Do local election results correlate with air quality?

Felix Schlechte
Matrikelnummer 5888672
felix.schlechte@student.uni-tuebingen.de

Pankaj Bora
Matrikelnummer 6946375
pankaj.bora@student.uni-tuebingen.de

Hendrik Jansen
Matrikelnummer 7019597
h.jansen@student.uni-tuebingen.de

Lennart Goldscheider
Matrikelnummer 5858772
lennart.goldscheider@student.uni-tuebingen.de

Abstract

Air pollution is bad for health. Political parties can implement policies that influence air pollution levels. We aim to find if there is a connection between air quality and election results at a local level. We use multivariate regression and fixed effects analysis to show that there might be a correlation between the local election results and change in air pollution levels over the next 5 years.

1 Introduction

In 2021, air pollution accounted for 8.1 million deaths globally [7]. In Germany, ambient PM2.5 was responsible for 27,040 deaths in 2019 [5]. Policies advocated by different political parties can have large variations, which can lead to contrasting effects on air quality. There have been studies on the effect of political institutions on air quality [2] and on how air pollution could influence elections [1]. However, there is no quantitative analysis on the impact of political parties on air quality at city level. We aim to find if there is a correlation between the results of local elections and the change in air quality after the election.
To do this, we create a panel dataset for German cities by combining relevant air quality and election results datasets. First we perform a multivariate linear regression, where for some parties and pollutants, we observe minor correlation. Then, to analyze the correlation at *entity* level - one entity corresponds to a specific station for the pollutants and the city for the Air Quality Index (index) - we do a fixed effects analysis. We observe correlation between the election results and PM10, PM2.5 and the index. Our analysis shows that there might be a correlation between local elections and air quality over the next 5 years. The complete code can be found here.

2 Data

2.1 Elections

The parties that we include in our data are: Die Linke (Linke), Bündnis 90/Die Grünen (Grüne), Sozialdemokratische Partei Deutschlands (SPD), Freie Demokratische Partei (FDP), Christlich Demokratische Union Deutschlands (CDU), Christlich-Soziale Union in Bayern (CSU) and Alternative für Deutschland (AfD). As the CSU only exists in Bavaria while the CDU exists in all states except Bavaria and the two parties form a single faction in the federal parliament (Bundestag), we

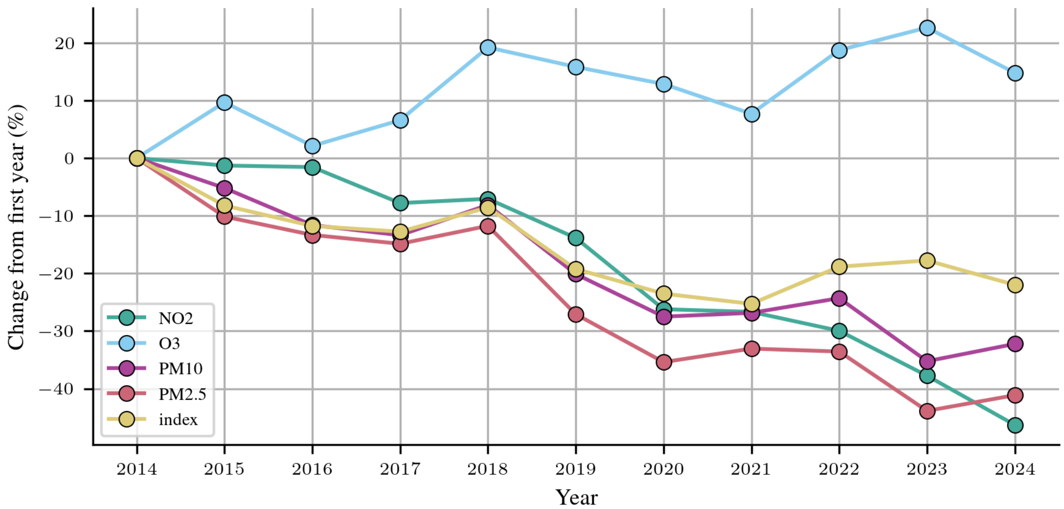

Figure 1: Trends of change in air pollution levels

consider them as the same party referred to as Union (Union).
The power a political party has in a city council is determined by the number of seats it holds, which are determined by local elections. As the AfD was founded in 2014 and to look at recent trends, we use data on the outcome of local elections in all cities with $\geq$ 20,000 inhabitants in the period of 2014 to 2024. This data is sourced from the State Statistical Offices (Statistische Landesämter). If available we use the percentage of seats won by a party, for states that don't report this number we use the percentage of votes cast for a party. Any elections where the result of these six parties combined is less than 67% are excluded from our data as outliers.

2.2 Air Quality

Our pollutants of interest are PM10, PM2.5, NO2 and O3. PM10 includes all inhalable particles smaller than 10 μm and PM2.5 all that are smaller than 2.5 μm. All four pollutants, especially PM2.5, are responsible for ailments such as asthma, COPD, lung cancer and can lead to death.
For Germany, the European Environment Agency (EEA) compiles air quality data for each pollutant [3] from the Umweltbundesamt as well as the state and regional environmental agencies. For the air pollution level we consider the Annual Mean/1 Year Average of the pollutant.
The EEA datasets have missing city values but provides the location coordinates for each measurement. We create a list of unique air quality stations across each pollutant with their coordinates. Using the GADM shapefile [4] for Germany, we can find the corresponding city for all the coordinates and fill the missing data. Additionally, we create a continuous index, based on the German Air Quality Index, in order to assess air quality as a whole.
Visualizing the percentage change in average air pollution levels, we observe a declining trend across all pollutants except O3 (fig. 1).

3 Methods

To show correlation without considering any external factors we use a multivariate regression model of the form [6, eq. 13.1]

$$Y_j = \alpha + \beta X_j + \epsilon_j, \tag{1}$$

where the index j stands for one election and pollution level pair.
To take into account some external effects, we use a fixed effects model of the form [6, eq. 16.3]

$$Y_{it} = \alpha_i + \alpha_t + \beta X_{it} + \epsilon_{it}. \tag{2}$$

The index i corresponds to one entity of our panel data and t to the specific time period. For regression we define $j := (i, t)$. We define X_{it} as a vector of election results for the air quality station i and time step t. So X_{it} corresponds to the independent variable. Let $year(X_{it})$ be the year of the election

	R^2	p-value	β-coefficients					
			Linke	Grüne	SPD	Union	FDP	AfD
Regression								
Index	0.235	**0.000**	**-0.54**	**0.48**	0.24	0.15	-0.74	**0.74**
PM10	0.187	**0.000**	**-0.48**	0.12	0.12	0.12	0.03	**0.56**
PM2.5	0.188	**0.000**	**-0.52**	0.32	0.10	0.35	-0.86	**1.06**
O3	0.135	**0.000**	-0.10	0.08	0.17	-0.03	0.18	**-0.39**
NO2	0.175	**0.000**	**0.19**	-0.05	**0.18**	**-0.53**	**0.30**	-0.11
Fixed Effects								
Index	0.5195	**0.000**	**-1.62**	0.59	-0.05	0.29	-0.24	0.34
PM10	0.1578	**0.000**	-0.75	-0.22	-0.15	-0.11	0.48	0.54
PM2.5	0.4452	**0.000**	**-1.96**	0.43	0.12	0.44	0.66	0.22
O3	0.0849	0.152	-0.20	0.07	-0.09	0.16	-0.37	-0.26
NO2	0.0441	0.095	-0.35	0.11	0.09	-0.07	0.24	0.24

Table 1: Resulting β coefficients of multivariate regression and fixed effects analysis. A coefficient indicates that for every percentage point the party gets, the percentage change of the pollutant changes by the amount of the coefficient. **Bold** values indicate $p < 0.005$ for the specific value.

and $pol\_level_i(year)$ the pollution level of entity i in the given year. To counter the declining trend, mentioned in fig. 1, we define the explanatory variable for a specific pollutant as the percentage change in the pollutant level

$$Y_{it} := \frac{pol\_level_i(year(X_{it}) + 5) - pol\_level_i(year(X_{it}))}{pol\_level_i(year(X_{it}))}. \tag{3}$$

We chose the offset 5, because most elections are held every five years and there might be a delay between implemented policies and their effects.
The error term ϵ_{it} captures effects, that vary between entity and time. The advantage of fixed effects is that we control for unobserved effects that stay constant over time for every entity (α_i) and for effects that stay constant for every entity, that just depend on the time step (α_t). We include these effects because factors such as the geography of a city, population, or certain laws might have an influence on air pollution, but are difficult to quantify. β are the coefficients that the analysis aims to estimate.
The H_0 for both methods is that $\beta = 0$. The p-values are computed by an F-Test. Since we do 10 tests in total, we adjust the significance level to 0.005 according to the *Bonferroni method* [8, p. 105]. Any data-points where the dependent variable is outside of two standard deviations of all data, are removed as outliers.
Both methods account for heteroscedasticity in the data with White's heteroscedasticity robust standard errors [9]. To test for multicollinearity, we compute the variance inflation factor (VIF) for each dependent variable in all analyses we run.

4 Results

In all cases the VIF stayed below 2.5, meaning that multicollinearity increased the variance of the estimated coefficients by a factor of less than 2.5. This is within commonly used bounds of acceptable multicollinearity.

4.1 Multivariate Linear Regression

The results for the index and all pollutants are significant, we therefore reject the H_0 that the coefficients of all parties are zero. There exists a correlation between local election results and the change in air quality in the following five years. However, there are only a few instances where we can reject that the coefficient of a particular party is zero, so the correlation between the result of an individual party and air quality change is not clear in most cases. We show data of significant results of individual parties for the index in fig. 2.

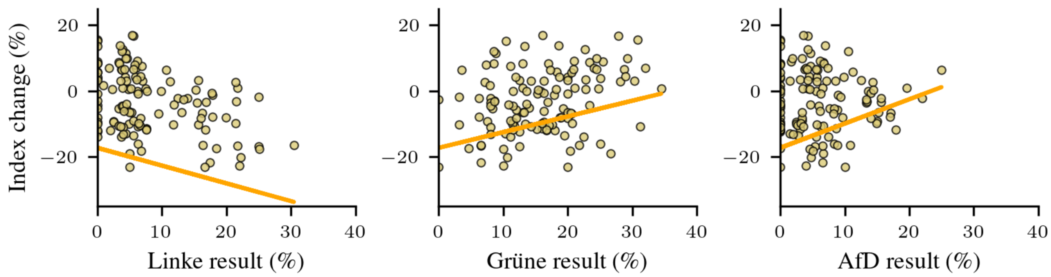

Figure 2: Plots of party election results (x-axis) with change of index five years after that election (y-axis) for parties with a significant coefficient. Also shows slice of the hyperplane fit by the multivariate regression (–).

4.2 Fixed Effects

In contrast to multivariate linear regression, there are only significant results for the index, PM10 and PM2.5, therefore we reject the H_0 for these pollutants. We find two significant coefficients for Die Linke in the analysis of the index and PM2.5 (tab. 1). Aside from that, the p-values are too high to make a statement about any correlation.

5 Limitations

A limitation of our data is that in some elections a party that should be included in our data forms a union with another small local party. In these cases the result for that party is reported as zero, which is not an accurate representation of its political power after the election. Also, our dataset might be too small, as we only consider cities with over 20,000 inhabitants and elections after 2014. Within these cities, some might not have any air quality stations. Also, the five year offset is arbitrary and using different offsets might change the results.
We assume a linear relationship between the election results and the air quality. In practice, the relationship between the distribution of city council seats, the policies implemented and air quality is much more complex.
The multivariate normality assumption and the independence of observations may be violated which might affect the accuracy of the coefficients. Also, since the expectation of the error term might not be 0 ([10, p.48] and [10, p.509]), both of our estimations might be biased.
Finally, we need to consider problems that arise from our analysis setting. One might be the omitted variable bias [10, p. 88]: Since we are excluding other relevant variables, such as being the decisive power of a government in the specific city, or the weather, our estimation might be biased. The consideration of control variables falls out of the scope of this paper.

6 Conclusion

Given the limitations, we observe some significant results for multivariate regression, which indicates a correlation between local election results and change in air pollution levels, but can't make any statements about the causes of that correlation. The fixed effects analysis provides evidence for some pollutants that this correlation also exist when taking the effects of individual years and stations/cities into account. All in all, we conclude that there might be correlation between local elections and air quality over the next 5 years.

7 Statement of Contributions

The data collection was divided equally between the group members. Hendrik and Pankaj performed the multivariate regression analysis. Felix and Lennart performed the fixed effects analysis. All members contributed equally to writing the report.

References

[1] Luna Bellani, Stefano Ceolotto, Benjamin Elsner, and Nico Pestel. "The political fallout of air pollution". In: *Proceedings of the National Academy of Sciences* 121.18 (2024), e2314428121. DOI: 10.1073/pnas.2314428121. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2314428121. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2314428121.

[2] Thomas Bernauer and Vally Koubi. "Effects of political institutions on air quality". In: *Ecological Economics* 68.5 (2009), pp. 1355–1365. ISSN: 0921-8009. DOI: https://doi.org/10.1016/j.ecolecon.2008.09.003. URL: https://www.sciencedirect.com/science/article/pii/S0921800908004345.

[3] European Environment Agency. *Air Quality Statistics Viewer*. [Online resource; accessed 14-February-2025]. European Union. 2025. URL: https://discomap.eea.europa.eu/App/AQViewer/index.html?fqn=Airquality_Dissem.b2g.AirQualityStatistics (visited on 02/14/2025).

[4] GADM. *GADM Data Download: Country-Level Administrative Boundaries*. Version 4.1. Available under Creative Commons Attribution 4.0. Global Administrative Areas. 2022. URL: https://gadm.org/download_country.html (visited on 02/14/2025).

[5] Omar Hahad, Jos Lelieveld, Sadeer Al-Kindi, Volker H. Schmitt, Lukas Hobohm, Karsten Keller, Martin Röösli, Marin Kuntic, and Andreas Daiber. "Burden of disease in Germany attributed to ambient particulate matter pollution". In: *Herz* 50.1 (2025), pp. 42–50. DOI: 10.1007/s00059-024-05269-8. URL: https://doi.org/10.1007/s00059-024-05269-8.

[6] Nick Huntington-Klein. "The effect: An introduction to research design and causality". In: (2021). URL: https://theeffectbook.net/.

[7] Health Effects Institute. "State of Global Air 2024. Special Report". In: (2024). URL: https://www.healthdata.org/research-analysis/library/state-global-air-2024.

[8] Neil J Salkind. *Encyclopedia of measurement and statistics*. SAGE publications, 2006.

[9] Halbert White. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity". In: *Econometrica* 48.4 (1980), pp. 817–838. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1912934 (visited on 02/14/2025).

[10] Jeffrey M Wooldridge. *Introductory Econometrics: A Modern Approach 5th ed.* 2013.