., & Bryant, B. (1991). Race, poverty & the distribution of environmental hazards: Reviewing the evidence. *Race, Poverty & the Environment*, 2(3/4), 3–27.
- Mohai, P., Pellow, D., & Roberts, J. T. (2009). Environmental justice. *Annual Review of Environment and Resources*, 34, 405–430.
- Mohai, P., & Saha, R. (2015). Which came first, people or pollution? Assessing the disparate siting and post-siting demographic change hypotheses of environmental injustice. *Environmental Research Letters*, 10(11).
- Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2), 243–251.
- Muehlenbachs, L., Spiller, E., & Timmins, C. (2015). The housing market impacts of shale gas development. *American Economic Review*, 105(12), 3633–59.
- Ortiz, A. (2015). Air quality in Europe - 2019 report. *European Environment Agency (EEA)*.
- Paloniemi, R., Apostolopoulou, E., Cent, J., Bormpoudakis, D., Scott, A., Grodzińska-Jurczak, M., Tzanopoulos, J., Koivulehto, M., Pietrzyk-Kaszyńska, A., & Pantis, J. D. (2015). Public participation and environmental justice in biodiversity governance in Finland, Greece, Poland and the UK. *Environmental Policy and Governance*, 25(5), 330–342.
- Pawlak, A. (2016). Płeć emigracji. kobiece i meskie wzorce emigrowania. *Rocznik Lubuski*, 42(1), 213–228.

- Pearce, J., Kingham, S., & Zawar-Reza, P. (2006). Every breath you take? Environmental justice and air pollution in Christchurch, New Zealand. *Environment and Planning A*, 38(5), 919–938.
- Piketty, T., & Saez, E. (2003). Income inequality in the United States, 1913–1998. *The Quarterly Journal of Economics*, 118(1), 1–41.
- Poczta, W., & Bartkowiak, N. (2012). Regionalne zróżnicowanie rolnictwa w Polsce. *Journal of Agribusiness and Rural Development*, 1(23).
- Prüss-Üstün, A., Wolf, J., Corvalán, C., Bos, R., & Neira, M. (2016). *Preventing disease through healthy environments: A global assessment of the burden of disease from environmental risks*. World Health Organization.
- Ringquist, E. J. (1997). Equity and the distribution of environmental risk: The case of TRI facilities. *Social Science Quarterly*, 78(4), 811–829.
- Roszkowska, E., & Karwowska, R. (2014). Wielowymiarowa analiza poziomu zrównoważonego rozwoju województw Polski w 2010 roku. *Ekonomia i Środowisko*, 2, 168–190.
- Sikora, J. (2012). Wielofunkcyjność obszarów wiejskich w Polsce. *Journal of Agribusiness and Rural Development*, 2(24).
- Smogowy, K. A. (2015). *Co wiemy o smogu? informowanie o zanieczyszczeniu powietrza w Polsce*. Krakowski Alarm Smogowy.
- Šrám, R. J., Binková, B., Dejmek, J., & Bobak, M. (2005). Ambient air pollution and pregnancy outcomes: A review of the literature. *Environmental Health Perspectives*, 113(4), 375–382.
- Stemmler, M., Eye, A., & Wiedermann, W. (2015). Dependent data in social sciences research. *Springer Proceedings in Mathematics & Statistics*, 145, 2015.
- UN. (2008). *The right to health. Fact sheet no. 31*. United Nations Geneva.
- Varga, C., Kiss, I., & Ember, I. (2002). The lack of environmental justice in Central and Eastern Europe. *Environmental Health Perspectives*, 110(11), A662–A663.
- Varró, M., Gombkőto, G., Szeremi, M., Et al. (2001). Risk factors of a mass lead exposure, Heves, Hungary. *Egészségtudomány*, 45, 167–80.
- Vasileva, K. (2012). Nearly two-thirds of the foreigners living in EU member states are citizens of countries outside the EU-27. *Eurostat Statistics in Focus*, 31, 2012.
- Wang, Z.-b., & Fang, C.-l. (2016). Spatial-temporal characteristics and determinants of PM_{2.5} in the Bohai Rim urban agglomeration. *Chemosphere*, 148, 148–162.
- WHO. (2016). World Health Organization global urban ambient air pollution database (update 2016). *Geneva, Switzerland: WHO*.
- Widziewicz, K., Rogula-Kozłowska, W., Loska, K., Kociszewska, K., & Majewski, G. (2018). Health risk impacts of exposure to airborne metals and benzo(a)pyrene during episodes of high PM₁₀ concentrations in Poland. *Biomedical and Environmental Sciences*, 31(1), 23–36.

- Wilczyńska-Michalik, W., Pietras, B., & Michalik, M. (2016). Smog w Krakowie-spojrzenie w przyszłość z perspektywy historycznej. *Aura*, (11), 3–8.
- Wu, Y., Li, R., Cui, L., Meng, Y., Cheng, H., & Fu, H. (2020). The high-resolution estimation of sulfur dioxide (SO₂) concentration, health effect and monetary costs in Beijing. *Chemosphere*, 241, 125031.
- Zieliński, E., Wielgus, A., Dreliszak, J., & Zukow, W. (2018). Air pollution-selected health effects in Poland. *Journal of Education, Health and Sport*, 8(12), 641–648.

A Appendix A

A.1 Pollutants description

Benzo(a)pyrene in PM₁₀ (BenPM₁₀). A carcinogenic organic compound that is connected to lung cancer and leukemia. It can be found in gasoline and its main anthropogenic sources are burning of fossil fuels, industrial production and cars (Widziewicz et al., 2018).

Nitrogen dioxide (NO₂). A chemical compound that is associated with elevated risks of cardiovascular diseases and hemorrhagic strokes. The main anthropogenic source is combustion of fossil fuels and traffic (Collart et al., 2018).

Nitrogen oxides (NO_x). A colourless free radical gas that is associated with site specific DNA methylation. The main anthropogenic source is combustion of fossil fuels and traffic. Agricultural soil management is also a large source of nitrogen oxide which is a component of NO_x (Edalatmanesh et al., 2018).

Particulate matter with diameters smaller than 10 µm and 2.5 µm (PM₁₀ and PM_{2.5}). Aerosol particles, usually not visible to naked eye, that literature connects to elevated risks of respiratory diseases, such as bronchitis, pneumonia or lung cancer. Moreover, sometimes those particles can get into the bloodstream of an individual, which can cause serious health issues. Main anthropogenic sources are automobiles, combustion of fossil fuels and industry (Hou et al., 2012).

Sulphur dioxide (SO₂). A chemical compound that is associated with increased cardio respiratory mortality and morbidity. It can damage respiratory system and hamper breathing. The main anthropogenic source is combustion of materials containing sulphur, which in most instances are fossil fuels such as oil and coal in power plants and industrial facilities (Wu et al., 2020).

A.2 Additional tables

Table 7: Descriptive statistics (2 km buffer)

	BenPM ₁₀	NO _x	NO ₂	PM ₁₀	PM _{2.5}	SO ₂
Concentration	4.32 (2.76)	17.95 (9.20)	30.06 (26.93)	32.29 (6.87)	23.57 (5.07)	5.21 (2.74)
Rel. gross sal.	88.15 (13.05)	92.81 (13.52)	92.89 (13.52)	90.20 (13.27)	93.63 (14.21)	91.65 (13.01)
Gross enrol. ratio	103.04 (10.18)	104.81 (11.19)	104.85 (11.22)	103.80 (10.31)	107.22 (10.95)	104.19 (11.24)
Soc. ben. per 10k	529.56 (221.80)	481.78 (231.01)	481.57 (231.20)	505.38 (226.80)	462.56 (213.38)	486.26 (226.46)
Unemp. share	4.29 (1.87)	4.03 (1.84)	4.05 (1.85)	4.11 (1.87)	4.06 (1.87)	4.05 (1.82)
Total pop.	188.52 (266.01)	232.39 (293.28)	234.24 (293.88)	213.11 (291.18)	277.52 (371.53)	206.74 (254.11)
Feminity ratio	107.04 (4.37)	108.33 (4.64)	108.38 (4.62)	107.57 (4.53)	109.52 (4.56)	107.97 (4.55)
Mun. rev. pc	5374.54 (1105.41)	5609.92 (1192.90)	5612.04 (1196.27)	5434.02 (1154.10)	5890.71 (1283.35)	5543.92 (1155.03)
House price	3639.84 (1333.72)	3898.44 (1545.11)	3892.88 (1555.07)	3737.57 (1446.60)	4035.25 (1548.50)	3777.76 (1492.95)
Agri. share	0.21 (0.18)	0.15 (0.17)	0.15 (0.17)	0.18 (0.17)	0.12 (0.15)	0.16 (0.17)
Observations	118	138	137	221	95	129

Standard deviations reported in parentheses. Rel. gross sal. is relative gross salary in relation to national average. Gross enrolment ratio is a percentage of population enrolled in the middle school education in the age group corresponding to this level of education, so given adult education it can be over 100. Soc. ben. per 10k is number of individuals receiving social benefits per 10,000. Mun. rev. pc is municipality revenue per capita. House price is median for 1 m² of residential premises. Agriculture share is share of workers in agriculture, fishery and forestry in total working population. All the pollutants except BenPM₁₀ are reported in µg in m³, while BenPM₁₀ is reported in ng in m³.

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

Table 8: Expanded model without SDI (join-by-location)

	BenPM ₁₀	NO ₂	NO _x	PM ₁₀	PM _{2.5}	SO ₂
	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE
Rel. gross sal.	-0.049 (0.03) [0.04]	0.097 (0.09) [0.10]	0.276 (0.29) [0.37]	0.036 (0.06) [0.06]	0.045 (0.08) [0.10]	0.038 (0.03) [0.04]
Gross enrol. ratio	-0.018 (0.03) [0.03]	0.025 (0.08) [0.07]	-0.008 (0.28) [0.21]	0.025 (0.06) [0.05]	0.003 (0.07) [0.06]	0.007 (0.03) [0.03]
Soc. ben. per 10k	-0.003 (0.00) [0.00]	-0.002 (0.00) [0.00]	0.010 (0.02) [0.01]	-0.009** (0.00)** [0.00]**	-0.005 (0.00) [0.00]	-0.003 (0.00) [0.00]
Unemp. share	-0.193 (0.18) [0.14]	-0.116 (0.49) [0.40]	-0.856 (1.63) [1.15]	-0.054 (0.33) [0.28]	-0.036 (0.38) [0.28]	-0.112 (0.18) [0.19]
Total pop.	0.000 (0.00) [0.00]	0.000*** (0.00)*** [0.00]*	0.000** (0.00)** [0.00]	0.000*** (0.00)*** [0.00]***	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]
Feminity ratio	-0.154 (0.11) [0.09]	0.271 (0.27) [0.20]	0.690 (0.92) [0.64]	-0.091 (0.19) [0.17]	-0.048 (0.24) [0.20]	-0.076 (0.10) [0.10]
Mun. rev. pc	-0.000 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.002 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.000 (0.00) [0.00]
House price	0.000 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.002** (0.00)** [0.00]**	-0.001 (0.00) [0.00]	-0.001** (0.00)** [0.00]
Agri. share	-3.653 (2.52) [2.54]	-14.615* (6.49)* [3.89]***	-36.340 (22.18) [9.94]***	-3.344 (4.34) [3.94]	-3.078 (6.95) [5.16]	-1.740 (2.48) [2.64]
Constant	32.492** (11.52)** [11.04]**	-16.790 (27.52) [21.08]	-60.527 (92.84) [66.51]	52.547** (19.44)** [18.19]**	33.922 (26.28) [24.17]	16.142 (10.54) [10.12]
Observations	118	138	137	221	95	129
Adjusted R^2	0.099	0.415	0.240	0.117	-0.025	0.113

Standard errors reported in parenthesis and robust standard errors reported in brackets. Mun. rev. pc is municipality revenue per capita. House price is median for 1m2 of residential premises. Agri. share is share of workers in agriculture, fishery and forestry in total working population.

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

Table 9: Basic model (2 km buffer)

	BenPM ₁₀	NO ₂	NO _x	PM ₁₀	PM _{2.5}	SO ₂
	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE
SDI BenPM ₁₀	-0.032 (0.17) [0.15]					
SDI NO ₂		-3.436*** (0.41)*** [0.47]***				
SDI NO _x			-7.603*** (1.32)*** [1.64]***			
SDI PM ₁₀				-0.975** (0.29)** [0.30]**		
SDI PM _{2.5}					-0.636 (0.33) [0.33]	
SDI SO ₂						-0.237 (0.16) [0.14]
Constant	4.325*** (0.25)*** [0.25]***	17.948*** (0.64)*** [0.64]***	30.062*** (2.07)*** [2.07]***	32.287*** (0.45)*** [0.45]***	23.567*** (0.51)*** [0.51]***	5.207*** (0.24)*** [0.24]***
Observations	118	138	137	221	95	129
Adjusted R^2	-0.008	0.337	0.191	0.044	0.027	0.010

Standard errors reported in parenthesis and robust standard errors reported in brackets.

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

Table 10: Expanded model (2 km buffer)

	BenPM ₁₀	NO ₂	NO _x	PM ₁₀	PM _{2.5}	SO ₂
	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE
SDI BenPM ₁₀	-0.532 (0.30) [0.30]					
SDI NO ₂		-1.148 (0.68) [0.60]				
SDI NO _x			-1.480 (2.35) [1.88]			
SDI PM ₁₀				-2.033*** (0.50)*** [0.51]***		
SDI PM _{2.5}					-1.666* (0.63)* [0.65]*	
SDI SO ₂						-0.986*** (0.26)*** [0.30]**
Total pop.	0.000 (0.00) [0.00]	0.000*** (0.00)*** [0.00]*	0.000** (0.00)** [0.00]	0.000** (0.00)** [0.00]***	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]
Feminity ratio	-0.159 (0.09) [0.09]	0.369 (0.22) [0.19]	0.645 (0.76) [0.58]	0.043 (0.16) [0.14]	0.090 (0.20) [0.17]	-0.067 (0.09) [0.08]
Mun. rev. pc	-0.001** (0.00)** [0.00]*	-0.000 (0.00) [0.00]	-0.000 (0.00) [0.00]	-0.002** (0.00)** [0.00]**	-0.001 (0.00) [0.00]	-0.001 (0.00) [0.00]
House price	0.000 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.002*** (0.00)*** [0.00]**	-0.001 (0.00) [0.00]	-0.001** (0.00)** [0.00]
Agri. share	-3.312 (2.35) [2.36]	-13.213* (6.01)* [3.59]***	-31.642 (20.80) [8.74]***	-2.513 (4.05) [3.57]	1.423 (5.72) [4.39]	-1.864 (2.27) [2.41]
Constant	27.266** (10.21)** [9.46]**	-17.911 (23.60) [19.86]	-35.570 (81.45) [58.69]	43.843* (17.06)* [15.07]**	22.714 (21.49) [19.10]	18.678* (9.13)* [9.32]*
Observations	118	138	137	221	95	129
Adjusted R^2	0.076	0.449	0.258	0.147	0.046	0.157

Standard errors reported in parenthesis and robust standard errors reported in brackets. Mun. rev. pc is municipality revenue per capita. House price is median for 1 m² of residential premises. Agri. share is share of workers in agriculture, fishery and forestry in total working population.

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

Table 11: Expanded model excl. 4 largest cities (2 km buffer)

	BenPM ₁₀	NO ₂	NO _x	PM ₁₀	PM _{2.5}	SO ₂
	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE	B/SE/rSE
SDI BenPM ₁₀	-0.558 (0.32) [0.30]					
SDI NO ₂		-1.064 (0.61) [0.56]				
SDI NO _x			-1.110 (1.87) [1.67]			
SDI PM ₁₀				-2.248*** (0.51)*** [0.53]***		
SDI PM _{2.5}					-1.934** (0.67)** [0.70]**	
SDI SO ₂						-1.049*** (0.29)*** [0.30]***
Total pop.	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]	0.000 (0.00) [0.00]	-0.000 (0.00) [0.00]	0.000 (0.00) [0.00]
Feminity ratio	-0.174 (0.10) [0.10]	0.361 (0.20) [0.19]	0.585 (0.63) [0.54]	-0.023 (0.17) [0.15]	0.088 (0.23) [0.22]	-0.065 (0.10) [0.09]
Mun. rev. pc	-0.001* (0.00)* [0.00]*	-0.000 (0.00) [0.00]	0.001 (0.00) [0.00]	-0.002* (0.00)* [0.00]**	-0.001 (0.00) [0.00]	-0.001 (0.00) [0.00]
House price	-0.000 (0.00) [0.00]	-0.001 (0.00) [0.00]	-0.003 (0.00) [0.00]	-0.002*** (0.00)*** [0.00]**	-0.001 (0.00) [0.00]	-0.001** (0.00)** [0.00]
Agri. share	-3.096 (2.47) [2.46]	-12.762* (5.13)* [3.35]***	-29.841 (15.63) [7.35]***	-2.963 (4.02) [3.67]	1.042 (5.78) [4.32]	-1.620 (2.38) [2.42]
Constant	28.491* (10.99)* [10.15]**	-17.643 (21.65) [20.13]	-32.427 (66.24) [54.13]	51.295** (17.80)** [16.25]**	23.912 (24.15) [22.80]	18.496 (10.07) [10.07]
Observations	109	123	122	202	80	119
Adjusted R^2	0.056	0.364	0.187	0.155	0.078	0.156

Standard errors reported in parenthesis and robust standard errors reported in brackets. Mun. rev. pc is municipality revenue per capita. House price is median for 1 m² of residential premises. Agri. share is share of workers in agriculture, fishery and forestry in total working population.

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

A.3 Additional figures

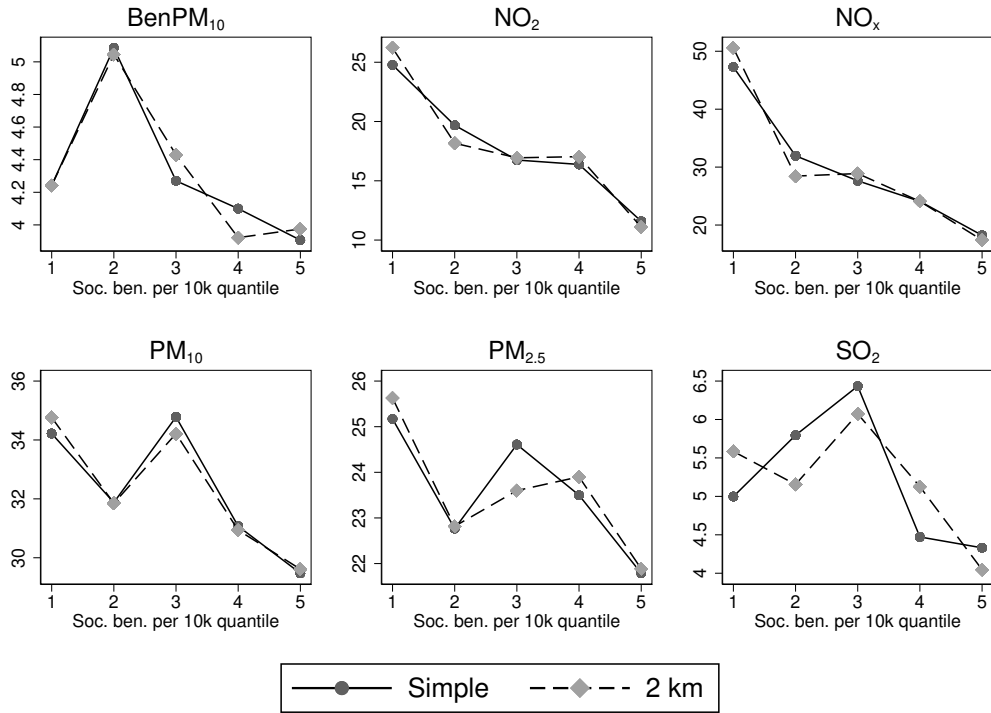


Figure 5: Quantile graphs for social benefits per 10k and air pollution

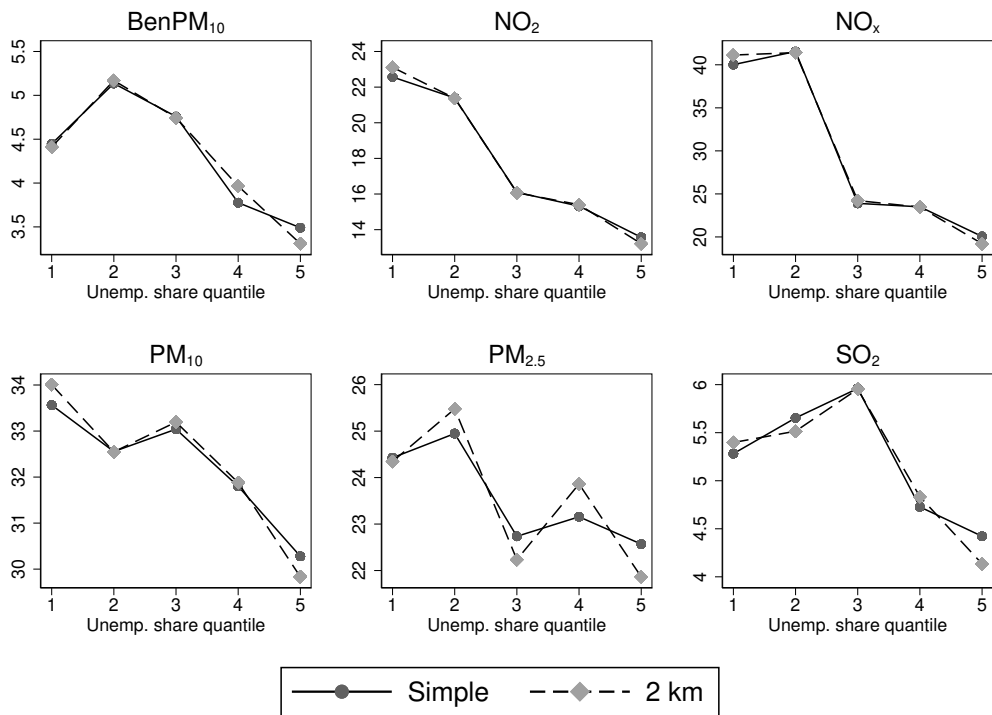


Figure 6: Quantile graphs for unemployment share and air pollution