AGGLOMERATIONS, AIR QUALITY AND URBAN TRANSFORMATION

INAUGURAL-DISSERTATION

zur Erlangung des akademischen Grades
eines Doktors der Wirtschafts- und Sozialwissenschaften
(Dr. rer. pol.)
an der Wirtschafts- und Sozialwissenschaftlichen
Fakultät
der Universität Potsdam

vorgelegt von

Philipp Schrauth

Berlin, 2022

Erstgutachter:
Prof. Dr. Rainald Borck.

Zweitgutachter:
Prof. Dr. Ronny Freier.

Drittgutachter:
Prof. Dr. Marco Caliendo.

Eingereicht im Dezember 2022.

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-58608
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-586087

Acknowledgements

I want to thank my two main supervisors Rainald Borck and Ronny Freier for their constant support and the valuable feedback they provided. I highly regard the friendly relationship with both of them. Rainald was moreover a co-author and a very pleasant boss, who always facilitated and supported the balancing of family and work. I am also very grateful to Marco Caliendo for taking on the third review and for his repeatedly encouraging feedback to my work.

My PhD studies would not have been as joyful without my colleagues and friends from the University and beyond. I especially want to thank Johannes Endler, Maximilian Andres, Felix Degenhardt, Kalle Kappner, Relika Stoppel, Niklas Gohl, and numerous others for discussions and shared time. A particularly deep thanks goes to Niklas, who proofread my work, discussed it intensively with me and gave valuable suggestions. Moreover, I had the pleasure and honour of writing a paper together with him. I also want to thank all the colleagues from CEPA and the PRSE for their fruitful comments on the papers. Noah Koegler, Jana Hamdan and Max Schall always helped with excellent research assistance and Stephanie Massar as well as Annett Wadewitz have been consistently supportive in the administrative challenges of university life.

I also want to commemorate my dear friend Patrick Petit who left this world way too early and is greatly missed.

Last but not least, I want to thank my family. My parents and siblings constantly supported me in all my decisions. Most importantly, returning home from work to Amrei, Lene and Lisi always fills me with joy. I love you.

Zusammenarbeit mit Koautoren und Vorveröffentlichungen

Die vier Kapitel dieser Dissertation basieren auf Arbeiten, die zum Teil an anderer Stelle bereits veröffentlicht wurden. Aus der folgenden Auflistung werden Koautoren des jeweiligen Papiers sowie Ort und Art der Veröffentlichung erkenntlich.

`Kapitel 2`:

`Borck, R., Schrauth, P. (2021).` Population density and urban air quality. *Regional Science and Urban Economics, 86, 103596.*

`Kapitel 3`:

`Borck, R., Schrauth, P. (2022).` Urban pollution: A global perspective. *BSE Discussion Papers (No. 8). Submitted to Journal of Environmental Economics and Management.*

`Kapitel 4`:

`Gohl, N., Schrauth, P. (2022).` Ticket to paradise? The effect of a public transport subsidy on air quality. *Center for Economic Policy Analysis (No. 50). R&R at Journal of Urban Economics: Insights.*

`Kapitel 5`:

`Schrauth, P. (2022).` The causal effect of cycling infrastructure on traffic and accidents: Evidence from pop-up bike lanes in Berlin. *Center for Economic Policy Analysis (No. 48).*

Contents

Chapter 1

Introduction

General remarks & Motivation

The renowned economist Edward Glaeser proclaimed the "Triumph of the city", arguing that urban agglomerations make people "richer, smarter, greener, healthier, and happier" (Glaeser, 2011). This assertion is empirically supported by a large body of literature suggesting a positive correlation between population density and, for example, wages, patent intensity, and lower energy use (Ahlfeldt and Pietrostefani, 2019). Nevertheless, cities come at the cost of entailing severe diseconomies such as higher congestion, pollution, and traffic accidents (Glaeser, 1998). In the year 2018, about 55% of the world population was residing in urban areas, a share that is expected to grow to 68% by 2050 (United Nations, 2019). Consequently, the majority of people not only enjoy the benefits of metropolitan areas, but face the corresponding disadvantages. These are scientifically less well studied (Ahlfeldt and Pietrostefani, 2019). A deeper understanding of the interplay between cities, their costs, and policies that may transform urban life in response to current and future challenges is therefore pivotal.

This dissertation empirically assesses diseconomies of cities, and explores policies that offer possible solutions to them. Across chapters, the emphasis shifts from a causal relationship between urban agglomerations and air quality to the consequences of policy measures that are designed to transform cities toward greater sustainability. The policy interventions considered relate to the advancement and promotion of modes of transport beyond motorized private vehicles and their impact on air pollution or underlying causes such as congestion.

Air quality The importance of assessing (poor) air quality and its influencing factors in urban areas are illustrated by various examples. Historically, the "Great Smog of London" in the winter of 1952 provides a well-known instance of the detrimental effects of high pollutant concentrations on health related outcomes. According to the most recent studies, the event led to excess mortality of 50-300% (Bell and Davis, 2001). Such high pollutant concentrations are often caused by thermal inversions that result in lower wind speeds. As a consequence, industrially and domestically emitted pollutants, e.g. from coal or wood-fired ovens, do not disperse in space. In the case of London in 1952, this led to pollutant concentrations 3-5 times higher in comparison to the standard levels at that time (Anderson, 2009). The London fog was not the first of its kind. Another prominent incident happened in 1930 in Meuse Valley, Belgium. It was triggered by industrial pollution combined with weather inversion as well and resulted in a tenfold increase in expected deaths (Firket, 1936). The earliest established links between bad air quality and mortality reach back to times as early as the 17th century (Brimblecombe, 2012).

However, not only relatively short exposure to extremely high pollution concentrations lead to serious health damage and even death. A large range of scientific research shows that also lower pollution concentrations result in adverse health effects (Pope III and Dockery, 2006; Mayer, 1999). Numerous studies have quantified the relationship between air pollution and health related outcomes like cardiovascular diseases, low birth weight and infant mortality (e.g. Currie and Walker, 2011; Knittel et al., 2016; Margaryan, 2021).

Policymakers recognized the importance of poor air quality to human health and for the quality of life long ago. In England, the first legislation in this regard was put into effect with the "Smoke Abatement Act" of 1273 (Fowler et al., 2020). Over the past several decades, there have been increasingly more regulatory measures on national and international levels to curb air pollution. One prominent example is the 1963 Clean Air Act and its subsequent amendments in the United States, which set emission standards and provided control mechanisms to ensure compliance. Such interventions have been largely successful in reducing emissions of highly hazardous pollutants like Sulphur Dioxide (SO_2) (Godish and Fu, 2019). On international level, the World Health Organization has set and constantly updated guidelines regarding both short-term and long-term exposure to certain levels of pollutants. These are intended to indicate pollution concentrations that should be prevented in order to diminish negative health effects and often taken as guidelines by governments (World Health Organization, 2021).

Despite such efforts, poor air quality still imposes a large burden on health and the economy. Welfare costs worldwide are projected to rise from 3.8 trillion in 2015 to about 30.9 trillion US Dollars in 2060 due to air pollution. Most of these costs are caused by health impacts (OECD, 2016). Thus, further actions are needed to address these challenges.

Policy interventions Accordingly, general regulatory measures, such as the Clean Air Act, are commonly complemented by policy instruments that target behavioural aspects and directly or indirectly affect emissions. One focus of policy makers lies on the transport sector. This is important since air pollution is an unintended negative external effect of driving, which often is not taken into account by the polluter. The link between motorized traffic and poor air quality and consequent negative health effects has been empirically demonstrated repeatedly and shows the high welfare costs of motorized private transport. (Currie and Walker, 2011; Knittel et al., 2016; Margaryan, 2021).

Following economic theory, the first best solution to address the social costs of driving would be to internalize the external effects of traffic, e.g. through a Pigouvian tax. There were several attempts across various nations to implement a pricing scheme. For instance, a congestion charge has been shown to improve air quality in London (Green et al., 2020), while an electronic road pricing scheme has been shown to improve congestion levels in Singapore (Santos et al., 2004). Additionally, there exist quantity-based measures in a multitude of countries, like driving restrictions for certain types of cars (Fageda et al., 2022). One example of such regulations are low emission zones in Germany, that caused improvements in air quality by restricting the access of particulate emitting cars to city centres (Gehrsitz, 2017). However, the implementation of an effective road pricing or quantity scheme requires the comprehension of associated damages (Vickrey, 1969) and often is not very popular in the public opinion (Jaensirisak et al., 2005). Therefore, the focus of policy makers often concentrates on second best solutions that affect the mobility behaviour of citizens. This includes incentives to redirect car users to alternative modes of transport such as trains, subways, and more recently also bicycles.

A growing body of literature finds that infrastructural changes towards improving the public transport sector indeed positively impacts air quality. The emphasis mostly lies on extensions or the creation of subway systems (Chen and Whalley, 2012; Gendron-Carrier et al., 2022). In contrast, the effects of different pricing schemes for public transport have hardly been studied except for one analysis about price increases (Yang and Tang, 2018). Another way to reduce air pollution, other than acting directly on the public transportation system, is to improve congestion levels (Anderson, 2014; Bauernschuster et al., 2017). The reason why highly congested streets lead to poorer air quality is stop-and-go driving, which triggers the excess consumption of fuel and thereby additionally increases Carbon dioxide (CO_2) emissions (Treiber et al., 2008). The reduction of congestion moreover bears positive effects for other outcomes like accidents and improved travel times (Green et al., 2016). Time spent in traffic jams is significant, especially in cities. In London, for example, the average motorist spent almost 150 extra hours on the roads due to congestion in the year 2021 (INRIX, 2021). Due to such economic costs and the detrimental effects on the environment, congestion is an interesting outcome to consider on its own.

While public transportation networks and their effects have been extensively analysed, there is hardly any evidence on the effects of improvements for cyclists. Among the few exceptions are Hamilton and Wichman (2018), who show that the provision of bicycle sharing improves congestion within cities. Regarding road safety, Li et al. (2017) find that cycle superhighways decrease per-cyclist accidents. However, scientific research on bicycle infrastructure and its effects is still in its infancy.

Contribution and outline

This dissertation is comprised of four main chapters, which are distinctive, self-contained studies. Each of them uses a variety of methodological tools to empirically assess the respective research question. Apart from the individual chapter's specific contribution, which will be detailed below, the work generally adds to topics of urban environmental problems and policy interventions in the transportation sector in three fundamental ways: Firstly, it contains the first peer-reviewed publication that causally quantifies the influence of population

Table 1.1: Overview of Chapters

	Chapter 2	Chapter 3	Chapter 4	Chapter 5
Title	Population density and air quality	Urban pollution: A global perspective	Ticket to paradise? The effect of a public transport subsidy on air quality	The causal effect of cycling infrastructure on traffic and accidents: Evidence from pop-up bike lanes in Berlin
Co-Author(s)	Rainald Borck	Rainald Borck	Niklas Gohl	
Primary outcome(s)	Air pollution	Air pollution	Air pollution	Traffic speed & volume, accidents
Geographical focus	Germany	World	Germany	Berlin
Data	Ground-level air pollution, weather, administrative regional information	Satellite & administrative measures: air pollution, weather, economic, agricultural & trade data	Linked data from various sources: Ground-level air quality, gasoline prices, weather	Measurements of traffic volume & speed, accidents, construction sites, speed limit changes
Methodology	OLS, IV, Long-differences, LPM & Probit models	Descriptives, Within-country estimates, IV, Spatial first differences	Difference-in-Differences	Two-way fixed effects, event study design, synthetic control groups

density and agglomeration size on air quality and then extends this knowledge to a global scale. Secondly, using a quasi-experimental approach, it fills a research gap by assessing the effect of a large-scale price reduction for public transport on air quality. Thirdly, it contains novel empirical evidence about the effect of new bike lanes that replace an existing car lane regarding several aspects of road dynamics. Table 1.1 provides a stylized overview about the key facts of analyses on the individual chapters.

In the following, I will describe content and contribution of each individual chapter in more detail.

Chapter 2 - Population density and urban air quality Despite the fact that air pollution is a serious problem in many cities, the actual effect of population density and agglomeration size on air quality had not been causally quantified. There are few publications in fields other than economics that have established a link between urban agglomerations and air pollution (Gudipudi et al., 2016; Oliveira et al., 2014). In this chapter (co-authored with Rainald Borck), we aim to fill this gap by estimating the effect of population density on ground-level pollution, more precisely on Nitrogen Dioxide (NO_2), Particulate matters with aerodynamic diameter of 10 micrometer or less (PM_{10}), Particulate matters with aerodynamic diameter of 2.5 micrometer or less ($PM_{2.5}$), Ozone (O_3), and an Air Quality Index (AQI), which is an aggregate measure subsuming different pollutants that is standard in the literature.

Theoretically, the effect of density may affect pollution in two directions. On the one hand, more people produce more emissions due to commuting and housing. On the other hand, densely populated areas potentially benefit from a countervailing effect as densely built-up houses are more energy efficient than stand-alone houses. In addition, commutes to work in cities may be shorter and there are more alternative modes of commuting such as bicycles or subways.

In order to identify the effect, we use German panel data spanning from 2002 to 2015, and resort to a variety of methodological approaches. Firstly, we run Ordinary Least Squares (OLS) estimations, which are, however, likely to be biased e.g. due to omitted variable bias or reverse causality. Thus, in a second step we estimate long difference and fixed effects models, which allows to control for unobserved heterogeneity that is constant over time. Thirdly, we use an Instrumental Variable (IV) approach, accounting for endogeneity caused by reverse causality as well as unobserved time-varying determinants affecting emissions and density jointly, both of which cannot be solved by the fixed effects approach. Similar to Combes et al. (2010), we use soil quality as well as historical population data as instruments. In various robustness checks, we confirm the validity of our main results. We supplement the paper by testing whether densely populated areas have more days with extreme pollutant concentrations as defined by the World Health Organization (WHO), which tend to pose elevated risks on health. In order to do so, we use a Linear Probability Model (LPM) as well as Probit regressions.

Our preferred estimates stem from IV regressions and suggest an increase of pollution with population density with an elasticity of about 0.12 as measured for the AQI. This approximately corresponds to the average of the elasticities estimated for the single pollutants and implies that a one standard deviation increase of population density affects the AQI to increase by about 5%. Moreover, agglomerated areas suffer significantly more from extreme values above WHO thresholds than less densely populated locations. The empirical results are in line with our theoretical model, which is an extension of the monocentric city model.

Chapter 3 - Urban pollution: A global perspective The aim of the second chapter (co-authored with Rainald Borck) is to enhance the knowledge from Chapter 2 by analysing the effect of agglomeration size and population density on air pollution exposure using a global data set. Consequently, it provides a more holistic view and additionally circumvents problems that may arise when analysing single countries, like institutional peculiarities. The data stems from a large variety of sources including satellite measurements, privately supplied information, and administrative authorities. The study contributes to a small, yet growing literature on questions surrounding agglomerations and air pollution worldwide.

One of the main objectives is to highlight heterogeneities between countries. Thus, we want to know for instance which country characteristics influence the gradient between population measures and air quality. This is important, since the relationship between air quality and population patterns is likely to hinge on geographic or institutional factors, like environmental policies, that vary by region. One focus thereby lies on differences between city forms, which may be sprawling or densely built.

While we conduct instrumental variable regressions in order to justify that the outcomes resemble those from ordinary least squares, the main analyses use predominantly the latter in order to make use of the entire dataset. Within country regressions provide us with population-pollution gradients for almost all countries worldwide. Thereby, we are able to differentiate between an entire country using gridded data spanning the whole world and city differences using a global dataset of functional urban areas.

We document that about 75 percent of the world population faces particulate matter concentrations above WHO thresholds that are considered harmful to health, the largest proportion of whom live in cities. We then estimate the average pollution-population elasticity to be about 0.03 for $PM_{2.5}$ and 0.16 for NO_2. In combination with Chapter 2 this means that the gradient in Germany lies slightly above the worldwide average. Interestingly, we find that on a global scale, the effects are driven more by agglomeration size than population density. Moreover, the gradient is determined by larger commuting zones rather than by the population in urban centres. By means of a counterfactual simulation, we show that redistributing the population among cities of equal size in each country can lead to substantially lower pollution exposure and, consequently, lower health care costs for citizens.

Chapter 4 - Ticket to paradise? The effect of a public transport subsidy on air quality Chapter 4 is joint work with Niklas Gohl. We identify the causal effect of a large scale price subsidy for public means of transportation on air quality. More precisely, we use the implementation of the 9-Euro-Ticket (9ET), which reduced monthly public transport prices by up to 90% in Germany as a quasi-natural experiment in order to causally estimate its effects on an Air Quality Index (AQI) as well as on the pollutants NO_2, PM_{10} and $PM_{2.5}$ separately. To the best of our knowledge, this is the first study to estimate the effect of a substantial ticket price reduction in the public transport sector on air quality. Thereby, we contribute to the knowledge about external effects of traffic and how these can be addressed, especially in spatially constrained cities that do now allow for large infrastructural interventions.

We employ a Difference-in-Differences (DiD) approach to estimate the effects of this large and national-wide price adaptation on air quality. Our treatment group is the month of June, which was treated in the year 2022 unlike in preceding years. The control group is provided by the month May, which shares similar characteristics compared to June. Thus, the implicit assumption is that developments in air quality between May and June of 2022 should have been the same compared to prior years in the absence of treatment.

We find an improvement in air quality, as suggested by the AQI, by about eight percent. The effect is largest in urban areas, during working days, and where the provision of public transport is well established. We additionally estimate that the effect size slightly falls over time and that air pollution increases again after the end of the policy measure in September, 2022. Based on our estimates and using prior findings about the relation between air quality and health, we document that a reduced fare price for public transport is likely to improve health-related outcomes. Monetizing those leads us to conclude that they have the potential to amortise the actual costs of the intervention.

Chapter 5 - The causal effect of cycling infrastructure on traffic and accidents: Evidence from pop-up bike lanes in Berlin. In the last chapter of the dissertation, I analyse the effects of new cycling infrastructure on congestion, traffic volume, and accidents. New bike lanes have become a measure by local authorities all around the globe (e.g. in New York, London, Bogotá, Paris and many others) to make cities more attractive for cyclists and non-motorized use. Despite this global trend, analysing the effects of such infrastructural projects is still on the fringe of academic research.

Since cities are limited in space, building new cycling infrastructure often requires the transformation of street lanes that were formerly used by cars. Such an intervention may then affect road dynamics and lead to unintended consequences. This chapter considers congestion and accidents as outcomes due to their large negative impact on city life. Congestion leads to time losses for commuters, wasted fuel consumption and as a consequence an increase in CO_2 emissions and air pollution (Knittel et al., 2016; Schrank et al., 2015; Treiber et al., 2008; Vickrey, 1969). Thus, even though air quality is not separately considered in this study, the analysed outcomes are closely related to it. Additionally, accidents lead to substantial external costs (Edlin and Karaca-Mandic, 2006).

In order to causally assess the effects of interest, I apply variations of the classical difference-in-differences approach. The main outcomes rely on two-way fixed effect estimations, which condense the main effects into single, easy to interpret estimates and allow for the utilization of the most granular data available. These are supplemented by an event study design in order to track heterogeneities in the outcomes and to verify the common trend assumption. Synthetic control group estimations provide robustness tests whenever applicable. For identification, I use Pop-Up Bike Lanes (PUBLs), which were installed unexpectedly and allocated quasi-randomly. Due to their easy-to-implement and low-cost nature, PUBLs are especially interesting to look at, since they may be considered a viable way of constructing new cycling infrastructure in the future.

The results suggest an increase of congestion, as measured by a decrease of average vehicle speed, by 8-12 percent. In peak traffic hours, the effect is even larger. Since vehicle volumes are not significantly different between treatment and control streets, it is likely that the status quo bias causes motorists to stick to their accustomed routes. In combination with fewer space for vehicles, this leads to more congestion. Regarding absolute accidents, I do not find any effect, independent from cars or bikes being involved. However, different research suggests an increase of cycling caused by PUBLs (Kraus and Koch, 2021), which would mean an improvement of per-cyclist accident rates. Overall, I find that cycling infrastructure that replaces existing car lanes does not come without costs. Nevertheless, in perspective they seem to be moderate.

Chapter 2

Population density and urban air quality[1]

Abstract

We use panel data from Germany to analyse the effect of population density on urban air pollution (nitrogen oxides, particulate matter, ozone, and an aggregate index for bad air quality [AQI]). To address unobserved heterogeneity and omitted variables, we present long difference/fixed effects and instrumental variables estimates, using historical population and soil quality as instruments. Using our preferred estimates, we find that the concentration increases with density for NO_2 with an elasticity of 0.25 and particulate matter with elasticity of 0.08. The O_3 concentration decreases with density with an elasticity of -0.14. The AQI increases with density, with an elasticity of $0.11-0.13$. We also present a variety of robustness tests. Overall, the paper shows that higher population density worsens local air quality.

[1] Co-authored with Rainald Borck. This paper has been published in a slightly different version as *"Population density and urban air quality." Regional Science and Urban Economics, 86, 103596.*

2.1 Introduction

Are bigger and more densely populated cities better or worse places to live? Over the last centuries, the world has become more and more urbanized, as agglomeration benefits have drawn households to bigger cities. The urban economics literature on these agglomeration benefits is huge. Yet, in order to predict equilibrium and optimal sizes of cities, robust evidence is needed on the costs as well as the benefits of agglomeration, and much less seems to be known about the costs.[2] Kahn (2010) documents that in the United States (US), larger cities have longer commuting times, higher pollution levels and higher crime rates. We follow this line of research and analyse a particular element of the cost of agglomeration, namely the impact of population density on air pollution. As we document below, there is hardly any evidence that credibly estimates the causal effect of density on pollution. We aim to fill this gap.

Air pollution is an acute phenomenon in many cities worldwide. Megacities in developing countries suffer from particularly high pollution levels. But even in developed countries, where urban air pollution has fallen over the last decades, high pollution levels keep occurring. Cities in Germany and other European Union countries have been subject to a variety of legal proceedings against transgressions of pollution thresholds. Therefore, the relation between urban structure and pollution concentration is an essential policy issue.

Air quality is an important element for life in the city. Polluted air causes severe health problems, most notably heart diseases, strokes, chronic obstructive pulmonary disease, lung cancer, and respiratory infections. According to the World Health Organization (WHO), air pollution caused 600,000 premature deaths in Europe alone in 2010 and costs European economies US$ 1.575 trillion per year (World Health Organization, 2015). The European Environment Agency estimates that in Germany, Particulate matters with aerodynamic diameter of 2.5 micrometer or less ($PM_{2.5}$) caused 66,000 premature deaths in 2013.[3] Such numbers illustrate the potential economic benefits of using policies to reduce air pollution. The first best policy would be to internalize pollution externalities, e.g. through Pigouvian taxes or pollution licenses. In the absence of first-best prices, the effect of urban structure on pollution is obviously relevant to social welfare.

[2] Ahlfeldt and Pietrostefani (2019) synthesize research on the benefits and costs of population density.

[3] See https://www.eea.europa.eu/themes/air/country-fact-sheets/germany.

The pollutants we study are produced in a variety of industrial and non-industrial activities. Nitrogen oxides are produced by various combustion processes, predominantly by traffic with a share of about 38%. Other sources are agriculture, as well as power generation plants and different industries. Particulate matter is also emitted by various industrial processes as well as the combustion of fossil fuels for heating or power generation.[4] In addition, particulates arise from the dispersion of dust on the streets and tire wear of cars. Ground level ozone is created by chemical reactions between oxygen and Nitrogen Oxides (NO_X) (emitted for instance by cars) as well as Volatile Organic Compounds (VOC)[5]. Human activity is therefore the major source of bad air quality.

Adverse health effects are the main reason to worry about pollutant exposure (World Health Organization, 2003). For particulate matter, all levels of exposure may lead to negative health effects, but long-term threshold levels of 20 (PM_{10}) and 10 ($PM_{2.5}$) Microgram per cubic meter ($\mu g/m^3$) were set by the WHO in order to highlight concentrations, that are particularly harmful. The resulting diseases are mostly related to the respiratory tract and lung.[6] While for Nitrogen Dioxide (NO_2) older studies focus primarily on inflicted health impacts in animals (World Health Organization, 2006), more recent ones also find significant effects in humans (Costa et al., 2014). Nitrogen dioxide is moreover an important precursor for several other pollutants including Ozone (O_3), which have been shown to negatively affect health (World Health Organization, 2006). Ground-level ozone is linked to breathing problems, asthma, reduced lung function and respiratory diseases (Mücke, 2014).

The effect of city size or population density on air quality has only recently become the subject of research in economics and other disciplines, and the findings have partly been contradictory (see the next section). In addition, much of the empirical literature uses cross-sectional data, sometimes from several countries, which thwarts the causal interpretation of estimated coefficients. In this paper, we estimate the effect of population density on ground-level pollution (NO_2, PM_{10}, $PM_{2.5}$, O_3, and an air quality index – [AQI]) for German cities, using rich panel data from 2002 to 2015. We focus on these four pollutants for several reasons.

[4]There are also natural sources such as volcanoes, dust storms or wildfires.

[5]These are produced for instance in paints or in gasoline exhaust fumes, but are also emitted naturally by trees in forests.

[6]See e.g. Pope III and Dockery (2006) for a summary of the health effects of particulate matter. For one among many recent studies, see Lagravinese et al. (2014).

As discussed above, particulate matter and ozone have been identified to be particularly damaging to human health. Nitrogen dioxide has been the source of ongoing discussions about driving bans in Germany and other European countries.[7] It is interesting to study the effect of population density on these pollutants for two reasons: first, they may affect health differently (see above), and second, they may be differently affected by population density, which is indeed what we find (see below). Since in addition to individual pollutants, city residents and policy makers will be interested in overall air quality, we also study the effect of density on the aggregate AQI, which is also an indicator of local health risk from pollution (Van den Elshout et al., 2012).

We start by presenting Ordinary Least Squares (OLS) estimates. However, these may be biased due to omitted variables or reverse causality. Consequently, we additionally estimate Long Differences (LD) (and Fixed Effects (FE)) regressions to control for unobserved heterogeneity that affects density and pollution. Lastly, we run Instrumental Variable (IV) regressions using historical population measures as well as soil quality as instruments for current population density (see Combes et al., 2010). According to our preferred estimates from the IV regressions, population density increases NO_2 and PM_{10} concentrations with elasticities of 0.25 for NO_2 and 0.08 for PM_{10}. The effect of density on smaller particulates ($PM_{2.5}$) is less precisely estimated, which is due to the more recent and less extensive measuring net. However, if we consider $PM_{2.5}$ measured by satellite data, we find a density elasticity of 0.08 for these smaller particulates as well. For O_3, we find a negative density effect with an elasticity of -0.14.[8] The AQI index for bad air quality increases with density, with an elasticity of 0.11 for background and 0.13 for traffic stations. In summary, we find that population density worsens air quality in German cities. We perform a variety of robustness checks, which confirm our main findings.

The rest of the paper is structured as follows. The next section reviews the related literature. Section 2.3 presents some theoretical considerations on the link between population density and pollution concentration. Section 2.4 presents the data and Section 2.5 the estimation

[7] Other pollutants like Sulphur Dioxide (SO_2) are associated primarily with industrial production and their importance has subsided over the last decades. For most other pollutants, there is no developed measuring system.

[8] While NO_2 is a chemical precursor of O_3, the preconditions for ozone formation are more favourable outside of cities, see Footnote 43 below. This explains why we find that denser cities have higher NO_2 but lower O_3 concentrations.

methods. Our regression results are shown in Section 2.6, and the last section concludes.

2.2 Related literature

We contribute to the growing literature that examines the interaction of city structure and environmental pollution. On the theoretical side, Borck and Pflüger (2019) analyse the channels through which population size affects pollution. In general, pollution may increase or decrease with population size (see also the model in the next section). Larson and Yezer (2015) simulate the implications of city size and density for energy use. They find that per-capita energy use falls when population density increases as a consequence of urban amenities like greenbelts or relaxed building height limits.[9]

On the empirical side, several recent contributions have looked at the relation between city size or density and pollution. Some papers have looked at household energy consumption and mostly found that residents of denser cities consume less energy per capita (Glaeser and Kahn, 2010; de Thé et al., 2021). Potential reasons are that residents of densely populated cities may consume less fuel due to the availability of public transport systems and possibly shorter commutes. Reduced residential energy consumption may be attributed to smaller dwellings and more energy-efficient high-rise apartment buildings. Indeed, per capita fuel consumption and auto-mobile utilization have been found to be significantly lower in more densely populated cities due to the availability of public transport and shorter commutes to work on average (Newman and Kenworthy, 1989; Karathodorou et al., 2010). Research predominantly analysing the role of public transport shows that its availability improves air quality (Bauernschuster et al., 2017; Borck, 2019).

Looking at global rather than local emissions, several studies have examined the relation between city size and CO_2 emissions with conflicting results, e.g. Gudipudi et al. (2016) and Oliveira et al. (2014). Both of these papers use cross-sectional data. Borck and Tabuchi (2019) use panel data from US metropolitan areas. They find that per capita CO_2 emissions decrease with city size.

Another set of papers examines the effect of population size and other explanatory factors

[9]For papers that study population density and pollution in a welfare maximization framework, see Borck and Brueckner (2018) and Schindler et al. (2017).

on air quality. Population size has been found to be positively correlated with pollution for particulates (Glaeser, 1998) and NO_X (Lamsal et al., 2013; Sarzynski, 2012). Population density, however, has been found to be negatively correlated with NO_X (Sarzynski, 2012; Ewing et al., 2003; Hilber and Palmer, 2014) and PM_{10} (Hilber and Palmer, 2014). On the contrary, Ahlfeldt and Pietrostefani (2019) and Carozzi and Roth (2020) found a positive effect of population density on particulate ($PM_{2.5}$) concentration with an elasticity of around 0.13. These papers use different samples and methods, but with the exception of Hilber and Palmer (2014) and Carozzi and Roth (2020), all are based on cross-sectional OLS estimates.

To our knowledge, the only two papers other than ours that seriously tackle causality are the unpublished paper by Carozzi and Roth (2020) and the now defunct working paper by Hilber and Palmer (2014). While Carozzi and Roth (2020) use IV estimates with geological instruments and Fixed Effects (FE), Hilber and Palmer (2014) only use FE regressions. FE estimations, however, may be biased if there are time varying omitted variables that affect both density and pollution. The only paper besides ours that also uses instrumental variables is Carozzi and Roth (2020), who study the effect of population density on particulate pollution in US cities. The instruments they use – aquifers, earthquake risk, and soil drainage capacity – differ slightly from ours. Moreover, their main analysis is based on satellite data while ours is based on local monitor readings. The latter presumably measure pollution more accurately and also exist for a variety of other pollutants besides particulates. Still, the placing of monitors may be non-random. This could bias the estimates, which we try to account for in our estimation procedure.[10] Even if the method is similar, the two papers present estimates from the US and Germany, two countries with different city systems, energy use and pollution patterns.[11] Finally, our paper takes into account more pollutants than only $PM_{2.5}$ (namely PM_{10}, NO_2, O_3 and overall air quality measured by an index), so the findings of the studies can be viewed as complementary.

[10] To mitigate the latter problem, we will include some station characteristics, such as distance to city centre, station type and distance to main roads in our regressions (see below). Interestingly, Carozzi and Roth (2020) also use monitor reading data as a sensitivity check and find a slightly reduced effect of density on pollution. We also rerun our main regressions using satellite data and find results that are similar to those obtained with the monitor data, see Section 2.6.2.

[11] European and American cities differ along a number of dimensions. For historical reasons and different policies (for instance, planning policies, energy taxation, public transport investment), city structure, population density, commuting behaviour including distances and mode shares, housing patterns and energy usage all differ between the US and Europe (Nivola, 1999; Gordon and Cox, 2012). Therefore, whether and how density affects energy use and pollution differently is an open question.

In summary, we think that many existing contributions to the literature have only limited value in identifying causal effects of population on pollution. In fact, in a survey of the economics of density, Ahlfeldt and Pietrostefani (2019) argue that pollution is one of the areas where more evidence on the effects of density is needed.

2.3 Theoretical considerations

In this section, we present results from a simple urban economic model of city structure and pollution, building on Borck and Brueckner (2018), Borck and Pflüger (2019), Borck and Tabuchi (2019) and Larson and Yezer (2015). We combine central aspects from these models, and extend them insofar as here we focus on pollutant concentration instead of emissions. More details are in Appendix 2.A. Consider a circular open monocentric city with N workers who commute to the Central Business District (CBD) for work. Households have utility $v(c, q, \mathcal{P})$ over consumption, c, square meters of housing floor space, q, and pollution concentration $\mathcal{P}$ (see below). A household who lives at x km from the CBD incurs commuting costs tx and pays land rent $p(x)$. Mobility ensures that all households attain the same utility level throughout the city.

Housing is produced by profit maximizing developers using capital and land under perfect competition. They pay land rent $r(x)$ at distance x and an invariant price i per unit of capital. In equilibrium, land rent at the city border, $r(\bar{x})$ must equal the opportunity cost of land r_A. Let γ be the share of land in the city that is available for housing development at each distance x. We will use this parameter to induce a change in population density: as more land is available for housing, developers build more housing at each distance from the CBD, which leads to increasing densities.[12]

Let u be the outside utility residents can obtain in the rest of the economy. Then, population in the city adjusts through migration such that city residents obtain the same utility u everywhere. This canonical model produces a city where in the city centre buildings are tall, dwellings small and population density high.

We assume that emissions equal the sum of emissions from commuting and residential en-

[12]More precisely, the increased availability of land leads to buildings that are lower, but since more space is available for housing, more will be built in total at each distance from the CBD.

ergy.[13] Commuting emissions are assumed to be proportional to the sum of total commuting distances for all households, weighted by the emissions intensity of commuting; likewise, residential emissions are assumed proportional to total residential energy demand (itself assumed proportional to housing floor space), weighted by the emissions intensity of energy use. Pollution concentration in a city is the sum of total emissions divided by land area.[14]

Suppose now that we increase γ, the share of land available for housing. For instance, government might use zoning policies to increase the floor-area-ratio. As a result, more housing will be built and population density increases. Since increased housing supply reduces rents, residents' utility would rise, which induces in-migration from the outside economy to restore utility to the reservation level u. In the new equilibrium, the city boundary decreases, as in-migration is not strong enough to offset the increased housing supply. Since the city population has increased, its average density also increases. See Appendix 2.A for details.

What happens to pollution concentration? Total emissions rise, as more housing is built which increases residential energy use. Moreover, since there are more residents, aggregate commuting rises due to in-migration. In sum, aggregate emissions increase. Furthermore, since the city's land area decreases, pollution concentration increases. Hence, the model predicts that cities with higher population density will have a higher pollutant concentration.

However, this model ignores some possible countervailing forces. First, due to their higher density, bigger cities tend to have a more extensive supply of public transit due to economies of scale and density. Since transit typically produces lower emissions per person kilometre than automobiles, this would tend to decrease traffic emissions, all else equal (Bauernschuster et al., 2017; Borck, 2019). In this vein, de Thé et al. (2021) find that denser cities have better transit networks and lower car-related emissions. Second, in denser cities households especially in the city centre tend to live in high-rise buildings that are more energy efficient than the detached single family homes that dominate in sparsely populated cities (see Borck and Brueckner, 2018). In Appendix 2.A, we show, however, that while including these two forces in a stylized way attenuates the density-pollution relation, it does stay positive for

[13]Borck and Pflüger (2019) in addition consider emissions from industrial and agricultural production, and intercity goods transport. Note also that we abstract from congestion, see e.g. Larson and Yezer (2015).

[14]This assumption is for simplicity. In reality, how emissions diffuse over space and time is obviously a more complicated process.

realistic parameter values.[15]

In summary, we predict that, unless energy efficiency or public transit shares increase very strongly with density, higher population density should be associated with higher pollution concentration.

The model also highlights some issues in the estimation of the relation between pollution and density. There may be unobserved shifters of density that could be correlated with pollution. For instance, cities might have differing land use and environmental policies. If we cannot observe these policies and they are correlated with density, the estimation would be biased. Secondly, density is endogenous. As the model shows, density reacts to changes in exogenous parameters, such as agricultural land rent, but also to shocks to pollution concentration, which affects residents' utility and therefore leads to migration into or out of the city. We will address these issues in our estimation below.

The empirical literature has – largely descriptively – shown positive as well as negative correlations between density and pollution. To shed more light on this question, we empirically examine the relation between density and pollution for a panel of German communities in the next sections.

2.4 Data

We use administrative panel data from Germany for the period 2002–2015. While we have hourly data collected by monitoring stations for our pollutants of interest in Germany, the regional data we use, in particular population density, is available on a yearly basis for the roughly 400 German districts (*Landkreise*).

2.4.1 Emission data

We obtained hourly emission data from the Federal Environmental Agency (FEA) (*Umweltbundesamt*) for the years 2002–2015. This data is collected via a network of measurement

[15]In particular, in our parametrization, we assume that both commuting and residential emissions are constant elasticity functions of average density. As long as these elasticities are not lower than -1, we find that pollution increases with density. In their survey, Ahlfeldt and Pietrostefani (2019) suggest values of -0.07 for both.

stations throughout Germany for different pollutants.[16] Measurement stations are special monitors that lie either at streets and transport axes and measure pollution caused mainly by vehicles (traffic stations), or are dispersed throughout cities to record the overall level of city pollution at representative places (background stations). There are also stations close to industrial sites (industrial stations), but these are less numerous than traffic and background stations. The FEA classifies the areas in which the stations are located into rural, urban and suburban areas, which we explicitly control for in our analysis. Pollutants taken into account in this paper are Nitrogen Dioxide (NO_2), particulate matter with diameter less than 10 $\mu g/m^3$ (PM_{10}), particulates with diameter of less than 2.5 $\mu g/m^3$ ($PM_{2.5}$), and ozone (O_3). We also follow common practice and construct an Air Quality Index (AQI) from the single pollutant concentrations as described in more detail below.

The availability of average hourly emissions enables us to control for differences in emission patterns, for example due to differences between peak and off-peak periods and workdays versus weekends. We account for such temporal heterogeneities by including indicator variables for each day of week and each hour of day in the regressions. Furthermore, hourly emission data can be matched to weather information in more detail than lower frequency data, so we are better able to control for weather effects on emissions. The specific matching approach and the importance of taking into account weather variables are explained in Subsection 2.4.2 and in Appendix 2.C.

We have an unbalanced panel of stations and keep only stations with more than two years of observations such that we can apply long difference and fixed effects estimations.[17] In order to rule out the possibility that results are driven by seasonal forces that differently affect stations, we add a month dummy to our set of control variables.

We furthermore delete outliers from the sample. These are particulate matter values above 500 $\mu g/m^3$. Such high concentrations only occur if there is a large fire or another idiosyncratic source of pollution (for instance New Year's eve fireworks).

[16]Below, we also compare our main outcomes to those with pollution measures obtained from satellite data. See Section 2.6.2 for the results and Appendix 2.C for a description of the satellite data.

[17]We repeated our OLS and IV regressions without this restriction, but results were not affected.

Air pollution thresholds. In addition to average pollution concentration levels, we will also look at instances of transgressions of official thresholds.

Thresholds set by the European Union (EU) have entered into force in 2005 (PM_{10}) and 2010 (NO_2). Global guidelines by the World Health Organization (WHO) were updated in 2005.[18] Threshold values and their transgressions may be of particular interest, as they are based on evidence of the health effects of pollution. If health related problems increase non-linearly after the threshold is crossed, analysing these transgressions is of particular interest. Even if there are no non-linear health effects, there is a public interest in complying with them, since official bodies can be sued if the limits are exceeded. This has happened in the recent past in Germany and other EU countries.[19]

The guideline values for pollutant concentrations published by the WHO are 20 $\mu g/m^3$ for the annual mean concentration of PM_{10}, and 50 $\mu g/m^3$ for the 24-hour mean concentration. $PM_{2.5}$ is more aggressive to human health, so the threshold values are lower. The WHO recommends the annual mean pollution level to lie below 10 $\mu g/m^3$ and the 24-hour mean to be lower than 25 $\mu g/m^3$. For NO_2, the guidelines contain an annual mean value of 40 $\mu g/m^3$ and an one-hour mean value of 200 $\mu g/m^3$.[20]

Air quality index. We calculate the annual Air Quality Index (AQI), following Van den Elshout et al. (2012). As is common practice, we calculated the AQI for traffic and background stations separately, as they use different sub-indices for their computations. For traffic stations, the index is the average of the NO_2 and PM_{10} concentrations divided by 40, and a subindex, which takes into account the number of days where the PM_{10} concentration is above 50 $\mu g/m^3$.[21] For background stations, the index additionally contains an ozone

[18] The WHO has updated the guidelines in 2021 to lower thresholds (see https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution), after the publication of our study in the Journal *Regional Science and Urban Economics*. I stick to the 2005 guideline for the dissertation since over the period analysed the old thresholds were in place.

[19] See e.g. https://www.right-to-clean-air.eu/recht-und-klageverfahren/deutschland/klagen-und-urteile/.

[20] The guidelines set by the EU are less strict but binding for its member states. The EU has published an annual threshold of 40 $\mu g/m^3$ and an 1-hour threshold of 200 $\mu g/m^3$ for NO_2. The latter is allowed to be exceeded up to 18 times per year. For PM_{10}, there is an annual threshold value of 40 $\mu g/m^3$, while the 24-hour-mean should lie below 50 $\mu g/m^3$ with an allowance of 35 exceedances annually. For $PM_{2.5}$ there is only an annual threshold of 25 $\mu g/m^3$. For a full list of air quality standards in the EU, see https://ec.europa.eu/environment/air/quality/standards.htm.

[21] The maximum number of days with average daily values above 50 $\mu g/m^3$ allowed by the EU is 35 at the moment. The value of this sub-index is $\frac{\log(\text{Nr. of days}+1)}{\log(36)}$. See Van den Elshout et al. (2012) for a brief

subindex which accounts for the number of days with an 8-hour average value greater than or equal to 120 $\mu g/m^3$.[22] The higher the total AQI, the worse is the overall air quality. An index of less than or equal to one indicates compliance with EU standards on average, while an index above one indicates that air quality is worse on average than mandated by EU guidelines.[23]

2.4.2 Weather data

Ambient concentration of emissions is affected by weather conditions. As Auffhammer et al. (2013) argue, it is necessary to include a high amount of available weather variables in a regression, since they are themselves correlated over time and space.[24] Particulate matter for example is literally washed away on very rainy days or blown out of the city on very windy ones. The concentration of NO_2, in contrast, depends more on temperature and sunlight as it is decomposed into its elements more quickly by chemical processes on very hot days. Accordingly, the formation of O_3 occurs mainly on hot and sunny days in summer, as it requires the oxygen atoms of NO_2 as a crucial precursor.[25] The German Meteorological Service (*Deutscher Wetterdienst* (DWD)) provides data from its various weather and precipitation stations. We thus have hourly data on temperature, air pressure, rainfall, snowfall, sunshine, and wind. While Auffhammer and Kellogg (2011) and Wolff (2014) control for daily weather, we are able to match hourly weather variables with hourly emissions. The matching of emission monitors and weather stations is described in Appendix 2.C.

2.4.3 Other control variables

We include various additional control variables in our regressions. An important determinant of recorded pollution concentration levels is the physical location of a monitoring station.

discussion why the subindex is calculated like this.

[22]The subindex is calculated as $\frac{\#\text{days with 8-hour average}\geq 120}{25}$, because the EU target is not to exceed 25 days a year with values above 120.

[23]Since the AQI is an annual index by district, we do not control for station characteristics or weather variables.

[24]It might be that some weather variables are themselves affected by population density, for instance, if denser cities are hotter or more or less windy. Therefore, we also ran regressions without weather controls. However, we do not find that weather changes our results, which is why we control for it in all presented estimations.

[25]As Auffhammer and Kellogg (2011) note, ozone creation needs a certain amount of NO_2 and of other VOC. If climatic preconditions are not given, NO_2 levels therefore stay high. Furthermore, at great heat, plants are less able to absorb ozone, which increases ozone concentration in the air on very hot days.

We are able to control for a set of station-specific factors such as the distance to the CBD,[26] whether the station lies in an urban or a suburban area, the station type (traffic, background, or industry, see Section 2.4.1), and the distance to the closest major road (*Bundesstraße* or a street of similar size).[27]

In Germany, over the course of the past two decades, many cities introduced Low Emission Zones (LEZ) (*Umweltzonen*). Those zones were established in order to lower high levels of particulate matter by restricting access to the city to cars with particulate filters.[28] Using maps, we assign to each monitoring station an indicator for whether or not it lies in a LEZ. Including the emissions zone indicator makes sense as policy schemes that differ between cities affect city level pollution. For instance, Gehrsitz (2017) and Wolff (2014) found that such badges significantly lower PM_{10}-levels (but not other pollutants) within cities after their introduction.

In order to control for economic drivers of pollution, we can control for district level Gross Domestic Product (GDP), unemployment rate and average private household income within a district. Moreover, we collected the vote share for the Green Party to capture the potential sorting of 'green' households into cities. Another variable that we control for in the robustness section is the area of green space in a district. Green space may affect pollution in several ways. On the one hand, plants can absorb particles and thus mitigate pollution, and green space may lower temperatures which can also affect pollutant concentrations; on the other hand, plants and trees can generate VOC, which then react with NO_2.[29] As a consequence, O_3 concentrations may rise, while NO_2 concentrations may fall. We also have a measure of the number of public transit users as a share of total inhabitants per year.[30]

Lastly, we can also control for the presence of coal-fired power stations in a district and the distance of a monitoring station to a coal-fired power station.[31] Since burning coal leads to

[26]Our main geographic units are districts, which often contain several cities or towns. Therefore, we define the CBD as the centroid of the most densely populated municipality within a district. For district–free cities (independent towns), the CBD is defined as the centroid of the city.

[27]To construct the distance, we use maps provided by the Federal Office of Cartography and Geodesy (see http://www.geodatenzentrum.de/geodaten/).

[28]There are three different levels of LEZ: green, yellow and red with green being the most and red the least restrictive. Thus, these zones differ in the requirements of the quality of the particle filters of cars. We have the exact dates at which a red, yellow or green low emission zone was implemented.

[29]Natural sources of VOC are e.g. broadleaf trees and conifer, which produce VOC via evaporation.

[30]This data can be obtained on request from the German Federal Statistical Office.

[31]Since we do not have exact geo-coordinates of the power stations, we calculate the distance of the

high pollution levels, this might take out some variation that is caused by the presence of coal mines.

Some of our potential control variables may be 'bad controls', since they are themselves affected by density. We therefore choose to include such variables only in robustness checks (see Section 2.5 below). We can also think of some of these as mediating variables through which the effect of density works on pollution. We address this issue in more detail in Sections 2.5 and 2.6.2.

2.4.4 Descriptives

Table 2.1: Descriptives

	NO_2	PM_{10}	$PM_{2.5}$	O_3
Overall Stations	623	533	147	409
Background	360	290	80	340
Industrial	43	45	15	26
Traffic	220	198	52	43
Districts	269	247	109	251
Urban Districts	88	85	51	72
Labour Market Regions	128	125	77	126
Functional Urban Areas	63	61	38	53
Stations in LEZ	93	94	26	34
Whole Sample				
Mean Pollution	28.42	23.20	14.83	47.24
S.D. Pollution	15.52	5.964	2.963	10.10
Mean Pop. density	2590.2	2543.6	2383.0	2249.0
S.D. Pop. density	1337.6	1325.7	1322.5	1262.2

Note: Own calculations. The table shows descriptive data for each pollutant separately. Station specific pollution data is provided by the FEA upon request. Labour Market Regions are defined by Kosfeld and Werner (2012) and Functional Urban Areas by (Moreno-Monroy et al., 2019). LEZ are low emission zones. The table lists the number of all stations that were in the sample at least once between 2002 and 2015. Pollution/population means and standard deviations are calculated for the year 2015.

Table 2.1 presents monitoring stations in our sample and how they are distributed. The coverage of monitoring stations varies widely with NO_2 being measured by the most extensive net of monitoring stations (623), while $PM_{2.5}$ is measured by only 147 monitors, as monitoring of this pollutant only started in the mid 2000s with an extending network since

monitoring station to the centroid of the closest postal code region accommodating a coal-firing power plant. Postal code regions are relatively small administrative units with an average size of about 65 km^2.

then. The number of monitoring stations within the samples is also reflected in the number of districts, which are our main regional unit of analysis. In Germany, there are currently 401 districts including urban districts.[32]

In addition to using districts, we use Labour Market Regions (LMRs)[33] as defined by Kosfeld and Werner (2012) and Functional Urban Areas (FUAs) from Moreno-Monroy et al. (2020) to check whether results are driven by the geographical delineation of cities. Both definitions are metropolitan regions containing several districts with large commuting flows between them. As the table suggests, there are many LMRs and FUAs with at least one monitoring station for NO_2, while $PM_{2.5}$ stations only exist in about half of each defined area delineation.

To get a first visual impression of the pollution patterns and the way pollution is recorded, Figure 2.1 shows annual mean concentration levels of NO_2 and PM_{10} in 2015 and the distribution of monitoring stations in each district. Analogous maps for $PM_{2.5}$ and O_3 are shown in Figure 2.D.1. The small districts in the maps are mostly urban municipalities, which are more densely populated than other parts of the country. These areas also show high concentrations of pollution. Furthermore, the historical industrial regions in West Germany and the automotive centre around Stuttgart show high values of PM_{10} and NO_2. The figures reveal pollution patterns that are clearly not related to high population densities. For instance, PM_{10} shows high concentration levels in less urbanized districts in East Germany. These high levels might be caused by the proximity to coal-fired power stations in these areas. We will control for the presence of coal fired power plants in the robustness section of the paper.[34] Another possibility is the proximity to the Polish and Czech borders, where pollution standards are likely still lower than in Germany.[35]

Figure 2.2 depicts scatter plots for the four pollutants (logged mean pollution on district level) against logged population density. The figure shows that density is positively correl-

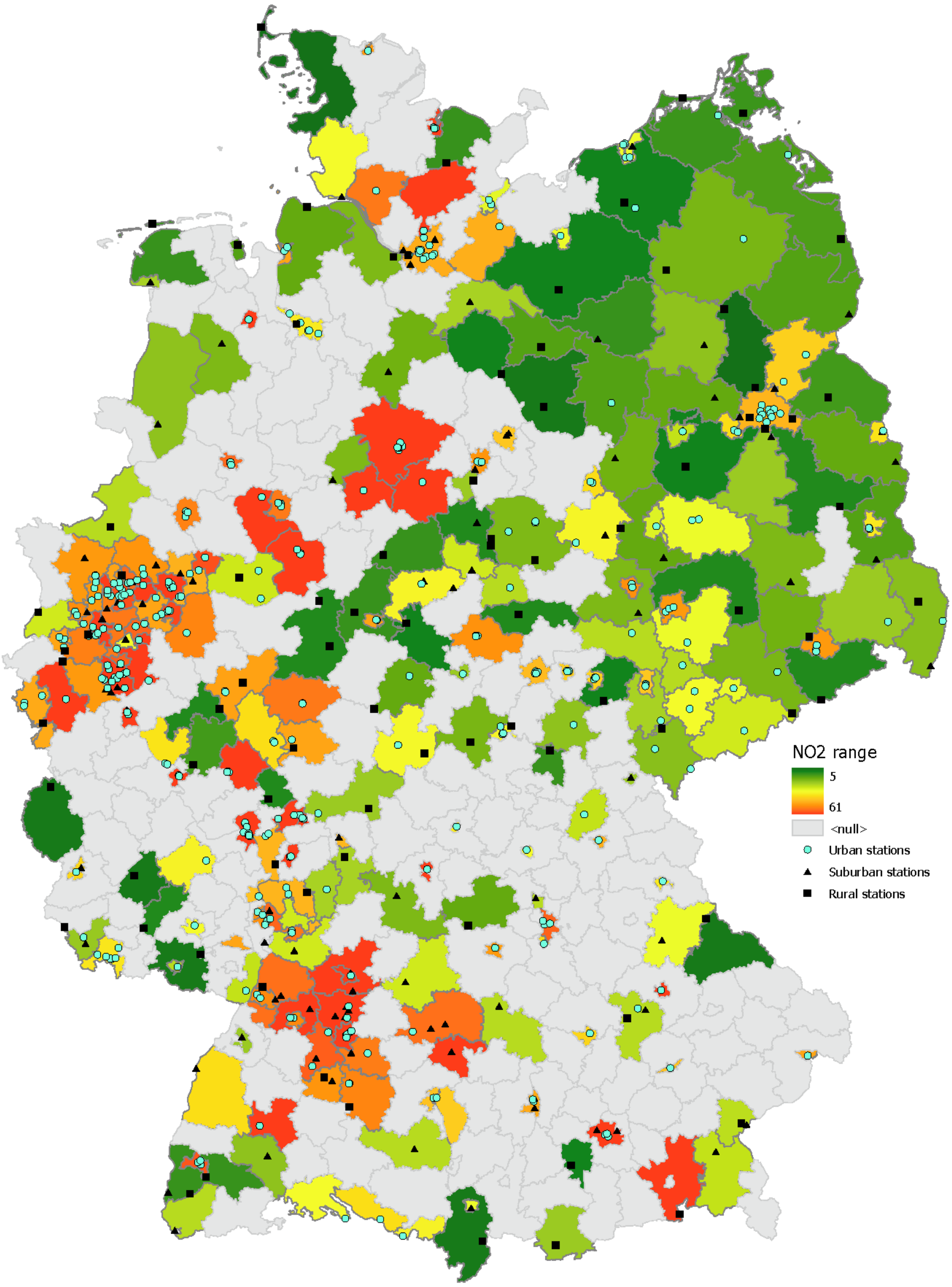
[32]The German administrative system distinguishes between districts (*Landkreise*) and district-free cities or urban districts (*kreisfreie Städte*). The latter are entities where the 'district' consists of a single (large) city, while *Landkreise* contain several jurisdictions.

[33]There are 141 LMRs of which we cover up to 128 in our analyses; the regions not covered do not contain a monitoring station for any pollutant.

[34]It is not immediately clear whether this variable should be included in the regressions: on the one hand, the energy mix might itself be driven by population, so one might want to leave the presence of coal fired power plants out. On the other hand, part of the location of these plants may be driven by the exogenous presence of coal mines. We therefore include coal fired power plants only in the robustness section; as will be seen, including this variable does not affect our results.

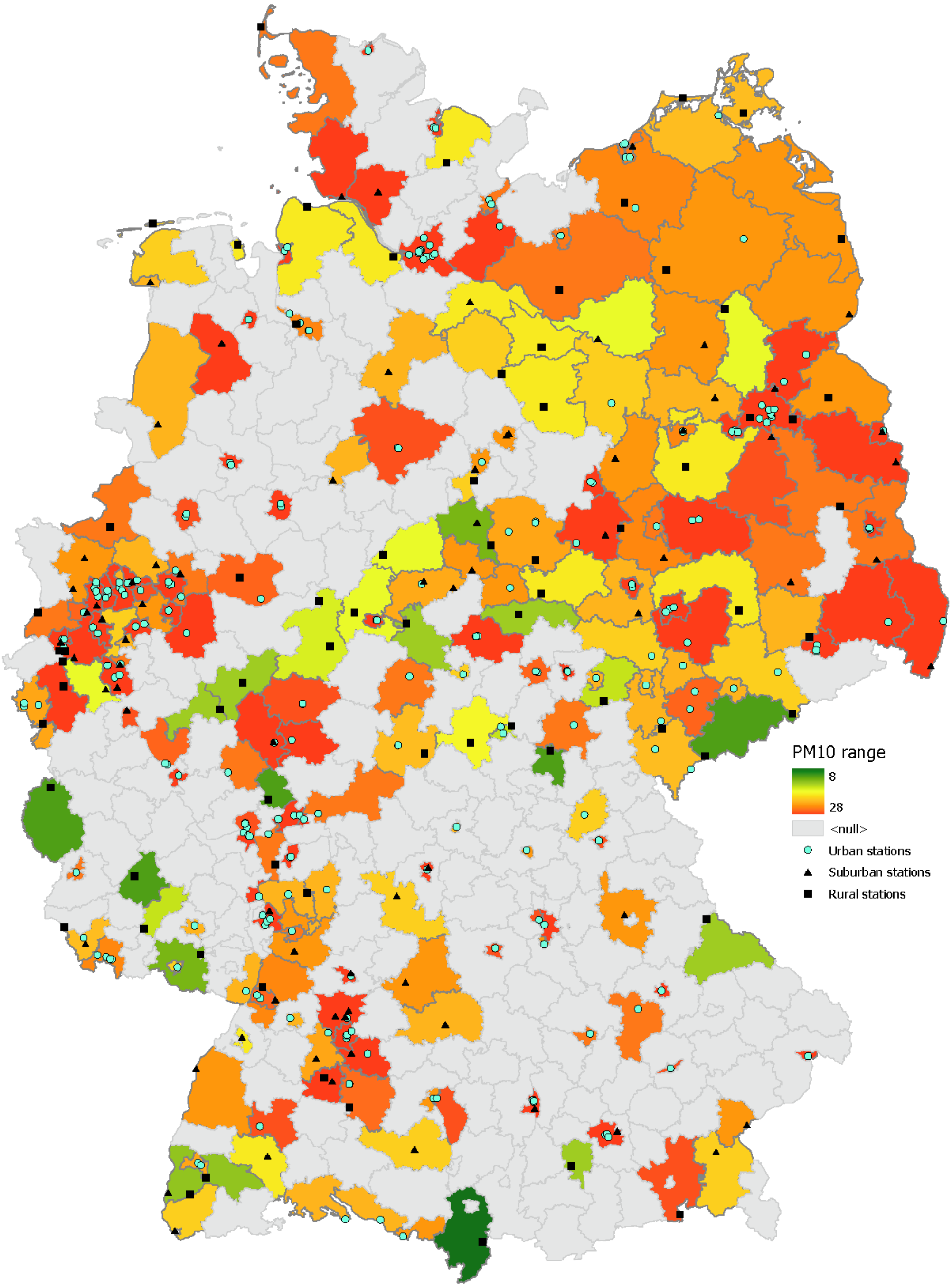
[35]See e.g. https://deutsch.radio.cz/streit-um-eu-grenzwerte-fuer-luftverschmutzung-8166941.

Figure 2.1: Monitoring stations and concentration levels in 2015 (NO_2)



(a) Sample of NO_2 stations

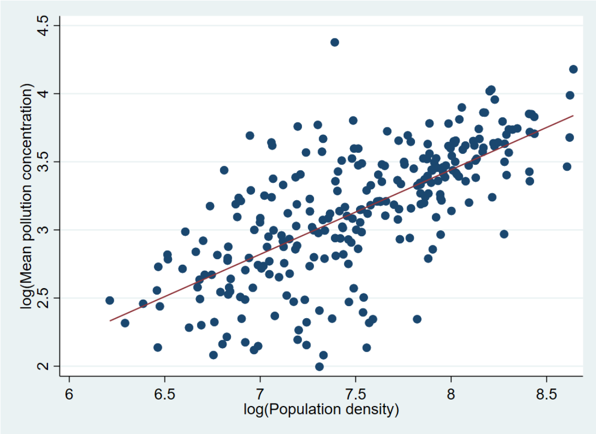
Figure 2.1: Monitoring stations and concentration levels in 2015 (PM_{10})



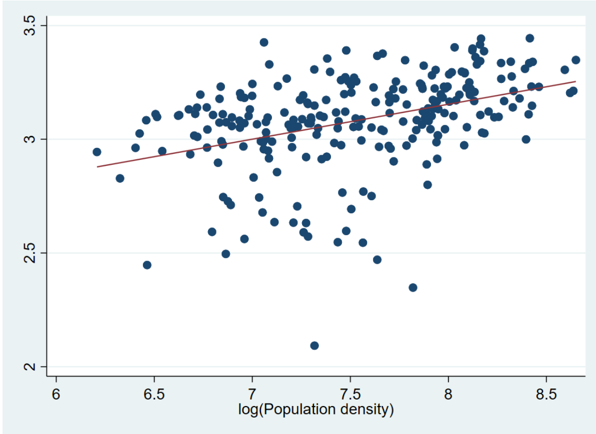
(b) Sample of PM_{10} stations

Note: Own calculations. The maps show average district-level pollution concentrations in 2015 for NO_2 and PM_{10} respectively. Pollution concentrations are relatively low in green coloured districts and relatively high in red coloured ones. Turquoise dots represent measuring stations that are marked as "urban" stations, black triangles lie in "suburban" districts and black squares depict "rural" stations. Grey coloured districts do not contain measuring stations.

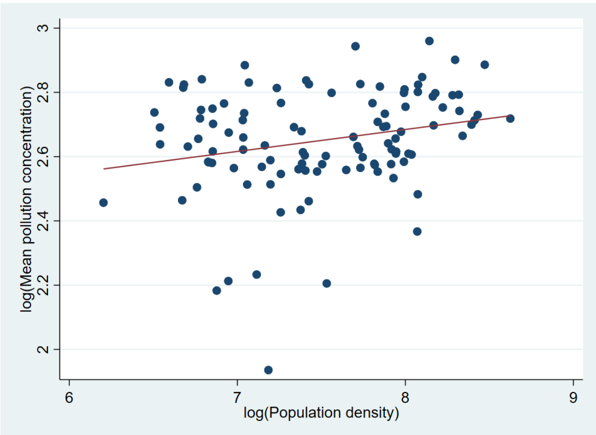
Figure 2.2: Scatter plots of log(Pollution) with log(Population density) and linear fits



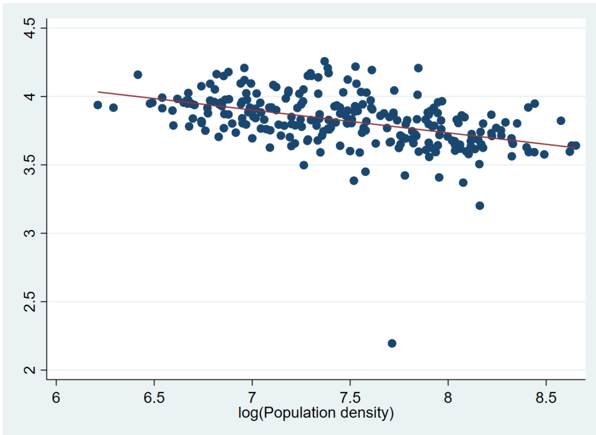
(a) NO_2



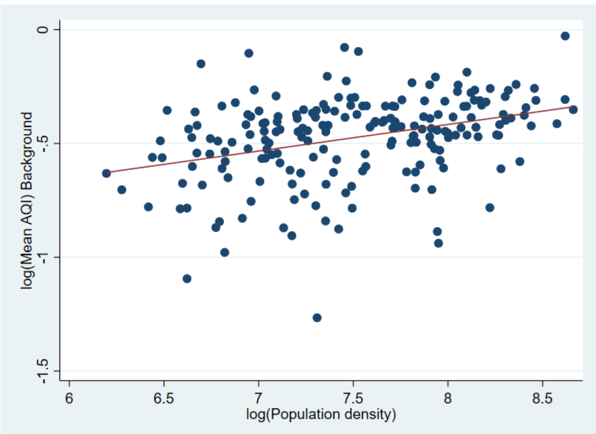
(b) PM_{10}



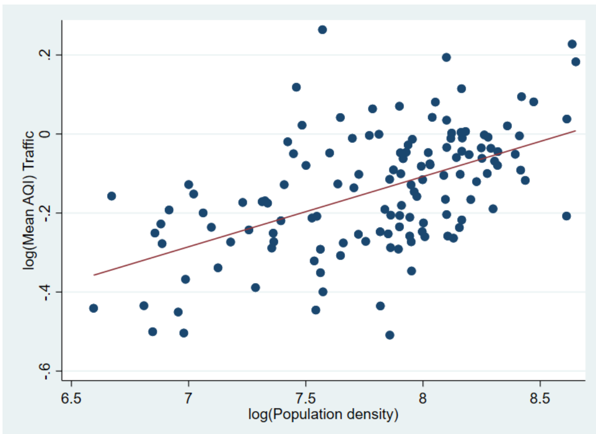
(c) $PM_{2.5}$



(d) O_3



(e) AQI at background stations



(f) AQI at traffic stations

Note: Own calculations. The graph depicts raw correlations and a linear fit of plotting the log of district-level mean concentrations of the respective pollutant against logged population density on district level for the most recent year in the data, 2015.

ated with NO_2 and particulates, although for the latter there is more noise. On the other hand, ozone concentration is negatively correlated with population density. Below, we will analyse whether those results hold up in multivariate regressions.

2.5 Estimation

2.5.1 Basic regressions

We now turn to estimating the model. While the pollution monitor readings are hourly data, our main variable of interest, population density, is available only annually. Therefore, and in order to reduce computational burdens, we first regress hourly pollution on hourly weather data, as well as time indicators. Following Auffhammer et al. (2013), the extensive set of weather and weather-interaction variables includes the hourly level of precipitation, sunshine, wind-speed, cloudiness, air pressure, and temperature at weather stations, as well as quadratic terms for precipitation, temperature, and wind, and a cubic temperature variable. We also interact temperature with wind and add hourly lags for temperature and precipitation. We include as further controls an indicator for day of week to take into account different patterns between days, an indicator for hour of day in order to control for special pollution patterns throughout the day (e.g. increased traffic during rush hours), and an indicator for the month of year for seasonal effects. We then take the residuals from this regression and aggregate them by year and district.

We proceed with these residuals by regressing pollution outcomes on a set of control variables using a simple OLS framework. Our first regression equation is

$$\ln(Y_{idt}) = \beta + \rho \ln(D_{dt}) + \gamma \mathbf{X}_{idt} + \alpha_t + \epsilon_{idt}. \tag{2.1}$$

Our dependent variable, Y_{idt}, is the residual concentration level (for a particular pollutant) in year t at station i in district d. Population density D_{dt} is available in yearly intervals at the district level. Our main parameter of interest is then ρ, which measures the elasticity of pollution concentration with respect to population density. Control variables $\mathbf{X}$ are attributes of the monitoring station like the station area (urban, suburban or rural), the distance to the next large road, and station type (background, traffic, or industrial). Therefore, we explicitly control for the location of a measuring station and its type. We also control for the distance of an emission station to the centre of the most densely populated municipality within a district. Effectively, our measure of pollution is then the pollutant concentration at the CBD. This should be a representative measure of city pollution. We also include year dummies α_t in order to control for business cycles and other time varying effects.

We can add economic controls at the district level like GDP, mean household income, and the unemployment rate. We are also able to take into account whether the station lies within an environmental zone with a red, yellow or green badge. The share of green party voters is used as a control for the sorting of households with 'green' preferences into low- or high-emission cities. There is moreover the concern that the presence of coal-fired power plants might cause bad air quality in some regions. Therefore, we constructed an indicator equalling one if such a power plant is located within the same district as the monitoring station. Additionally, we have a continuous measure of the distance between a measuring station and the closest coal-fired power plant.

In choosing whether to include control variables, we face two issues. On the one hand, leaving important drivers out of the regression will lead to omitted variable bias. On the other hand, some of these variables may be endogenous and therefore constitute "bad controls" that should be left out of the regression. For instance, income may be affected by density through agglomeration effects (even though the large empirical literature tends to find modest agglomeration economies, e.g. Combes et al., 2010). This also holds for most other potential controls. Green voting clearly may differ with a district's urbanity and also responds to local pollution. Coal fired power plants may be present in densely populated districts with large energy demand. The presence of LEZ and the share of transit users are also likely affected by population density. Therefore, we choose to present our basic regressions with controls only for the urban/suburban/rural indicator, station type, distance to major street, and distance to the CBD. As a sensitivity check, we analyse in Sec. 2.6.2 how our results change when we successively add controls.

We cluster standard errors on the labour market region-year level in our OLS regressions. According to Cameron and Miller (2015, p. 333), the consensus is to be conservative and avoid bias by using "bigger and more aggregate clusters when possible, up to and including the point at which there is concern about having too few clusters". When using clusters at district-year level, significance of the results does not change. However, we prefer spatial clustering by labour market regions as otherwise we have very few observations (monitoring stations) within some clusters.

OLS estimates would be unbiased and consistent if population density is not correlated with the error term, conditional on controls. However, this seems unlikely. For instance, densely populated cities may differ from less densely populated ones in their geography, industrial structure, or other unobserved variables that affect emissions.[36] Therefore, we also estimate Long Differences (LD) and Fixed Effects (FE) regressions of the form

$$\Delta \ln(Y_{idt}) = \rho \Delta \ln(D_{dt}) + \gamma \Delta \mathbf{X}_{idt} + \Delta \alpha_t + \Delta \tilde{\epsilon}_{idt}, \tag{2.2}$$

where $\Delta \ln Y \equiv \ln Y_T - \ln Y_F$. We run a LD estimation where $t = F$ represents the year 2002 and $t = T$ is 2015. In other regressions, we include all years in the sample to estimate FE. Our main LD regressions control for unobserved heterogeneity at the district level, but we also consider long differences at the station level (see Appendix 2.D). In addition to the controls described above, we again add year dummies α_t to the estimation. Unlike LD, fixed effects regressions take into account all years in the sample. However, we prefer the LD estimator, since the yearly within variation of population density is small. Nevertheless, the results of FE regressions differ only slightly in the size of the estimated coefficients.

Long difference estimation will be unbiased if the unobserved heterogeneity affecting density and pollution is time invariant. However, if there are unobserved time varying factors that affect emissions and are correlated with density changes over time, the LD estimates will be biased. For instance, it may be that household sorting leads to large cities getting 'greener' over time, which could be reflected by more use of bicycles instead of cars. In this case, density may still be correlated with the error term. Moreover, density and pollution may be simultaneously determined. For example, households may migrate out of very polluted cities, which leads to endogeneity of population density. Moreover, because variations in density and pollution are relatively small within short time periods, fixed effects estimates can be

[36]For instance, Stuttgart, one of the most densely populated cities, lies in a valley which makes it prone to high pollution concentrations.

imprecise. This is why we additionally estimate Instrumental Variable (IV) regressions:

$$\ln(D_{dt}) = B_0 + B_1\mathbf{X}_{idt} + B_2 Z_{dt} + \eta_{idt} \tag{2.3}$$

$$\ln(Y_{idt}) = \Gamma + \rho\widehat{\ln(D_{dt})} + A_1\mathbf{X}_{idt} + \hat{\epsilon}_{idt}. \tag{2.4}$$

In the first stage regression (2.3), density is regressed on one or more instrumental variable(s) Z. The IV estimations will be valid if the instrument strongly predicts density but is not correlated with the error term in the second-stage regression (2.4).

Following Combes et al. (2010), we take both historical population data and soil quality as instruments. There is a long tradition of using historical population data, beginning with Ciccone and Hall (1996). Our instrument for current density is the log of historical population density from 1910.[37] Historical population data are relevant, since urban population tends to be strongly persistent over time. The exclusion restriction requires that historical density be correlated with current emission levels only through its effect on current density. We believe this to be the case, since pollution in the early 20th century was driven largely by industry. Today's urban pollution is much more driven by auto-mobile traffic, which was close to non-existent in 1910.[38] The German emperor Wilhelm II is purported to have said around 1900: "I believe in the horse. Automobiles are no more than a transitory phenomenon".[39] Furthermore, industry structures have changed dramatically over time. Therefore, it seems unlikely that historical population patterns should directly affect current pollution. We address this concern further below.

In addition to historical population measures, we instrument current population densities with data on soil characteristics. Some soil types are better suited for the construction of tall and heavy buildings for a large number of dwellings or offices. Furthermore, households have historically been attracted to settle in areas with fertile land. Henderson et al. (2018)

[37]The authors would like to thank Uli Schubert from http://www.gemeindeverzeichnis.de/ for sharing his data on population in 1910.

[38]See, e.g. Koh et al. (2013) and Redding and Sturm (2008) who use similar historical data for Germany. Note that there is no consistent population data for earlier years covering *all* districts, so instead of using incomplete data going further back in time we choose 1910 to have a complete IV.

[39]There is a dispute about the correctness of the quote. While it is referenced in the Mercedes-Benz-Museum in Stuttgart (see https://commons.wikimedia.org/wiki/File:Mercedes-Benz-Museum_in_Stuttgart,_Zitat_am_Beginn_des_Rundganges.jpg), there is doubt about the historical correctness (see https://falschzitate.blogspot.com/2018/01/das-auto-ist-eine-vorubergehende.html). However, the latter source claims that the "transitory nature" of cars has demonstrably been stated as early as the 1960s.

argue that agricultural variables are the most important drivers of agglomeration. Therefore, soil characteristics should be important determinants of historical and current population patterns. For these variables, the exclusion restriction may be easier to justify (Combes et al., 2010). First, geology is determined by nature and is thus independent of human economic activity. Second, since agriculture accounts for less than 2% of current employment in Germany[40], soil characteristics should not be important drivers of current pollution levels. We return to the issue of instrument exogeneity below and provide some additional tests in support of it.

Soil characteristics are taken from the European Soil Database (ESDB). For the choice of variables, we follow Combes et al. (2010) who look at French regions. We consider only variables that tend not to be influenced by human activity and therefore should be exogenous to it. In particular, we use soil characteristics that describe the mineralogy of the topsoil and the subsoil as well as the dominant parent material of the soil. The dominant parent material describes the bedrock of the soil, which is the underlying geological material. Mineralogy captures the presence of minerals in the different layers of soil. We also include information about the topsoil organic carbon content and the soil profile differentiation.[41] Lastly, we include the ruggedness of a district, which is a measure of the average local variance in elevation within a district (Nunn and Puga, 2012).[42] More detail on the soil data can be found in Appendix 2.C. In all of our instrumental variable regressions we cluster standard errors at the labour market region level. Since all our instruments are time invariant, clusters at year level are excluded.

2.5.2 Threshold regressions

To test whether extreme values of PM_{10}, $PM_{2.5}$ or NO_2 correlate with population density, we run further regressions. We use the same basic approach as in Section 2.5.1, but now

[40]See e.g. https://www.destatis.de/DE/Themen/Wirtschaft/Konjunkturindikatoren/Lange-Reihen/Arbeitsmarkt/lrerw13a.html.

[41]Due to the limited number of observations, we combine the high and medium categories of soil carbon contents into one category. The categorical values of topsoil mineralogy cannot be combined, as they are not ordinal. Thus, we replace the value which occurs only once in the dataset with a missing.

[42]All variables that we consider as instruments have at least one category significant at the 10-percent level or higher in the first stage regression. Variables with no significant category (water capacity at the topsoil and the subsoil, depth to rock, soil erodibility class, and hydrological class) are not included in the regressions. However, including those variables does not alter the quality of the second-stage results, but leads to the instruments getting weaker and sometimes over-identified.

our dependent variable is a dummy indicator switching to one when the threshold value was violated and zero otherwise. Threshold annual mean values indicate constant long-term exposure to air pollution. However, there could also be elevated pollution concentrations on individual days during the year, leading to serious health effects. Thus, we furthermore examine whether densely populated areas tend to have more days with threshold violations (24-hour means). In order to do so, we created dummy variables which equal one when a station exceeded a predetermined number of days within a year. The outcome then is the probability of exceeding the pollution thresholds by a certain number of days within a year. The limit values we consider are those set by the WHO air pollution guidelines (see Section 2.4.1), while the number of allowed threshold exceedances are defined by the EU. The hourly NO_2 threshold is allowed to be exceeded on 18 days during a year and the PM_{10} threshold on 35 days. For $PM_{2.5}$, there is no short-term threshold in the EU (unlike the WHO) and thus no daily violations limit exists. In this case, we resort to the allowed number of daily exceedances for PM_{10}. Local authorities could try to take short-term measures to avoid illegal threshold exceedances. Even then, however, they would still be exposed to high pollution levels, so looking at threshold exceedances just below the limits is of interest. This is why we examine the probability of transgressing the limit values on a set of days just below the EU allowances (17, 14, and 9 days for NO_2, 34, 29, and 24 days for PM_{10} and for $PM_{2.5}$).

We use Linear Probability Model (LPM) to estimate our outcomes of interest. With this approach, we can easily apply instrumental variable regressions. The LPM does a decent job in estimating the probabilities, as the occurrence of transgressing the threshold is relatively dispersed over the sample. However, we also run Probit regressions to account for potential non-linearities in the probability of transgressions (see Section 2.6.2).

2.6 Results

2.6.1 Basic results

OLS regressions. We present our basic cross-sectional OLS results in columns (1) and (5) of Table 2.2 and Table 2.3. The tables present coefficients for our parameter of interest, log of population density, as well as coefficients of basic controls (distance to CBD, distance to major street, whether the station lies in an urban or suburban area – rural is the reference category –, and the traffic and industrial station dummy – the reference category being background).

Table 2.2: OLS and IV regressions for NO_2 and PM_{10}

	NO_2				PM_{10}			
	(1) OLS	(2) IV Density 1910	(3) IV Soil	(4) IV 1910 & Soil	(5) OLS	(6) IV Density 1910	(7) IV Soil	(8) IV 1910 & Soil
log(pop density)	0.280***	0.191***	0.292***	0.251***	0.0749***	0.104***	0.0662*	0.0787***
	(0.0141)	(0.0544)	(0.0600)	(0.0575)	(0.00691)	(0.0239)	(0.0366)	(0.0247)
Distance to CBD	-0.00321***	-0.00477***	-0.00299**	-0.00378**	0.000947***	0.00115	0.000809	0.000744
	(0.000447)	(0.00152)	(0.00139)	(0.00147)	(0.000297)	(0.000906)	(0.000937)	(0.000884)
Distance to Street	-0.105***	-0.102**	-0.103***	-0.101***	-0.0405***	-0.0350*	-0.0392**	-0.0333*
	(0.0115)	(0.0397)	(0.0377)	(0.0387)	(0.00648)	(0.0188)	(0.0199)	(0.0192)
Suburban	0.281***	0.281***	0.283***	0.279***	0.0761***	0.0702***	0.0782***	0.0728***
	(0.0169)	(0.0500)	(0.0501)	(0.0509)	(0.00948)	(0.0251)	(0.0263)	(0.0256)
Urban	0.445***	0.476***	0.443***	0.458***	0.140***	0.119***	0.144***	0.129***
	(0.0216)	(0.0678)	(0.0693)	(0.0716)	(0.0108)	(0.0324)	(0.0330)	(0.0315)
Industrial	0.0898***	0.0890**	0.0972***	0.0969***	0.136***	0.131***	0.139***	0.134***
	(0.0128)	(0.0368)	(0.0330)	(0.0345)	(0.0127)	(0.0364)	(0.0372)	(0.0374)
Traffic	0.648***	0.656***	0.647***	0.652***	0.255***	0.257***	0.257***	0.260***
	(0.0124)	(0.0402)	(0.0385)	(0.0402)	(0.00672)	(0.0186)	(0.0182)	(0.0187)
N	5575	5301	5547	5273	4648	4407	4620	4379
R^2	0.755	0.751	0.754	0.754	0.474	0.463	0.476	0.468
Districts	269	269	269	269	247	247	247	247
Soil Characteristics	No	No	Yes	Yes	No	No	Yes	Yes
First-stage F-statistic	–	318.5	11.79	46.72	–	383.3	10.93	48.64
Hansen p-statistic	–	–	0.0681	0.0657	–	–	0.230	0.427

Note: The table presents OLS and IV model outcomes of regressing the respective pollutant concentration (NO_2 and PM_{10}) on log(population density). Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), average district-level GDP, average income, share of green party voters, the unemployment share, a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial), and whether the station lies within an environmental zone or not. The instruments used are historical population from 1910, soil characteristics, and a combination of both. Standard errors in parentheses are clustered at labour market region - year (OLS) and labour market region (IV) level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As shown by the OLS regression in column (1) of Table 2.2, the density elasticity of NO_2 concentration is 0.28 and the estimate is significant at 1%. The mean value of population density in 2015 was 2590.2 with a standard deviation of 1337.6. Thus, a one standard deviation increase in population density within a city increases the NO_2 concentration by 12.4 percent. For PM_{10}, we find a smaller elasticity of 0.075, which is significant at 1% (column (5) of Table 2.2). A one standard deviation increase in population density increases the PM_{10} concentration by 3.2%.

For $PM_{2.5}$, the estimated elasticity is 0.035 which is significant at the 5% level (column (1) of Table 2.3). A one standard deviation increase in population density increases the $PM_{2.5}$ concentration by 1.48 percent. Note that the coefficient is now lower than the corresponding

Table 2.3: OLS and IV regressions for $PM_{2.5}$ and O_3

	$PM_{2.5}$				O_3			
	(1) OLS	(2) IV Density 1910	(3) IV Soil	(4) IV 1910 & Soil	(5) OLS	(6) IV Density 1910	(7) IV Soil	(8) IV 1910 & Soil
log(pop density)	0.0353**	0.0820	0.0489	0.0204	-0.177***	-0.0931***	-0.217***	-0.143***
	(0.0161)	(0.0579)	(0.0658)	(0.0519)	(0.00945)	(0.0317)	(0.0503)	(0.0342)
Distance to CBD	0.00122*	0.00164	0.00137	0.000618	0.00230***	0.00354***	0.00172*	0.00279***
	(0.000651)	(0.00158)	(0.00166)	(0.00149)	(0.000315)	(0.00102)	(0.00102)	(0.000982)
Distance to Street	-0.0408**	-0.0431	-0.0341	-0.0274	0.0522***	0.0474**	0.0533**	0.0495**
	(0.0172)	(0.0394)	(0.0421)	(0.0440)	(0.00727)	(0.0241)	(0.0246)	(0.0243)
Suburban	0.124***	0.115*	0.135**	0.144**	-0.139***	-0.143***	-0.135***	-0.138***
	(0.0294)	(0.0617)	(0.0654)	(0.0648)	(0.0117)	(0.0347)	(0.0371)	(0.0349)
Urban	0.161***	0.123*	0.159*	0.180**	-0.193***	-0.219***	-0.180***	-0.204***
	(0.0330)	(0.0745)	(0.0820)	(0.0735)	(0.0138)	(0.0431)	(0.0484)	(0.0443)
Industrial	0.0619***	0.0659	0.0814**	0.0928**	-0.0754***	-0.0501	-0.0855***	-0.0651**
	(0.0210)	(0.0426)	(0.0403)	(0.0390)	(0.0116)	(0.0333)	(0.0286)	(0.0291)
Traffic	0.109***	0.109***	0.117***	0.112**	-0.231***	-0.238***	-0.232***	-0.238***
	(0.0173)	(0.0418)	(0.0423)	(0.0450)	(0.0164)	(0.0411)	(0.0380)	(0.0393)
N	795	758	780	743	3776	3588	3756	3568
R^2	0.254	0.227	0.256	0.239	0.445	0.426	0.440	0.440
Districts	109	109	109	109	251	251	251	251
Soil Characteristics	No	No	Yes	Yes	No	No	Yes	Yes
First-stage F-statistic	–	114.5	7.909	25.12	–	385.4	12.88	47.54
Hansen p-statistic	–	–	0.0447	–	–	–	0.0124	0.00767

Note: The table presents OLS and IV model outcomes of regressing the respective pollutant concentration ($PM_{2.5}$ and O_3) on log(population density). Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), average district-level GDP, average income, share of green party voters, the unemployment share, a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial), and whether the station lies within an environmental zone or not. The instruments used are historical population from 1910, soil characteristics, and a combination of both. Standard errors in parentheses are clustered at labour market region - year (OLS) and labour market region (IV) level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

one for PM_{10}. However, the net of monitoring stations is both more recent and less dense, so the estimates for $PM_{2.5}$ are much less precise. Using a z-test, we cannot reject the hypothesis that the two coefficients are identical. Furthermore, below in the robustness section we re-estimate the regressions using satellite measurements of pollution concentration. Since these cover the entire country, the estimates are much more precise. Interestingly, the IV estimate for $PM_{2.5}$ with satellite data is 0.08, just like the IV estimate for PM_{10} using station data.

Ozone concentration is negatively correlated with population density. This is likely due to the fact that the chemical prerequisites for ozone formation are more favourable outside large cities.[43] The density elasticity of ozone concentration is -0.18. A one standard deviation increase in population density decreases the O_3 concentration by 7.1 percent. This result is

[43] This is because Nitrogen Monoxide (NO), which is contained in car exhaust fumes, reacts with ozone to NO_2. Ozone is split into O_2 and NO_2 such that ozone pollution in city centres is significantly lower. Moreover, the ozone precursors are transported out of cities by wind and contribute to the formation of ozone away from their actual sources. See https://www.umweltbundesamt.de/daten/luft/ozon-belastung#textpart-1.

interesting, as it shows that not all pollutant concentrations are higher in denser cities.

So far, we have considered the effect of population density on individual pollutants. For city dwellers, however, total air quality, which takes into account all pollutants at once and therefore better indicates imminent health threats, is more interesting.[44] To assess overall air quality effects of density, we resort to the AQI as described in Section 2.4.1. Results in Table 2.4 show an AQI-density elasticity of 0.14 for both background and traffic stations in the OLS regressions (see columns 1 and 5). A one standard deviation increase in population density increases the AQI by 5.8 percent. This is another important finding: we have seen above that NO_2 and PM_x concentrations increase, whereas ozone concentration decreases with density. However, using a common index, we find that overall air quality decreases with density.

Table 2.4: OLS and IV regressions for the Air quality index

	Background Stations				Traffic Stations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV Density 1910	IV Soil	IV 1910 & Soil	OLS	IV Density 1910	IV Soil	IV 1910 & Soil
log(pop density)	0.135***	0.0694	0.224***	0.128***	0.135***	0.0912	0.145**	0.114*
	(0.0128)	(0.0436)	(0.0730)	(0.0399)	(0.0162)	(0.0840)	(0.0684)	(0.0596)
Unemployment share	-0.269	0.0167	-0.835	-0.388	0.394*	0.709	0.285	0.523
	(0.187)	(0.470)	(0.677)	(0.485)	(0.236)	(0.652)	(0.634)	(0.591)
Av. GDP	0.00360	0.0694	-0.0549	0.0265	-0.0195	0.0191	-0.0253	0.00596
	(0.0145)	(0.0484)	(0.0572)	(0.0466)	(0.0187)	(0.0651)	(0.0595)	(0.0576)
Av. Income	0.0896*	0.129	0.0131	0.0652	0.173***	0.145	0.149	0.110
	(0.0524)	(0.150)	(0.130)	(0.137)	(0.0630)	(0.166)	(0.152)	(0.162)
Green Voters	-0.281*	-0.0869	-0.745	-0.353	0.916***	1.091*	0.911	1.034*
	(0.169)	(0.434)	(0.608)	(0.444)	(0.180)	(0.593)	(0.554)	(0.539)
N	2142	2001	2137	1996	1147	1087	1139	1079
R^2	0.495	0.481	0.483	0.486	0.365	0.366	0.372	0.379
Depend. Var.	log(AQ)	log(AQ)	log(AQ)	log(AQ)	log(AQ)	log(AQ)	log(AQ)	log(AQ)
Districts	200	182	195	181	134	120	128	118
First-stage F-statistic		123.1	5.897	13.36		35.52	12.54	19.17

Note: The table presents OLS and IV model outcomes of regressing the background-station and traffic-station Air Quality Index (AQI) on log(population density). Control variables included are: Average district-level GDP, average income, share of green party voters, and the unemployment share. Station-specific controls are ignored since the index is constructed using district-level averages of pollutants. The instruments used are historical population from 1910, soil characteristics, and a combination of both. Standard errors in parentheses are clustered at labour market region - year (OLS) and labour market region (IV) level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

IV regressions. We now turn to the IV regression results. To judge the relevance of the instruments, in Table 2.D.1 we regress population density on each of the instruments and report the R^2. Historical population density is the strongest predictor of current population

[44]There are several websites, such as https://www.airnow.gov/, that provide up-to-date information on air quality in different regions and the extent to which it poses a health threat.

density, while the strength of the explanatory power varies for the geological instruments. For example, soil differentiation and subsoil mineralogy by themselves explain only about 1-3% of the variation in current population density, while the carbon content in the soil and the dominant parent material explain between 15 and 28% of the variation. In our main regressions we will additionally present F-statistics and partial R-squares in order to gauge the instruments' relevance.

In Table 2.E.1 we report results from regressions of population density on our instruments individually and combined. Again, some of the geological instruments by themselves seem weak, with low values of R^2 and F-statistics. However, historical density as well as the geological instruments combined are strong instruments.[45] Therefore, our instruments are relevant in that they explain a large share of the variation in current population density.

IV results are shown in Tables 2.2 and 2.3. As previously stated, historical density is a stronger instrument, but the exclusion restriction for soil characteristics is easier to defend. In the absence of a simple decision criterion, we rely in the remainder of the study on the results obtained with both sets of instruments; however, using only one of them does not change the interpretations dramatically in most cases.

The estimation coefficients suggest a density elasticity of 0.25 for NO_2, 0.08 for PM_{10}, and 0.02 for $PM_{2.5}$, although again, the latter is imprecisely estimated. For O_3, the IV estimate is -0.14. Finally, the IV coefficient for the AQI is 0.13 for background and 0.11 for traffic stations. Thus, a one standard deviation increase of population density increases the NO_2 concentration by 11%, and the PM_{10} concentration by 3.3%. The O_3 concentration decreases by 5.8% for a one-standard deviation increase in density. Last, the AQI increases by 5.5% for background stations and 4.9% for traffic stations. In general, the IV results using soil characteristics as instruments are fairly similar to the OLS results, while there is a bit more variation if we use the historical density instrument. In summary, it seems that the bias from omitted variables in OLS regressions is small, a point also made by Combes et al. (2010).

The argument for the exogeneity of historical instruments is that agglomeration tends to persist. Consequently, large and dense cities of the past tend to be large and dense today.

[45] The results are shown only for the sample of NO_2 stations. We repeated these regressions for the subsamples of stations covering the other pollutants, but results do not differ significantly between them.

Moreover, if enough time has elapsed, industry structures and other unobserved factors that might be correlated with current pollution levels should have changed sufficiently over time; therefore, the assumption that historical density does not affect current pollution other than through its effect on current density seems plausible. However, there remains concern that some unmeasured historical characteristic that correlates with past density and persists over time will influence current pollution. For instance, cities that were large industrial centres and densely populated in the past may still contain a lot of dirty industry today.

Likewise, places with fertile soil have historically become dense settlements because they could sustain large populations. The exclusion restriction is that these characteristics do not affect current pollution directly. Since agriculture makes up less than two percent of employment[46] and contributes around 10% to total air pollution, this seems plausible.[47]

To address the concern of potentially endogenous instruments, we include two additional tests. First, we control for historical shares of workers in industry and crafts and agriculture in our basic IV regressions. Results are in Table 2.D.2. We control for the share of workers in industry and crafts in the IV estimations with historical density (odd columns), and for the share of workers in agriculture in the IV regressions with soil instruments (even columns). We find that the coefficients on population density change only slightly, and to a lesser extent for the historical instrument.

Second, we regress the current shares of employment in industry, agriculture, and manufacturing on past population density. The results are in Table 2.E.2. Interestingly, they show that cities with high historical density contain *less* industrial and agricultural employment today. Thus, it seems like over a century or so, structural change led large centres of industry to shift into services; likewise, historically dense cities today contain less agricultural employment. It seems, therefore, that this kind of structural change renders a correlation of historical population patterns with current pollution unlikely.

Moreover, the over-identification tests for the soil instruments and the combination of soil and

[46]See e.g. https://www.umweltbundesamt.de/themen/boden-landwirtschaft/umweltbelastungen-der-landwirtschaft.

[47]More precisely, agriculture contributes 10% to NO_X emissions, 5% to $PM_{2.5}$ and 15% to PM_{10} emissions, see https://www.umweltbundesamt.de/themen/luft/emissionen-von-luftschadstoffen/quellen-der-luftschadstoffe.

historical ones do not seem to indicate instrument endogeneity (see the Hansen p-statistics in Tables 2.2 and 2.3).

Fixed effects and long differences. Fixed effects regressions may be a proper response to time invariant unobserved heterogeneity that causes cities to be more or less dense and more or less polluted at the same time. For instance, if dense cities provide amenities which attract 'green' households and these influence local environmental policies, the correlation of density and pollution might be driven by household selection. Using fixed effects at the district level could mitigate this selection bias. However, the variation in density and pollution within districts over time is much lower than the between variation, so fixed effects take out a lot of the interesting variation and the coefficient of interest is less precisely estimated. We present long-difference estimates for the years 2002-2015 as well as district fixed effects outcomes in Table 2.5. Fixed effects have the advantage of providing more observations (all years between 2002 and 2015), while the within variation of density and pollution is lower than for the long differences.

Table 2.5: Long difference and fixed effects estimations from 2002 to 2015

	NO_2		PM_{10}		$PM_{2.5}$		O_3	
	(1) FE All years	(2) LD 2002-15	(3) FE All years	(4) LD 2002-15	(5) FE All years	(6) LD 2002-15	(7) FE All years	(8) LD 2002-15
log(pop density)	0.337**	0.356**	-0.0223	-0.101	0.308	0.615	0.254	0.358
	(0.133)	(0.178)	(0.0913)	(0.157)	(0.228)	(1.768)	(0.157)	(0.228)
Distance to CBD	-0.00995***	-0.00851***	-0.00359**	-0.00225	-0.000513	-0.00113	0.00500***	0.00254
	(0.00231)	(0.00268)	(0.00145)	(0.00233)	(0.00392)	(0.00590)	(0.00163)	(0.00232)
Suburban	0.345***	0.312***	0.0986***	0.0139	0.0311	0.0575	-0.153***	-0.141*
	(0.0589)	(0.0877)	(0.0364)	(0.0616)	(0.101)	(0.241)	(0.0576)	(0.0767)
Urban	0.574***	0.536***	0.188***	0.118*	0.127	0.168	-0.232***	-0.248***
	(0.0488)	(0.0527)	(0.0392)	(0.0620)	(0.102)	(0.222)	(0.0656)	(0.0798)
Industrial	0.131***	0.158***	0.127***	0.0474	0.0542	-0.0122	-0.0875*	-0.0876
	(0.0433)	(0.0411)	(0.0414)	(0.0513)	(0.0393)	(0.0810)	(0.0455)	(0.0821)
Traffic	0.717***	0.668***	0.276***	0.262***	0.263***	0.231**	-0.213***	-0.223**
	(0.0337)	(0.0404)	(0.0151)	(0.0238)	(0.0337)	(0.104)	(0.0468)	(0.0869)
N	5575	781	4648	545	795	135	3776	549
R^2	0.896	0.897	0.761	0.804	0.794	0.932	0.824	0.834
Districts	269	258	247	235	109	105	251	248

Note: The table presents fixed effects estimations (uneven columns) for all years between 2002 and 2015 as well as long difference estimations (even columns) considering only the first year (2002) and the last year (2015) in the sample. Outcome is the respective pollutant and main parameter of interest is log(population density). Since the composition of measuring stations in the sample may change over time, we include all controls into the estimations. Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), average district-level GDP, average income, share of green party voters, the unemployment share, a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial). Standard errors in parentheses are clustered at labour market region level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As Table 2.5 shows, the estimated coefficient on population density becomes insignificant in all but the NO_2 regressions. This may be caused by the lower precision of the estimates due to the lower within variation of population densities. The coefficient in the NO_2 regression is 0.356 in the long difference regression, and 0.337 with district fixed effects, and both coefficients are significant at the 5 percent level.

We also ran station fixed effects regressions, which are presented in Table 2.D.3. The magnitude of the coefficient in the NO_2 estimations is very similar to the one in the OLS and IV regressions.

The NO_2 results point to the important role of car traffic for air pollution in German cities, since this pollutant is predominately emitted by motorized vehicles. The recent discussion on threshold violations (see Section 2.6.3) and driving bans for Diesel cars underlines the political dimension of this debate. In fact, running the long difference regressions by station type results in a large and significant effect of density on NO_2 concentration for traffic stations only.[48] This emphasizes the relevance of car traffic for NO_2 pollution.

2.6.2 Additional results and robustness

In this subsection, we perform a number of robustness checks to see how sensitive the results are to various specifications and to shed light on some interesting issues. First, we examine how the inclusion of control variables affects the estimates. We then perform estimations with variations in the definition of city and population density and use satellite data as alternative pollution measurements. Finally, we discuss several potential mechanisms that may drive our results.

Inclusion of controls. First, we check how sensitive the results are to the inclusion of controls. On the one hand, this may give some indication of whether our results are prone to suffer from omitted variable bias. On the other hand, we have several variables that may themselves be endogenous, but which may serve as mediating variables through which density affects pollution. This issue will be separately discussed below. We start with population density and year fixed effects as the only explanatory variables and successively add further control variables to the OLS regressions. Results are shown in Tables 2.E.3 to

[48] Results are not presented here.

2.E.6. With respect to the NO_2 results, we see in column (2) that adding station-specific control variables (urban/suburban, distance to CBD, distance to major road, station type) cuts the coefficient on population density in half. Column (3) adds an indicator for the state (*Bundesland*) in order to control for state-specific policies. This reduces the coefficient size further.[49]

The presence of a coal-fired power plant may be a driving force for air pollution in some regions since the emitted pollutants may be transported over long distances (Zhou et al., 2006). When we include an indicator for the existence of a coal-fired power plant in the district, the coefficient remains basically unchanged (column 4 of Table 2.E.3). In column (5) we replace the indicator variable with a measure of the distance to the closest coal-fired power station. While the former captures whether a dirty power plant exists in the same district as the monitoring station, a distance measure is independent of administrative boundaries and thus may be better suited to capture the role of such power plants for pollution. Again, however, the density coefficient is only slightly reduced.

Adding economic variables (log GDP per capita, log of average household income and share of unemployment in a district) in column (6) lowers the coefficient a little further. The density estimate remains relatively stable in magnitude across the range of included control variables, and always remains highly significant. In summary, once we add a basic set of control variables, which account for station-specific attributes, the coefficient does not change significantly.

The picture is similar for the PM_{10} outcomes (Table 2.E.4). Here, in particular adding state fixed effects reduces the density coefficient; it remains highly significant throughout all of the specifications though.

For $PM_{2.5}$ (Table 2.E.5), the density coefficient becomes insignificant as soon as we add indicators for the presence of coal-fired power plants or when adding state fixed effects (column 3, and 4).[50] However, comparing the sample distributions in Figure 2.1 (Panel b)

[49]One possible reason for the reduced effect may be that Berlin and Hamburg, Germany's two largest cities, are also states and the coefficient captures the within-state effect. Running regressions without these two states/cities leads to a coefficient of about 0.22 with NO_2 as outcome.

[50]When we control for distance to the next postal code with a coal-fired plant, the coefficient remains marginally significant.

and in Figure 2.D.1 (Panel a), we see that the $PM_{2.5}$ sample fails to cover many of the regions that are in the PM_{10} and the NO_2 sample. In particular, many of the densely populated areas like Berlin, the Stuttgart metropolitan area, and parts of the metropolitan areas of Hamburg and Munich are missing.

For O_3, the picture is similar to the NO_2 outcomes (Table 2.E.6): adding basic controls cuts the coefficient in half, but it remains highly significant throughout our specifications (except when adding state fixed effects in column 3). In contrast to the other pollutants, the density coefficient is negative, so densely populated cities seem to suffer less from O_3 pollution.

City definitions and population density measures. Our next set of results aim to provide sensitivity checks to particular definitions we make throughout the study. We first examine whether different spatial units, rather than districts, affect the results before turning to alternative definitions of population density.

A common definition of a city is based on the economic relations between locations, usually measured by commuting flows. We therefore rerun our basic regressions for LMR as defined by Kosfeld and Werner (2012). Similar to other concepts such as Metropolitan Statistical Areas (MSAs) or Functional Urban Areas (FUAs), LMRs are defined as collections of districts with significant commuting flows between them. There are 141 LMRs, of which 128 contain at least one pollution monitor. Results are shown in Table 2.D.4 (the IV regressions use historical density and soil as instruments). The results are very close to the coefficients for districts. For $PM_{2.5}$, both the OLS and IV estimates turn significant.[51] Table 2.E.7 contains results using municipalities (*Gemeinden*) as spatial unit.[52] The general pattern that emerges is that across pollutants, both the OLS and IV coefficients are smaller in absolute size than those obtained with districts or LMRs.

In order to completely free ourselves from any unit definition, be it a city or some sort of administrative area, we used Geographic Information System (GIS) software to create buffers of one and of five kilometres around monitoring stations. Then, we captured population

[51] Additionally, we ran regressions using the definition of FUAs as described by Moreno-Monroy et al. (2020). Results obtained with FUAs are very similar to the ones using districts, but are not presented here.

[52] In Germany, there are between 11.000 and 12.000 municipalities of which only a very small share contains a pollution measuring station (366 municipalities in our NO_2 sample and even less in the samples with the other pollutants). Furthermore, number and form of municipalities changed quite substantially over time. From 2000 until 2015, the total number was reduced from about 14.000 to about 11.000.

density within these buffers.[53] In Table 2.E.8 we see that this station specific population density measure leads to lower estimates the smaller the buffer size. For NO_2, PM_{10}, and O_3 regressions, the estimate doubles in size when increasing the buffer from 1 to 5 kilometres.

While we cannot say for sure which spatial scale is most appropriate, we think that municipalities and similarly smaller sized buffers are not the 'correct' units, since taking this approach neglects the economic density of nearby geographic units which affect pollution, e.g. through commuting and other economic activities (such as power plants or industrial spillovers) that produce spillover pollution. In other words, the smaller the spatial unit, the larger will the disparity between the generation and the local exposure to pollution be. In our view, the interpretation of linking pollution exposure to economic density is thus probably best viewed at scales larger than the community level.

The next issue we examine is the definition of population density. So far, our measure of interest was total district population divided by total built up area. However, some studies have used other measures of agglomeration (see e.g. Ahlfeldt and Pietrostefani (2019) for a discussion). We therefore rerun our basic estimations with different density measures, see Table 2.E.9. In particular, instead of population density we now use population divided by the entire district area (instead of built up area only), total population in the district or the total employment per km^2 (all in logs).

As is to be expected, the results differ somewhat from our main results quantitatively but not qualitatively. Using the first alternative population density measure cuts the coefficients in half for all pollutants. The coefficient on total population is a bit smaller than the one for density in the case of NO_2 but larger for PM_{10} and $PM_{2.5}$. Obviously, the population size can be very high in large districts, while the population density is low. Then, the interpretation of the coefficient is somewhat different. Looking at employment density opens another angle on the pollution-density relation. Focusing on population density emphasizes residential energy use and short-distance commuting, while examining employment density rather addresses longer-distance commuting and possible agglomeration effects on industrial pollution.

As the third row of each panel in Table 2.E.9 suggests, the coefficients are close to the

[53] In order to do so, we used population data gridded in 1x1 kilometre raster cells (Breidenbach and Eilers, 2018).

ones from our baseline results, especially for NO_2. The distinction between employment and residential density should be particularly important for smaller spatial units such as municipalities. Indeed, labour market regions are constructed by maximizing commuting flows within the unit, so that the distinction between residential and employment density becomes unimportant.

Between models (OLS or IV with our different instruments) the coefficients are relatively stable for almost all of the independent variables we look at.

Heterogeneities. Another interesting question is whether the effect of density on pollution is driven by traffic or 'background' activities such as residential energy use or perhaps industrial fumes that disperse over the entire city area. Table 2.E.10 presents outcomes including an interaction of population density with the station type indicator. The density coefficient now corresponds to the average effect of population density on pollution at background stations; it remains positive for NO_2 and particulates. Overall, we find that the density effect on air pollution seems more pronounced at traffic and industrial stations.[54]

Additionally, we test whether the density effect differs between growing and shrinking cities. Sluggish responsiveness of infrastructure and housing stock may imply that growing and shrinking cities have different density-pollution relationships. On the one hand, the gradient may be steeper in growing cities if the creation of new road infrastructure does not keep up with increases in traffic, which might lead to congestion and higher pollution. On the other hand, the effect might be stronger in shrinking cities, if infrastructure and housing stock do not shrink in par with the population, which could imply higher energy use at given densities compared to growing urban agglomerations. Figure 2.E.1 shows the density coefficients for cities where population increased between 2002 and 2015 and those were it decreased. With the exception of $PM_{2.5}$, we find that (in absolute terms), the effect of density on pollution seems to be smaller in growing than in shrinking districts.

Up to now, we have assumed the effect of density to be linear. This need not be the case, as an increase of density might affect pollution differently depending on the level of it. For instance, traffic congestion may only pick up if density rises so much that traffic volume

[54]Results for O_3 (not presented here) show that the effect is not significantly different at traffic stations compared to background stations but is significantly lower at industrial sites.

outpaces infrastructure supply. Hence, we test for non-linear effects of density in a stylized way. We divide the sample into five population density quantiles and test whether the effect is different between them (assuming constant density within quantiles). Results are shown in Figure 2.D.2. For NO_2 and O_3, the outcomes indicate that the effect is driven by denser cities: pollution increases for NO_2 and decreases for O_3 when moving up the density distribution. For PM_{10}, the effect of density increases when moving from the first to the second quantile, but stays constant thereafter; for $PM_{2.5}$, there is no clear cut relationship (as before, the estimates are more noisy than for the other pollutants).

Satellite outcomes. As a final robustness check, we rerun our regressions using satellite data instead of monitor readings as measuring source of pollution. Satellite readings are available for NO_2 and $PM_{2.5}$. These data have been used in many recent studies (e.g., Freeman et al., 2019; Achakulwisut et al., 2019). Compared to on-site monitor readings, the satellite data contain pros and cons. On the one hand, they are potentially subject to measurement error, as pollution is not directly measured but retrieved indirectly from related physical observations (e.g. aerosol optical thickness, which is inferred from atmospheric reflections that absorb visible and infrared light). On the other hand, the satellite data are available as grid cells covering the entire country instead of only a subset of districts like the monitors. The grid sizes are 0.1° by 0.1° (approximately 10 km x 10 km at the equator) for NO_2 and 0.01° by 0.01° (1 km x 1 km) for $PM_{2.5}$. Related to this, monitors may be placed endogenously in high pollution/high density locations. While these concerns should be mitigated by our long difference and IV estimates, it is still interesting to check how monitor and satellite based results match up. The correlation between pollution measurements by station readings and by satellite data is 0.4 for NO_2 and 0.7 for $PM_{2.5}$.[55]

Table 2.D.5 shows the results. Satellite based $PM_{2.5}$ outcomes are very close to the main PM_{10} findings based on monitor readings. They are now precisely estimated, which shows the upside of the much more complete coverage of satellite data. For NO_2, the results are similar to the baseline, but show a somewhat larger effect of density on emissions.[56]

Satellite and monitor data may, however, represent different samples and thus not be directly

[55] More detail on the satellite data is found in Appendix 2.C.

[56] Carozzi and Roth (2020) also find lower estimates for the $PM_{2.5}$-density elasticity using in-situ monitor readings.

comparable. We therefore align the satellite and the monitoring station samples. That is, we now keep only grid cells (or districts) containing at least one in-situ monitoring station. OLS estimates are in Table 2.E.11. It shows a significantly higher density coefficient when using station readings rather than satellite data. For $PM_{2.5}$, the difference between satellite and station estimates is smaller and not significant. Since station-based measurements are more accurate, we prefer estimates using station data for NO_2. Regarding $PM_{2.5}$ we find satellite data to be more appropriate, since the network of measuring stations is not well developed. Overall, the results using satellite data are broadly in line with our baseline estimates.

Discussion of mechanisms. We now turn to potential mechanisms that may be responsible for our findings. Urban economic models like the one presented by Borck and Pflüger (2019) analyse how urban pollution is driven by industrial and agricultural production, transport, and residential energy use. An alternative mechanism may be the sorting of "green" residents into cities. We therefore want to discuss what we can learn about potential channels from our analysis. One by one, we include the following variables. Firstly, we look at the share of public transport users and car density. Secondly, we consider industrial composition. Thirdly, we add some variables meant to measure the 'greenness' of cities: green space area, an indicator for environmental zones and the share of green party voters. Lastly, we take into account housing density. If the population density coefficient changes significantly, the variable may be viewed as a driver of the density-pollution gradient. Results are in Tables 2.E.12–2.E.15.[57]

In the first two columns of Tables 2.E.12–2.E.15, we add the number of public transit users (col. (1)) and number of cars (col. (2)) in a district as controls. Denser districts are likely to have both more transit users and more cars (although the number of cars per capita is lower). Intuitively, we find that NO_2 and PM_{10} pollution is positively correlated with the number of cars and negatively with the number of transit users. Hence, commuting by car contributes to the positive density effect, while commuting by transit reduces it.[58] For O_3,

[57] We repeated the exercise with IV estimates, and the results are very similar to the OLS ones. These results are not explicitly shown here.

[58] To get a complete picture, we would have to know how total vehicle kilometres for driving and transit change with density, as well as the emissions intensity of the two modes. This is beyond the scope of the paper, but we conjecture that while denser cities have a higher transit shares and lower car density, they still have more traffic in total, which contributes to pollution even though the average commute may be cleaner than in less densely populated cities.

the picture is again the opposite to NO_2: O_3 is reduced by the number of cars and increased by the number of transit users.

Next, we look at industry composition. It might be that dense cities are more polluted because they are specialized in dirty industries. To investigate this mechanism, we add the employment shares of industry types (industrial production, construction, agriculture, financial industries, public sector, and trade) to our regressions. The sectors that are most polluting are industry, construction and agriculture, whereas services and public sector should be much less polluting. We add the share of 'dirty' and 'clean' industries in total production separately as one block each (see cols (3) for 'dirty' and (4) for 'clean' industries). We can extract the following: at least for NO_2 and PM_{10}, controlling for 'dirty' industries increases the density coefficient, while employment in these industries should be positively correlated with pollution. In fact, denser cities are less specialized in dirty industries and more in relatively clean ones. So it seems that the industry composition actually causes pollution to decrease with density. For the 'clean' industry shares, we find small effects on the density coefficients, so they don't seem to strongly affect the pollution-density relationship.

We then add variables measuring the 'greenness' of cities in a physical sense as well as concerning preferences of the local population. Our first variable is the total area of green space in a city. If densely populated cities have less green space, pollution concentration may be higher than in less densely populated cities, as trees and plants may capture or filter air pollution and lower temperatures. Conversely, plants and trees emit VOC, which are precursors to O_3. The results are shown in column (5) in the respective tables. Green space increases O_3 and decreases NO_2 concentration; this is consistent with the fact that plants emit VOC, which react with NO_2 to form O_3. Since, in fact, denser districts have more green space, controlling for green space slightly increases the density coefficient for NO_2, while the estimate is smaller in the O_3 regressions. The effect of green space on PM_{10} is rather small.

Another indicator for the environmental friendliness of a city is whether or not it contains a LEZ. City authorities may designate a LEZ within their jurisdiction such that only cars fulfilling certain environmental requirements, then indicated by specific coloured badges, may enter. The designation is a political decision and hence may mirror the environmental preferences of the residents. In column (6) of Tables 2.E.12–2.E.15, we see that LEZs do not alter the density coefficient even though they are mostly positively correlated with pollution concentration.[59]

Moreover, we consider whether the density effect might be mediated by sorting of households according to environmental preferences. For instance, families might move to less dense and greener locations to avoid adverse health effects for their children. Conversely, cities may attract "green" voters with a strong preference for the environment. We follow the last track and include the share of green votes in our regressions (col (7)). For particulates and O_3, we find that the share of green voters reduces pollution. Since the green voter share is positively correlated with density, this reduces the density coefficient. This would imply that selection of green voters into cities makes denser cities greener.[60]

Lastly, we look at the number of buildings as a proxy for residential energy use. The results are in the last column of Tables 2.E.12–2.E.15. Denser districts obviously have more buildings. As the tables show, buildings are positively correlated with NO_2 and PM_{10} emissions. Hence, more residential energy use in denser cities seems to contribute to higher pollution.[61] Again, the opposite is true for O_3.

In summary, we find that the density effect is hardly driven by industry composition or the composition of the population in denser cities. Dense cities seem to have cleaner industries and are inhabited by 'greener' residents. However, consistent with our simple model, they have more total commuting and more total residential energy use. Our interpretation of the results is that denser cities are thus more polluted, even though each resident may produce lower emissions due to higher transit shares and more efficient energy use.

2.6.3 Threshold results

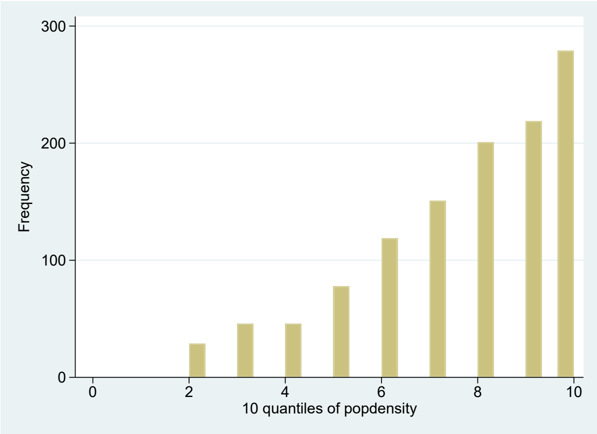
We now turn to the analysis of threshold violations. These have been the primary focus of recent policy debates, as cities and national governments in Germany and other European countries have been sued for violations of legally binding thresholds.

[59] Of course, this may be due to reverse causality: if pollution is high, political pressure for introducing an LEZ mounts.

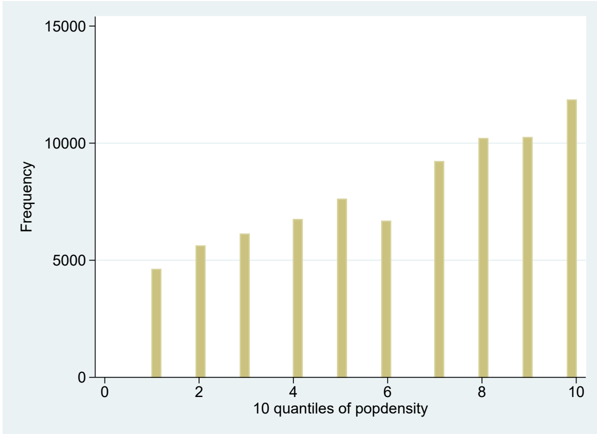
[60] Interestingly, for NO_2, the green vote share seems to positively correlate with pollution. However, once we control for the vote shares of other major parties (CDU, SPD, the Left and FDP), the outcome mirrors that for the other pollutants.

[61] Intuitively, building density is also larger in denser cities. Since this implies a higher energy efficiency (Borck and Brueckner, 2018), dense buildings flatten the density-pollution gradient.

Figure 2.3: Histograms of threshold transgressions by deciles of population density



(a) Transgressions of annual NO_2



(b) Transgressions of daily PM_{10}

Note: Own calculations. The graph depicts histograms of population density deciles and transgressions of WHO thresholds on measuring station level. Population density is divided into 10 equal deciles. The $y-$axis in the left thus shows the number of transgressions of the annual threshold of 40 $\mu g/m^3$ for NO_2 in each decile from 2002 until 2015. The right panel shows the quantity of transgressing the daily PM_{10} threshold of 50 $\mu g/m^3$ per population density decile within the same time frame.

For a first visual impression, Figure 2.3 shows the number of transgressions of the daily mean threshold of PM_{10}, and the NO_2 annual mean by population density decile. The histograms suggest a clear positive association between density and threshold transgressions.

In Table 2.6, we present results for the probability that the yearly mean was exceeded for NO_2, PM_{10} and $PM_{2.5}$. We concentrate on linear probability models (LPM), again using the historical and the soil IVs in some specifications. For $PM_{2.5}$, there is no significant relation between density and annual threshold violations. For NO_2 and PM_{10}, all results are positive and highly significant. The probability that the annual NO_2 threshold of 40 $\mu g/m^3$ is transgressed is significantly higher in more densely populated areas. Coefficients, except for the one with soil characteristics as instruments, are similar in NO_2 and PM_{10} regressions. Repeating the estimations with Probit IV models yields similar results, see Table 2.E.16.

Results for the transgressions of daily (PM_{10} and $PM_{2.5}$) as well as hourly (NO_2) means are shown in Table 2.7. Thus, the probability of exceeding the one-hour NO_2 threshold on more than 17 days a year is significantly higher in denser areas, even though the point estimate is relatively small, at 0.026. The lower we set the number of days, the higher the coefficient. Regarding PM_{10}, the probability of exceeding the threshold for 34 days is also significantly higher in denser cities with a point estimate of 0.05. For $PM_{2.5}$, we find an insignificant

Table 2.6: Probability of transgression of annual thresholds

	NO_2		PM_{10}		$PM_{2.5}$	
	(1) LPM	(2) LPM IV	(3) LPM	(4) LPM IV	(5) LPM	(6) LPM IV
log(pop density)	0.148***	0.143***	0.100***	0.0866***	-0.0128	-0.0229
	(0.0232)	(0.0313)	(0.0214)	(0.0264)	(0.0153)	(0.0164)
Distance to CBD	0.00143*	0.00136	0.00138	0.000820	0.000380	0.0000524
	(0.000818)	(0.000841)	(0.000993)	(0.00102)	(0.000809)	(0.000864)
Suburban	-0.0376*	-0.0311	0.218***	0.207***	0.185***	0.136***
	(0.0191)	(0.0199)	(0.0403)	(0.0414)	(0.0643)	(0.0497)
Urban	-0.0378	-0.0303	0.275***	0.264***	0.206***	0.158***
	(0.0244)	(0.0285)	(0.0459)	(0.0474)	(0.0710)	(0.0558)
Industrial	-0.0000506	-0.00415	0.260***	0.255***	0.0395	0.0403*
	(0.0206)	(0.0208)	(0.0411)	(0.0426)	(0.0272)	(0.0225)
Traffic	0.549***	0.563***	0.311***	0.316***	0.0124	0.0142
	(0.0427)	(0.0457)	(0.0303)	(0.0320)	(0.0140)	(0.0120)
N	5663	5341	4812	4520	795	743
R^2	0.494	0.507	0.421	0.421	0.240	0.179
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Districts	266	245	244	225	108	101

Note: The table presents the effect of log(population density) on a dummy variable of whether the annual threshold value was transgressed on district level. The thresholds are: 40 $\mu g/m^3$ for NO_2, 10 $\mu g/m^3$ for $PM_{2.5}$, and 20 $\mu g/m^3$ for PM_{10}. The regressions are performed using Linear Probability Models LPM and LPM IV. Marginal effects are calculated at means. Control variables are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, year fixed effects, a categorical variable for the station area classification (rural [base category], suburban, urban), and station type (background [base category], traffic, industrial). The instruments are historical population from 1910 and soil characteristics combined. Standard errors in parentheses are clustered at district level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

effect of density on daily threshold violations (note again, however, the smaller sample size). As shown in *Panel B* of Table 2.E.16, using a probit model does not change the results.

In summary, the evidence suggests that threshold violations occur more frequently in more densely populated cities.

2.7 Conclusion

In this paper, we have used panel data for German districts to estimate the effect of population density on air pollution. Our theoretical model predicts that denser cities should have higher pollution concentrations, although there are some countervailing forces. The evidence

Table 2.7: Probability of transgressing thresholds by specific number of days

	NO_2			PM_{10}			$PM_{2.5}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	> 17	> 14	> 9	> 34	> 29	> 24	> 34	> 29	> 24
log(pop density)	0.0255*	0.0277**	0.0312**	0.0503***	0.0625***	0.0546**	-0.0334	-0.00681	-0.00931
	(0.0132)	(0.0138)	(0.0147)	(0.0171)	(0.0196)	(0.0224)	(0.0389)	(0.0348)	(0.0276)
Distance to CBD	0.000295	0.000330	0.000354	0.000840	0.000633	0.000476	0.000367	0.000719	0.00124
	(0.000242)	(0.000247)	(0.000282)	(0.000701)	(0.000733)	(0.000868)	(0.00219)	(0.00148)	(0.00135)
Suburban	-0.00563	-0.00595	-0.00730	0.0135	0.0251	0.0519*	0.376***	0.316***	0.275***
	(0.00526)	(0.00550)	(0.00575)	(0.0200)	(0.0240)	(0.0268)	(0.0844)	(0.0807)	(0.0787)
Urban	-0.00938	-0.0100	-0.0117*	0.0148	0.0246	0.0579*	0.394***	0.341***	0.311***
	(0.00591)	(0.00620)	(0.00640)	(0.0225)	(0.0285)	(0.0314)	(0.0872)	(0.0853)	(0.0768)
Industrial	0.00835	0.00858	0.00789	0.114**	0.129**	0.167***	0.0623	0.113**	0.0559
	(0.00627)	(0.00656)	(0.00705)	(0.0452)	(0.0542)	(0.0637)	(0.0543)	(0.0436)	(0.0358)
Traffic	0.0439**	0.0470**	0.0592***	0.239***	0.270***	0.306***	0.161***	0.140***	0.0699**
	(0.0176)	(0.0182)	(0.0198)	(0.0219)	(0.0245)	(0.0236)	(0.0465)	(0.0404)	(0.0317)
N	5663	5663	5663	4817	4817	4817	795	795	795
R^2	0.053	0.057	0.063	0.268	0.298	0.333	0.311	0.253	0.219
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Districts	269	269	269	247	247	247	109	109	109

Note: The table presents outcomes of the effect of log(population density) on a dummy variable indicating whether the daily threshold value was transgressed on a pre-specified number of days within the district. The respective thresholds are: 200 $\mu g/m^3$ for NO_2, 25 $\mu g/m^3$ for $PM_{2.5}$, and 50 $\mu g/m^3$ for PM_{10}. The number of days (e.g. 17, 14, and 9 for NO_2) lie just below the number of exceedances allowed by the EU, which are 18 for NO_2 and 35 for PM_{10}. The regressions are performed using Linear Probability Models LPM. Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, year fixed effects, a categorical variable for the station area classification (rural [base category], suburban, urban), and station type (background [base category], traffic, industrial). The instruments used are a combination of historical population from 1910 and soil characteristics. Standard errors in parentheses are clustered at district level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

to date has been largely inconclusive. To mitigate concerns about unobserved heterogeneity and omitted variables, we have used both long difference regressions and instrumental variables. Our preferred estimates come from the IV regressions, where we instrument population density with historical population and/or soil characteristics. We find that increasing population density by one percent increases NO_2 by 0.25 percent and PM_{10} by 0.08 percent. The results for $PM_{2.5}$ are less precisely estimated using monitoring station readings, but of similar magnitude to PM_{10} when considering satellite data. For O_3, we find denser cities to have lower concentrations, with an elasticity of -0.14. Air quality as measured by the aggregate AQI decreases with population density, with an elasticity of about 0.12 on average.

The study thus contributes to our knowledge about the economic costs of agglomeration. The benefits of agglomeration due to labour market pooling, spillovers, matching etc. are by now well documented. However, there is much less robust evidence on the costs of agglomeration.[62] Thus, our study makes some headway towards a more complete picture of agglomeration benefits and costs. This seems important for urban policies. In Appendix 2.B, we use a simple numerical example to show that, based on our estimates, local pollution may reduce optimal city size by 7%. Knowledge of the elasticity of pollution with respect to population leads to a more complete understanding of the benefits and costs of agglomeration.

As far as we know, together with Carozzi and Roth (2020), this is the only study that seriously tries to estimate the causal effect of population density on pollution. More evidence from other countries surely will add to a more complete picture about this issue. For instance, whether or not population and pollution interact differently in developing and developed countries seems like an interesting and important question. More research on the interaction of urban structure and pollution thus seems warranted.

[62]See Combes et al. (2018) for a recent study on the costs of agglomeration implied by high land and housing prices. The interpretation of these costs is different however, as long as land and housing markets are competitive.

Appendices

2.A A model

Consider a circular, open monocentric city where N residents commute to the Central Business District (CBD) for work. A household living at x km from the CBD incurs round-trip commuting costs tx. Household utility is $v(c, q) = c^{1-\alpha} q^{\alpha} \mathcal{P}^{-\beta}$, where c is non-housing consumption, q consumption of housing floor space in square meters, and $\mathcal{P}$ is the concentration of local pollution in the city. Households are completely mobile in the city, so they achieve utility level u regardless of their location.

The household maximizes utility subject to the budget constraint, $w = c - tx + pq$, where w is wage income and p the price of housing per square meter. Maximizing utility subject to the budget constraint gives the household's optimal housing demand $q = \alpha u^{\frac{1}{\alpha}} \mathcal{P}^{\frac{\beta}{\alpha}} (y - tx)^{1-\frac{1}{\alpha}}$, and the bid rent, i.e. the maximum willingness to pay per unit of housing floor space, $p = u^{-1/\alpha} (y - tx)^{\frac{1}{\alpha}} \mathcal{P}^{-\frac{\beta}{\alpha}}$.

Housing floor space is produced by profit maximizing developers, using capital K and land L as inputs. We assume a Cobb-Douglas production function written in intensive form $h = S^{\theta}, \theta > 0$, where $S \equiv K/L$ is structural density (capital deployed per unit of land) and h is the amount of floor space per unit of land. We normalize the price of capital to one. The developer maximizes profits per unit of land

$$\pi = S^{\theta} - S - R,$$

where R is the land rent paid to (absentee) landowners. Solving the developers' problem gives structural density, $S = \theta^{\frac{1}{1-\theta}} u^{\frac{1}{\alpha(\theta-1)}} (y - tx)^{\frac{1}{\alpha - \alpha\theta}} \mathcal{P}^{\frac{\beta}{\alpha(\theta-1)}}$, and the land rent function at distance x, $R = \left(\theta^{\frac{\theta}{1-\theta}} - \theta^{\frac{1}{1-\theta}}\right) u^{\frac{1}{\alpha(\theta-1)}} (y - tx)^{\frac{1}{\alpha-\alpha\theta}} \mathcal{P}^{\frac{\beta}{\alpha(\theta-1)}}$.

We consider a small open city. Residents are freely mobile between the city and the rest of the economy. Letting u be the exogenous utility level that can be attained in the rest of the economy, the equilibrium is defined by the two equations

$$\begin{aligned} R(\bar{x}, u, \mathcal{P}) &= R_A && (2.A.1) \\ \int_0^{\bar{x}} \gamma D(x, u, \mathcal{P}) 2\pi x dx &= N, && (2.A.2) \end{aligned}$$

where $\bar{x}$ is the distance from the city border to the CBD, R_A is the agricultural land rent, and γ denotes the share of developable land at any distance x. $D = \frac{h(x,u,\mathcal{P})}{q(x,u,\mathcal{P})}$ is the population density at distance x from the CBD.

Solving (2.A.1) and (2.A.2) gives the endogenous city border $\bar{x}$ and number of residents N.[63]

Pollution is composed of emissions from commuting, $\mathcal{C}$, and residential energy use, H, weighted by the respective emissions factors. Transport emissions are assumed to be proportional to the aggregate commuting distance, and residential emissions are proportional to total housing floor space in the city. Letting the emissions intensities of commuting and housing be e_C and e_H, total emissions are

$$\begin{aligned} E &= e_C \mathcal{C} + e_H H && (2.A.3) \\ \mathcal{C} &= \int_0^{\bar{x}} x \gamma D(x) 2\pi x dx && (2.A.4) \\ H &= \int_0^{\bar{x}} \gamma h(x) 2\pi x dx. && (2.A.5) \end{aligned}$$

Finally, assume for simplicity that the concentration of air pollution is given by total emissions divided by land area.[64] Then concentration is given by $\mathcal{P} = E/(\pi \bar{x}^2)$.

How then does pollution concentration change with population density? To answer this question, we vary the parameter γ. For instance, government may increase γ by more liberal zoning policies (e.g. increasing the floor-area ratio by allowing more housing to be built per sq. meter of land). When γ increases, the city shrinks spatially as the city border $\bar{x}$ moves inward, for given population. For given population, this has two effects on residents' utility: First, utility would increase, since housing has become less scarce. Second, however,

[63]We use the following parameter values: $\gamma = 0.75$, $e_H = e_C = 1, r_A = 50,000, w = 50,000, t = 500, \beta = 0.05, \theta = 0.75, \alpha = 0.25$. These values are similar to those used in Borck and Brueckner (2018).

[64]In reality, concentration is given by emissions per cubic meter of air, but we can slightly simplify by assuming all pollution is at ground level and thus concentration equals emissions over land area.

pollution concentration rises: first, reduced competition for land raises aggregate housing consumption and residential energy use, and second, since the city shrinks, average and total commuting distance falls. The combined effect – assuming equal emissions coefficients on residential and transport emissions – is an increase in emissions, and an increase in concentration, since total emissions now diffuse over a smaller area.[65]

In our simulation, we find that residents' utility increases, as the housing effect dominates the increase in pollution concentration.[66] This will lead to in-migration from the rest of the economy, which increases the number of residents and the urban boundary $\bar{x}$, although not to its previous level. As a result of the increased population level, the effect on pollution concentration is reinforced, and density rises as well. Our simulation shows that the end result in the open city is that the increase in γ increases density and pollution concentration. The positive relation between concentration and density emerging from the model is shown by the upper blue curve in Fig. 2.A.1.

Extension. We now examine a couple of extensions that affect the relation between density and pollution. First, denser cities typically have higher mode shares for public transport because of economies of scale and traffic density. Second, denser cities have taller buildings that are more energy efficient. We include these two aspects in a simple reduced form fashion. In particular, we assume that transport emissions fall with average city density because density shifts transport mode choice towards cleaner public transit. Second, we assume that because of higher energy efficiency, density reduces the emissions associated with residential energy use.[67]

We amend Equations (2.A.4) and (2.A.5) in the following way:

$$\mathcal{C} = \int_0^{\bar{x}} x\gamma D(x)dx\bar{D}^{-\kappa} \tag{2.A.6}$$

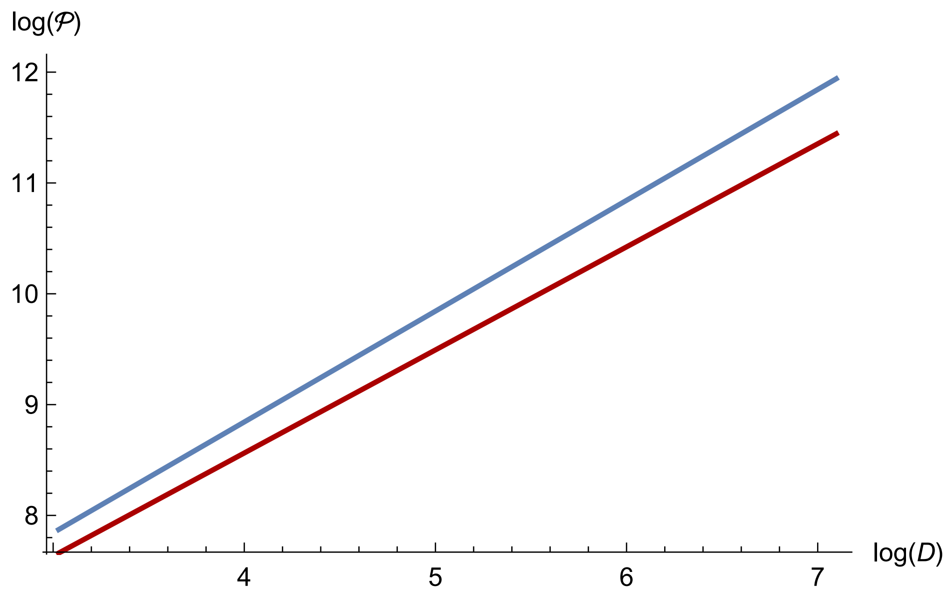
$$H = \int_0^{\bar{x}} \gamma h(x)dx\bar{D}^{-\mu}, \quad \kappa, \mu > 0, \tag{2.A.7}$$

[65]If we increase the emissions coefficient for transport emissions sufficiently, we find that raising γ decreases aggregate emissions, but concentration still increases.

[66]This will be true as long as β, the parameter which governs the strength of pollution damage, is not too large. We choose a value similar to recent literature here, but even a 3-fold increase would not change our results.

[67]Borck (2019) contains a model with pollution and mode choice without scale economies. Borck and Brueckner (2018) use a micro-founded model where energy use is related to a building's surface, which implies that more densely populated locations with tall buildings are more energy efficient.

Figure 2.A.1: Population density and pollutant concentration



Note: Own calculations. The figure shows the correlation between population density and pollution concentration as calculated in a monocentric city model. Both lines are produced with the following parameter values: $\gamma = 0.75$, $e_H = e_C = 1, r_A = 50,000, w = 50,000, t = 500, \beta = 0.05, \theta = 0.75, \alpha = 0.25$. The red line extends the model assuming that denser cities are more energy efficient and have lower transport emissions. Equations (2.A.4) and (2.A.5) are then extended by $\bar{D}^{-\kappa}$ and $\bar{D}^{-\mu}$, where $\bar{D} = N/(\pi\bar{x}^2)$ is the endogenous average density and $\kappa = \mu = 0.07$ are the elasticities of transit mode shares/energy efficiency with respect to density.

where $\bar{D} = N/(\pi\bar{x}^2)$ is the (endogenous) average city density, and κ and μ are the density elasticities of transport emissions and residential energy emissions.

Ahlfeldt and Pietrostefani (2019) find that both transit mode share and energy efficiency rise with density, with both elasticities being about -0.07. Setting $\kappa = \mu = 0.07$ produces the red curve in Fig. 2.A.1. Thus, the relation between density and pollution remains positive but is flattened compared to the benchmark case where density economies are absent. In fact, the relation will remain positive as long as $\kappa, \mu < 1$. Given the estimates in Ahlfeldt and Pietrostefani (2019), this restriction seems very likely to hold.

2.B Optimal city size

Consider a simplified version of the model above, where we abstract from housing construction. We modify the model in two respects: first, we assume that there is congestion, which implies that commuting costs per km are given by $t = N^{\psi}$. Second, we follow the literature on agglomeration economies and assume that the wage is given by N^{δ}. Suppose that pollution concentration can be described as in our empirical model by the relation $\mathcal{C} = N^{\rho}$. Solving the model (using $\alpha = 1/4$) gives the following functional form for utility as a function of population size:

$$\log u = (\delta - \beta\rho) \log N - \frac{1}{4} \log \left(\frac{256}{27} \left(N^{1+\psi} + R_A \right) \right). \tag{2.A.8}$$

Suppose that $\beta = 0$, so households don't care about pollution. Furthermore, in line with the large empirical literature on agglomeration, let $\delta = 0.05$ (Combes and Gobillon, 2015), and, following Duranton and Puga (2019), let $\psi = 0.07$. Then, maximizing u with respect to city size gives $N^* = 6235$.

Suppose, however, that $\beta = 0.02$, and let $\rho = 0.15$ (between our estimates for PM_{10} and NO_2, and close to our coefficient for the AQI).[68] Then, we find an optimal city size of $N^{**} = 5810$, 93% of N^*. Hence, pollution can significantly affect the balance of agglomeration benefits and costs.

2.C Data

Weather. Since weather and emission stations are usually not at the exact same spot, we have to match both types of stations such that we get the most accurate information about the weather at each emission station. Following the approach of Auffhammer and Kellogg (2011), for each emission station we searched for the ten closest weather and precipitation stations within a range of 50 kilometres and a maximum station altitude difference of 200 meters.[69] Out of those stations, we choose a primary one, which is the closest weather or precipitation station to the emission station with at least 50 percent of hourly observations non-missing. All emission stations that could not be assigned a primary station were deleted from the sample. Throughout a year, there are gaps between recordings such that many weather and precipitation stations do not have a full record of observations. Such missing observations were imputed by regressing the non-missing values of, say, sunshine on the

[68]The value of 0.02 is close to the value used by Borck and Tabuchi (2019) based on a calibration to the social cost of carbon.

[69]There are many more precipitation stations in Germany (more than 4000) than stations which provide information on all other weather variables other than rainfall and snowfall (a little more than 700). This is why we separately merged precipitation and other weather stations to each emission station.

sunshine records of all the other adjacent stations. The estimated coefficients of those other stations were then used to impute values for missing observations.

About 80 percent of particulate matter and nitrogen dioxide emission stations were matched to the closest available weather station. Less than four percent of PM_{10}–stations and two percent of NO_2–stations were matched to a weather station ranked 5th or higher regarding the ranking of distance between the two station types. In both cases (PM_{10} and NO_2), less than one percent of emission stations could not be assigned a weather station.

Historical industry data. To construct the historical data for workers in industry and crafts, we proceeded as follows. We had maps for administrative units now and in 1925 and for 1925 the total number of workers in industry and crafts as well as the total population of a historical district. Due to the fact that administrative assignment changed over time, we had to assign historical administrative units to current units. If the historical area matched with current districts by more than 60 percent of the area, those areas were assigned the recent district. In many cases this is true for more than one historical district. For example, southern and northern Dithmarschen correspond to the current Dithmarschen. In these cases, we just summed the number of workers and the number of inhabitants in 1925 and assigned the sum to the current administrative unit. From these variables we then calculated the shares of workers in industry and crafts over the whole resident population. A number of current districts could not be assigned to workers because there were no historical districts matching by at least 60 percent of the area. This is true for example for Wolfsburg, a city that was established after 1925 and did not exist back then. Other cases like Mainz or Worms were larger districts in the past and were assigned as district-free cities after 1925. In such cases, the recent district almost completely lies within a historical district and we assigned the value of the respective historical district. As these are only relatively few cities and districts, we performed this matching by eyeballing the maps and looking which area fits best to the current district.

Geology. We use the same 12 variables from the European Soil Database (ESDB) used by Combes et al. (2010).[70] The data comes in raster format of 1km$\times$1km rasters, which

[70]These data can be freely downloaded for research purposes from the European Soil Data Centre (Panagos et al., 2012).

we aggregate to the district level. For each district we use for instance the value of the dominant parent material which occurs most often within the district. Especially in urban areas like Berlin, we need to impute some of the values because of the lack of information in the data. In such cases, the dominant value often is described as a non-soil or just missing. In these cases we use the second most common value occurring within the district. The variables we use describe the mineralogy of the topsoil and the subsoil as well as the dominant parent material of the soil at different levels of aggregation. The dominant parent material describes the bedrock of the soil, which is the underlying geological material. At the broader level of aggregation, these are e.g. sedimentary rocks, igneous or metamorphic rocks, while the finer level of aggregation further classifies them. For instance, sedimentary rocks may consist of different types of limestone (hard, soft, marly, chalky etc.), marlstone or other types of stones. Mineralogy captures the presence of minerals in the different layers of soil (the topsoil being usually 5 to 15 cm deep and the subsoil being the intermediate layer between the topsoil and the bedrock).

We also include information about the water capacity of the topsoil (from low to very high) and the subsoil (from very low to very high), the depth to rock (from shallow to very deep), the soil erodibility class (from very weak to very strong), the topsoil organic carbon content (from low to very high), the soil profile differentiation (no differentiation, low and high differentiation) and the hydrological class, which consists of four categories describing the circulation and retention of underground water. The last variable we use is the ruggedness of a district, which is calculated as the difference between the mean of maximum altitudes of all the rasters within a district and the mean of minimum altitudes across all rasters within the same district.

We include the information on mineralogy, hydrological class and parent material as dummies in the regressions. All other variables, which differ in the quality of a characteristic (e.g. from low to high) remain in their continuous form. All variables are included as dummies in the regressions, except for ruggedness, which is the only continuous variable among the soil characteristics.[71]

[71]Note that we do not use water capacity of the topsoil and the subsoil, the depth to rock, the soil erodibility class, and the hydrological class in our main analyses. However, we also ran regressions including those variables as instruments and did not find the second stage outcomes to change significantly.

Satellite data. Gridded ground-level pollution data for NO_2 and $PM_{2.5}$ is obtained from the Atmospheric Composition Analysis Group (Geddes et al., 2015; Van Donkelaar et al., 2016). It comes at resolutions of 0.1° by 0.1° (NO_2) and 0.01° by 0.01° ($PM_{2.5}$). In order to get pollution at the district level, we took a map of German districts and calculated the mean concentration of ground-level pollution per district and year using the gridded pollution data.

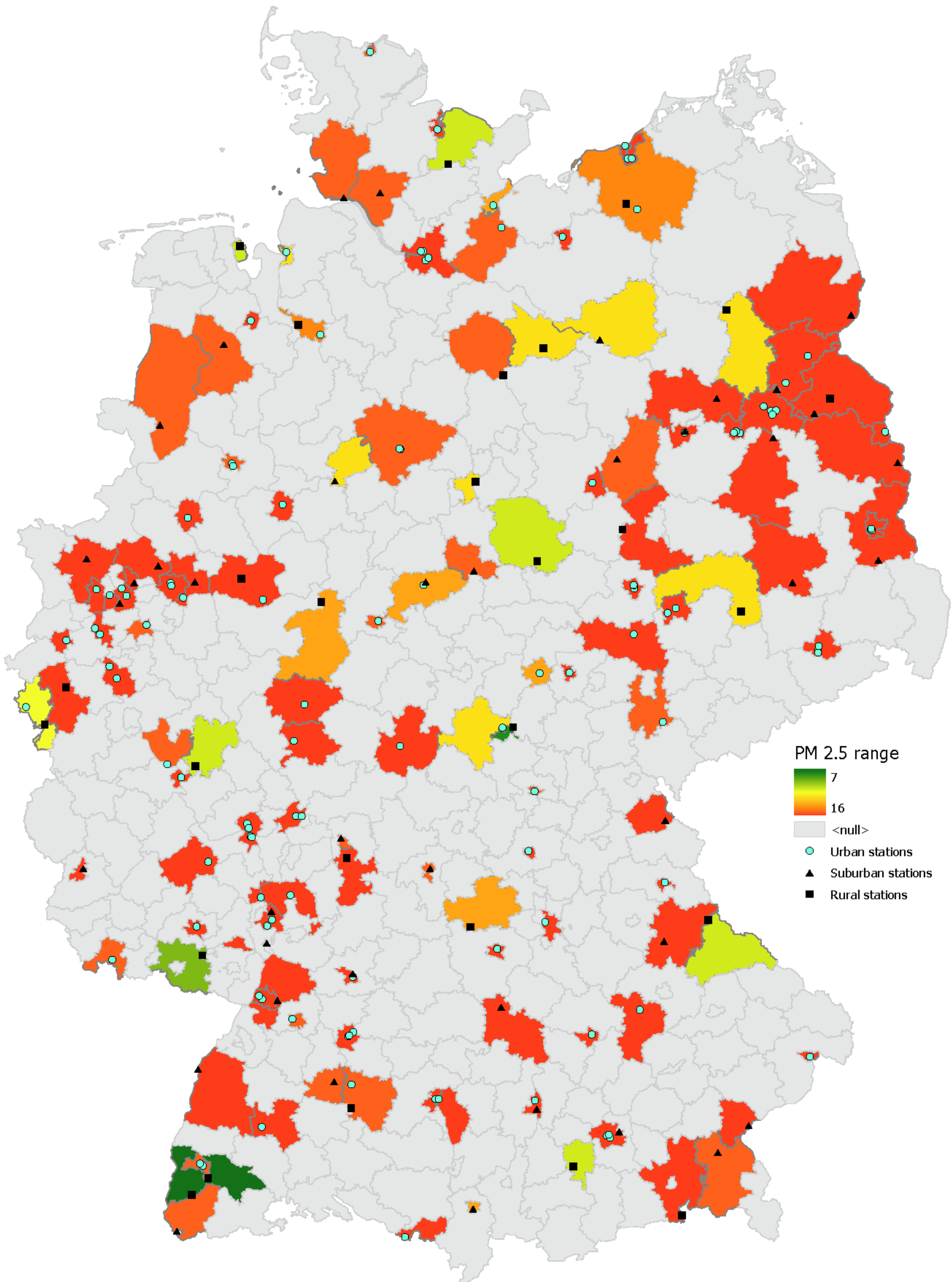
$PM_{2.5}$ data itself is obtained from different satellite instruments from NASA (MODIS, MISR, and SeaWIFS), which observe backscattered solar radiation and thereby Aerosol Optical Depth (AOD).[72] These observations are then transformed into ground-level data using Chemical Transport Models (here GEOS-Chem), which simulate the geophysical relationship between AOD and $PM_{2.5}$. Afterwards, those estimations are calibrated using monitoring stations where possible. This is mostly the case in economically developed areas like the US or Europe. The resulting dataset provides annual pollution data from 1998 until 2016. An exact explanation of how the data is produced is provided by Van Donkelaar et al. (2016).

For the construction of ground-level NO_2 data, again a chemical transport model is used and the approach is generally similar to the one explained above. The data set covers the time span from 1996 through 2012 and is described by Geddes et al. (2015).

[72] Generally, small particles floating in the atmosphere are called aerosols.

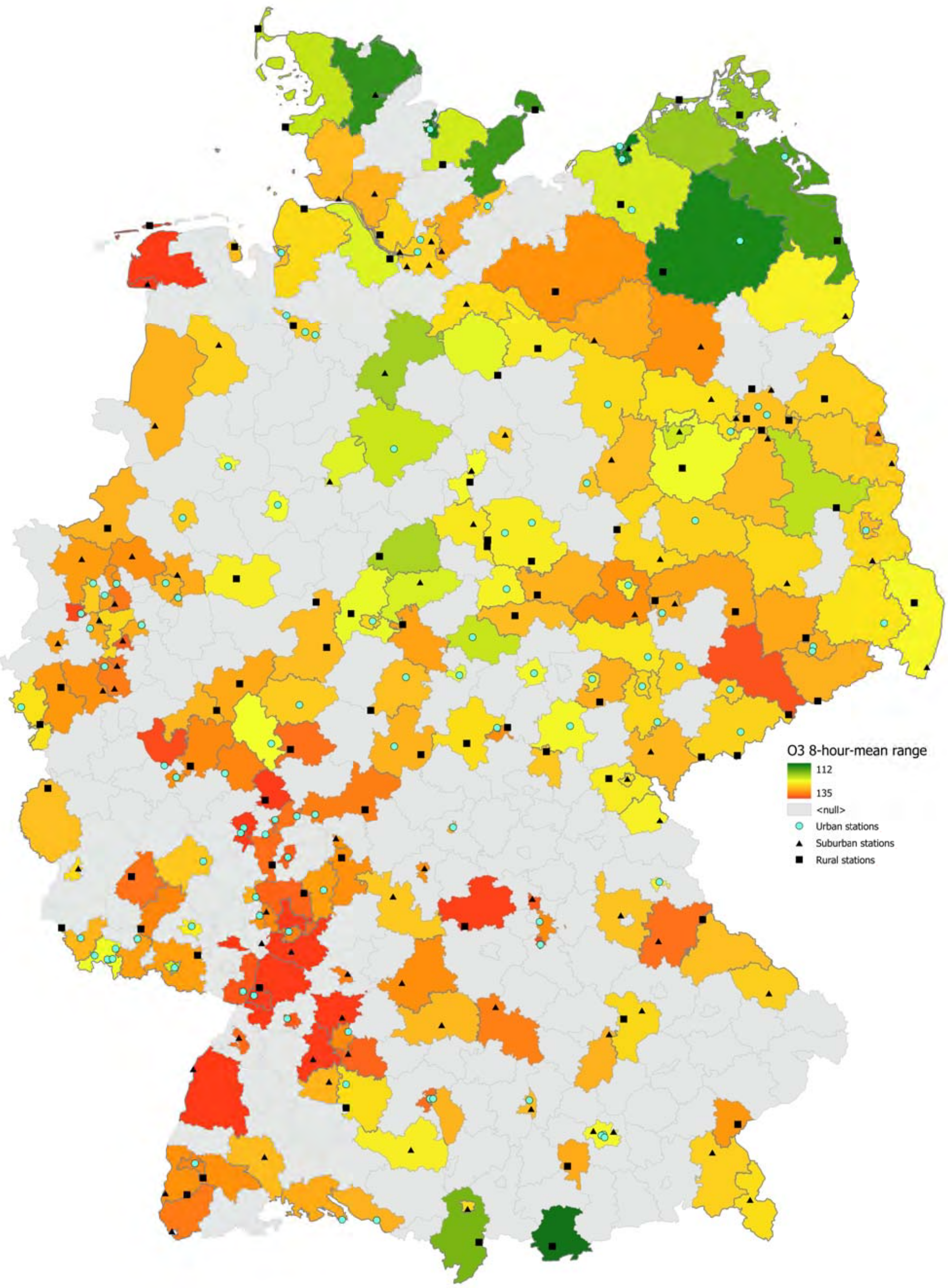
2.D Additional results and robustness checks

Figure 2.D.1: Monitoring stations and concentration levels in 2015 ($PM_{2.5}$)



(a) Sample of $PM_{2.5}$ stations

Figure 2.D.1: Monitoring stations and concentration levels in 2015 (O_3)



(b) Sample of O_3 stations

Note: Own calculations. The maps show average district-level pollution concentrations in 2015 for $PM_{2.5}$ and O_3 respectively. Pollution concentrations are relatively low in green coloured districts and relatively high in red coloured ones. Turquoise dots represent measuring stations that are marked as "urban" stations, black triangles lie in "suburban" districts and black squares depict "rural" stations. Grey coloured districts do not contain measuring stations.

Table 2.D.1: R^2 of regressions of population density on instruments

	NO_2	PM_{10}	$PM_{2.5}$	O_3
Population density in 1910	0.70	0.70	0.70	0.70
Ruggedness	0.07	0.05	0.06	0.03
Soil differentiation	0.03	0.02	0.01	0.01
Soil carbon content	0.24	0.22	0.28	0.21
Dominant parent material	0.15	0.18	0.23	0.16
Topsoil mineralogy	0.04	0.04	0.02	0.03
Subsoil mineralogy	0.03	0.02	0.03	0.02

Note: The table presents the R^2 of bivariate regressions of regressing population density on different instruments. Population density in 1910 represents the historical population density instrument. Soil characteristic instruments are: ruggedness, soil differentiation (3 categories), soil carbon content (4 categories), dominant parental material (7 categories), topsoil mineralogy (4 categories), subsoil mineralogy (3 categories). For more details on the instruments see Section 2.C. The regressions are run for each sample of the respective pollutant in the year 2015. Observation numbers are as follows: 392 in NO_2-sample, 331 in PM_{10}-sample, 132 in $PM_{2.5}$-sample, and 253 in O_3-sample.

Table 2.D.2: IV regressions with historical working population as control variable

	NO_2		PM_{10}		$PM_{2.5}$		O_3	
	(1) IV Density 1910	(2) IV Soil	(3) IV Density 1910	(4) IV Soil	(5) IV Density 1910	(6) IV Soil	(7) IV Density 1910	(8) IV Soil
log(pop density)	0.178***	0.370***	0.0801***	0.0366	0.0452	-0.00907	-0.0911***	-0.295***
	(0.0569)	(0.0579)	(0.0273)	(0.0410)	(0.0545)	(0.0619)	(0.0329)	(0.0495)
Share employed in Ind.	0.0598		0.184**		0.376***		0.00615	
	(0.122)		(0.0826)		(0.145)		(0.0997)	
Share workers in Agr.		0.376***		-0.159**		-0.190		-0.355***
		(0.102)		(0.0727)		(0.134)		(0.0814)
Distance to CBD	-0.00514***	-0.00282**	0.000914	0.000739	0.00199	0.00126	0.00366***	0.00124
	(0.00153)	(0.00130)	(0.000881)	(0.000885)	(0.00154)	(0.00159)	(0.000997)	(0.000925)
Distance to Street	-0.0957**	-0.102***	-0.0280	-0.0410**	-0.0313	-0.0366	0.0420*	0.0513**
	(0.0399)	(0.0375)	(0.0193)	(0.0199)	(0.0430)	(0.0416)	(0.0248)	(0.0248)
Suburban	0.280***	0.288***	0.0688***	0.0685***	0.0953	0.116*	-0.143***	-0.142***
	(0.0507)	(0.0518)	(0.0247)	(0.0256)	(0.0583)	(0.0692)	(0.0363)	(0.0370)
Urban	0.474***	0.474***	0.122***	0.131***	0.123*	0.161**	-0.221***	-0.214***
	(0.0708)	(0.0720)	(0.0330)	(0.0325)	(0.0705)	(0.0821)	(0.0455)	(0.0468)
Industrial	0.0847**	0.113***	0.115***	0.129***	0.0659*	0.0895**	-0.0496	-0.107***
	(0.0382)	(0.0351)	(0.0356)	(0.0378)	(0.0362)	(0.0390)	(0.0339)	(0.0336)
Traffic	0.659***	0.649***	0.254***	0.250***	0.137***	0.123***	-0.235***	-0.246***
	(0.0407)	(0.0387)	(0.0184)	(0.0185)	(0.0436)	(0.0430)	(0.0444)	(0.0399)
N	5091	5336	4272	4485	747	769	3414	3581
R^2	0.751	0.763	0.486	0.489	0.294	0.273	0.420	0.459
Districts	269	269	247	247	109	109	251	251

Note: The table presents second stage IV estimates including historical shares of workers in different sectors (industry, agriculture, manufacturing) for each pollutant separately. Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, average district-level GDP, average income, share of green party voters, the unemployment share , a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial). The instruments used are historical population from 1910, and soil characteristics. Standard errors in parentheses are clustered at labour market region - year (OLS) and labour market region (IV) level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.D.3: Station fixed effects for all pollutants

	NO_2		PM_{10}		$PM_{2.5}$		O_3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(pop density)	0.290***	0.286**	0.00294	0.0241	0.336*	0.338	0.268**	0.0144
	(0.0942)	(0.119)	(0.0912)	(0.109)	(0.193)	(0.225)	(0.106)	(0.127)
Av. GDP		0.120*		0.0566		-0.108		-0.0858
		(0.0624)		(0.0548)		(0.148)		(0.0519)
Av. Income		0.238*		0.148		0.00201		-0.152
		(0.132)		(0.149)		(0.533)		(0.160)
Unemployment share		0.612*		0.452		0.528		0.654**
		(0.369)		(0.372)		(1.061)		(0.317)
Green Voters		-0.457		-1.328***		0.863		-0.243
		(0.333)		(0.493)		(1.369)		(0.364)
LEZ		-0.00340		-0.0167***		-0.00577		0.00836
		(0.00591)		(0.00597)		(0.00854)		(0.00697)
N	5575	4905	4648	4137	795	719	3776	3438
R^2	0.094	0.107	0.010	0.028	0.075	0.102	0.040	0.043
Districts	269	269	247	247	109	109	251	251
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Weather	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table presents station fixed effects estimations for all years between 2002 and 2015 without (uneven columns) and with (even columns) controls. Outcome is the respective pollutant and main parameter of interest is log(population density). Control variables included are: Weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, average district-level GDP, average income, share of green party voters, the unemployment share, and whether the station lies within an environmental zone or not (may change over time). Standard errors in parentheses are clustered at labour market region level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.D.4: Estimations with Labour Market Regions

	NO_2		PM_{10}		$PM_{2.5}$		O_3	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
log(pop density)	0.311***	0.303***	0.0958***	0.109***	0.0718*	0.0847**	-0.233***	-0.222***
	(0.0648)	(0.0722)	(0.0248)	(0.0262)	(0.0394)	(0.0411)	(0.0439)	(0.0492)
Distance to CBD	-0.00670***	-0.00672***	0.000140	0.000188	0.000839	0.000872	0.00387***	0.00392***
	(0.00140)	(0.00141)	(0.000741)	(0.000734)	(0.00120)	(0.00117)	(0.000996)	(0.000995)
Distance to Street	-0.0971**	-0.0974**	-0.0384*	-0.0382*	-0.0347	-0.0343	0.0480**	0.0479**
	(0.0387)	(0.0384)	(0.0206)	(0.0203)	(0.0393)	(0.0383)	(0.0233)	(0.0231)
Urban	0.502***	0.503***	0.154***	0.152***	0.167***	0.163***	-0.231***	-0.232***
	(0.0602)	(0.0605)	(0.0298)	(0.0294)	(0.0591)	(0.0580)	(0.0407)	(0.0404)
Traffic	0.658***	0.658***	0.261***	0.261***	0.110***	0.111***	-0.252***	-0.251***
	(0.0364)	(0.0362)	(0.0173)	(0.0172)	(0.0413)	(0.0401)	(0.0361)	(0.0360)
N	5575	5575	4648	4648	795	795	3776	3776
R^2	0.753	0.753	0.481	0.480	0.268	0.268	0.470	0.470
Labour Market Regions	127	127	124	124	76	76	125	125
Weather	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic		9081.8		327398.9		132.2		25302.3

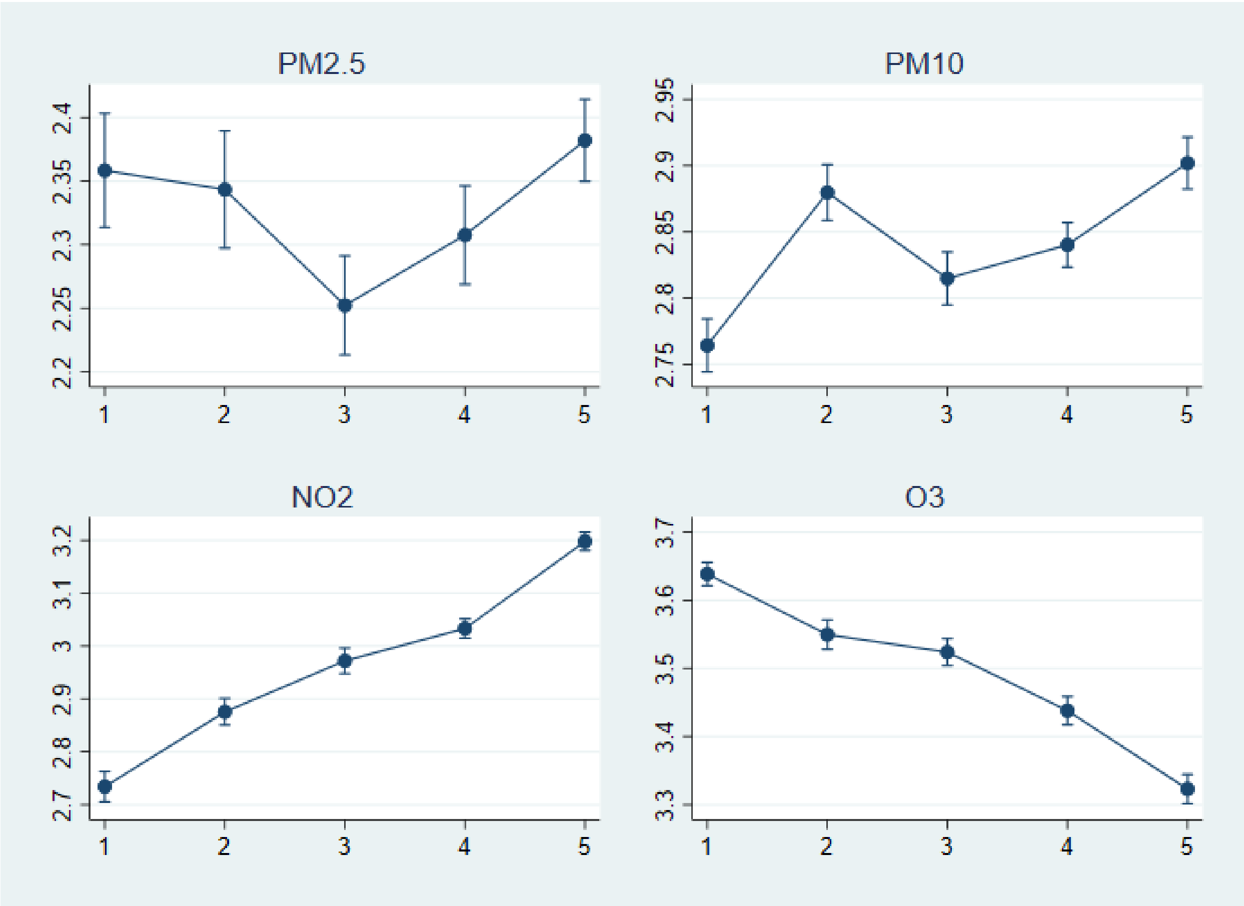
Note: The table presents OLS and IV model outcomes of regressing the respective pollutant concentration on labour market region level on log(population density). Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), average district-level GDP, average income, share of green party voters, the unemployment share, a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial), and whether the station lies within an environmental zone or not. The instruments used are the combination of historical population from 1910 and soil characteristics. Standard errors in parentheses are clustered at labour market region - year (OLS) and labour market region (IV) level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.D.5: Satellite data regressions

	(1) OLS	(2) IV Density 1910	(3) IV Soil	(4) IV 1910 & Soil	(5) FE	(6) LD
Panel A: NO_2						
log(pop density)	0.374***	0.418***	0.729***	0.385***	-0.00591	-0.133
	(0.0109)	(0.0158)	(0.0237)	(0.0141)	(0.109)	(0.144)
N	4414	4161	4414	4161	4414	803
R^2	0.247	0.249	0.059	0.252	0.364	0.338
Districts	402	402	402	402	402	402
Soil Characteristics	No	No	Yes	Yes	No	No
Panel B: $PM_{2.5}$						
log(pop density)	0.0485***	0.107***	0.0598***	0.0823***	0.421***	0.376***
	(0.00318)	(0.00451)	(0.00643)	(0.00406)	(0.0432)	(0.0802)
N	6022	5677	6022	5677	6022	803
R^2	0.497	0.466	0.496	0.483	0.704	0.641
Districts	402	402	402	402	402	402
Soil Characteristics	No	No	Yes	Yes	No	No

Note: The table presents OLS and IV model outcomes of regressing the respective pollutant concentration captured by satellites (NO_2 and PM_{10}) on log(population density). Columns (5) and (6) show fixed effects and long difference estimations on district level. The instruments used are historical population from 1910, soil characteristics, and a combination of both. Standard errors in parentheses are clustered at labour market region - year (OLS) and labour market region (IV) level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

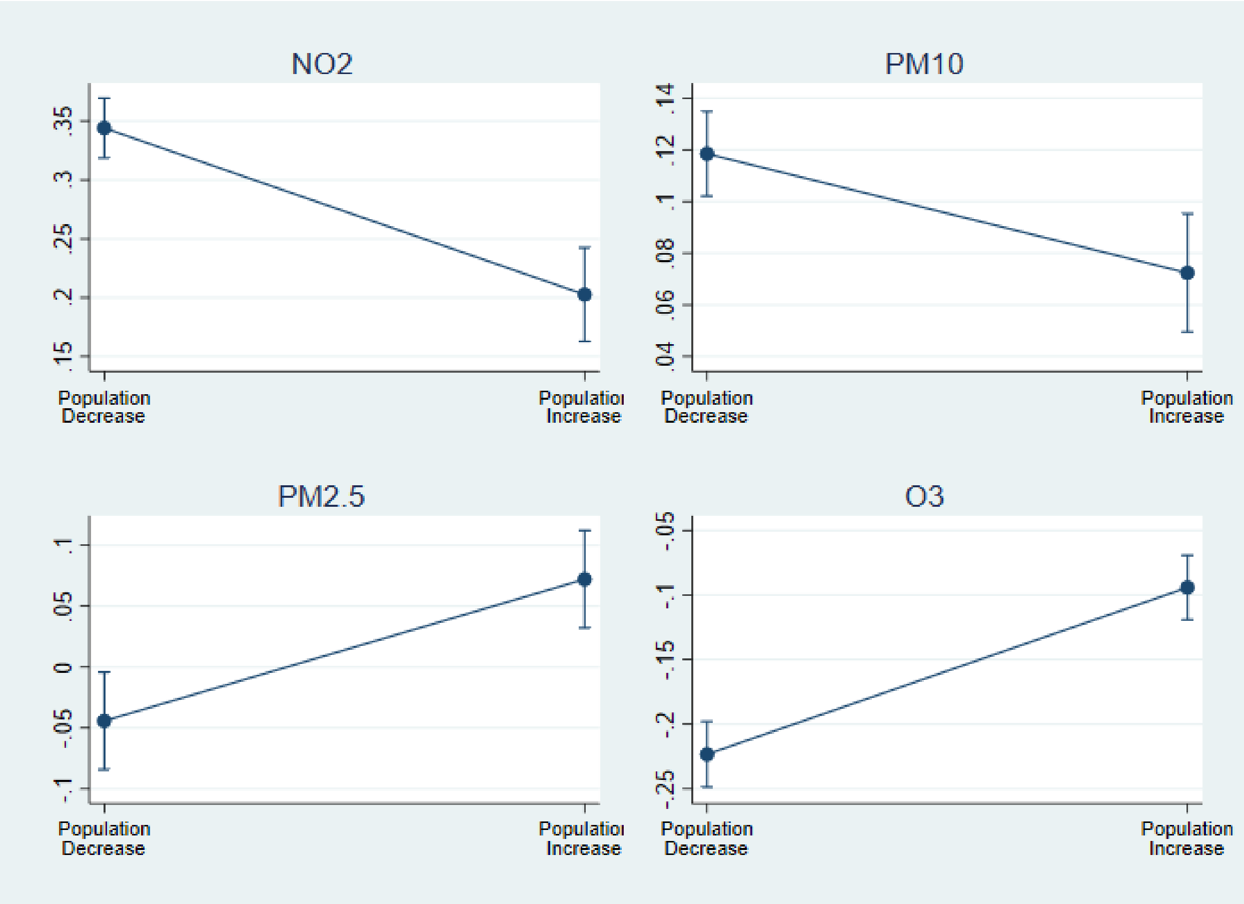
Figure 2.D.2: Quantiles of population density using the whole sample



Note: The figure shows coefficients of the effect of log(population density) on the respective pollutant concentration after dividing the density measure into five equal quantiles. Control variables included in the respective OLS regressions are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), a categorical variable for the station area classification (rural [base category], suburban, urban), and station type (background [base category], traffic, industrial). Standard errors are clustered at labour market region - year level.

2.E Additional tables and figures

Figure 2.E.1: Interaction effect of population increase and population density



Note: The figure shows the difference of OLS outcomes between growing and shrinking cities when regressing the respective pollutant on log(population density). Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), a categorical variable for the station area classification (rural [base category], suburban, urban), and station type (background [base category], traffic, industrial). Standard errors are clustered at labour market region - year level.

Table 2.E.1: First stage partial regressions of population density on different instruments (NO_2)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
log(pop density 1910)	0.257***		0.249***							
	(0.0144)		(0.0190)							
Ruggedness		-0.000396**	0.000506***	-0.000992***	-0.000579**					
		(0.000175)	(0.000187)	(0.000234)	(0.000239)					
N	372	390	370	390	392	392	392	392	390	392
R^2	0.722	0.593	0.811	0.536	0.329	0.431	0.308	0.382	0.299	0.295
F-Statistic	88.05	33.78	71.78	44.50	37.59	50.38	32.47	27.79	31.48	27.75
Adjusted R^2	0.717	0.570	0.799	0.512	0.316	0.419	0.293	0.362	0.285	0.280
Districts	227	227	227	227	227	227	227	227	227	227
Basic controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ruggedness	N	Y	Y	Y	Y	N	N	N	N	N
Soil carbon content	N	Y	Y	N	N	Y	N	N	N	N
Soil differentiation	N	Y	Y	Y	N	N	Y	N	N	N
Dominant parent material	N	Y	Y	Y	N	N	N	Y	N	N
Topsoil mineralogy	N	Y	Y	Y	N	N	N	N	Y	N
Subsoil mineralogy	N	Y	Y	Y	N	N	N	N	N	Y

Note: The table presents first stage regressions of population density on different instruments and varying combinations of them. 'Y' stands for 'Yes' and means the inclusion of the respective instrument. 'N' means 'No' and the exclusion of the instrument. Standard errors in parentheses are clustered at labour market region level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.2: Regressions of current sectoral shares on past population density

	NO_2		PM_{10}		$PM_{2.5}$		O_3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Agriculture:								
log(pop density 1910)	-0.00441***	-0.00571***	-0.00461***	-0.00604***	-0.00442***	-0.00557***	-0.00462***	-0.00576***
	(0.000448)	(0.000560)	(0.000477)	(0.000590)	(0.000545)	(0.000746)	(0.000461)	(0.000570)
N	209	209	186	186	97	97	191	191
R^2	0.417	0.371	0.440	0.396	0.493	0.423	0.415	0.351
Districts	227	227	211	211	105	105	206	206
Manufacturing:								
log(pop density 1910)	-0.0161***	-0.0224***	-0.0165***	-0.0229***	-0.0245***	-0.0289***	-0.0170***	-0.0219***
	(0.00450)	(0.00341)	(0.00441)	(0.00337)	(0.00522)	(0.00396)	(0.00475)	(0.00357)
N	209	209	186	186	97	97	191	191
R^2	0.045	0.149	0.049	0.162	0.112	0.278	0.052	0.138
Districts	227	227	211	211	105	105	206	206
Production:								
log(pop density 1910)	-0.0191***	-0.0222***	-0.0202***	-0.0229***	-0.0295***	-0.0297***	-0.0201***	-0.0220***
	(0.00455)	(0.00343)	(0.00446)	(0.00340)	(0.00555)	(0.00401)	(0.00479)	(0.00364)
N	209	209	186	186	97	97	191	191
R^2	0.065	0.148	0.074	0.163	0.156	0.291	0.075	0.141
Districts	227	227	211	211	105	105	206	206

Note: The table presents bivariate regressions of the importance of sectoral shares (agriculture, manufacturing, production) in 2015 on log(population density) in 1910 for each pollutant-sample separately. The importance is measured in shares of gross value (uneven columns) or as the share of workers (even columns) in the respective sector. Standard errors in parentheses are clustered at labour market region level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.3: OLS with varying sets of controls (NO_2)

	(1)	(2)	(3)	(4)	(5)	(6)
log(pop density)	0.591***	0.280***	0.190***	0.283***	0.240***	0.218***
	(0.0176)	(0.0141)	(0.0129)	(0.0147)	(0.0147)	(0.0127)
Distance to CBD		-0.00321***	-0.00237***	-0.00322***	-0.00352***	-0.00280***
		(0.000447)	(0.000394)	(0.000446)	(0.000429)	(0.000412)
Distance to Street		-0.105***	-0.0861***	-0.105***	-0.102***	-0.109***
		(0.0115)	(0.00998)	(0.0113)	(0.0113)	(0.0104)
Suburban		0.281***	0.323***	0.280***	0.284***	0.285***
		(0.0169)	(0.0158)	(0.0168)	(0.0167)	(0.0161)
Urban		0.445***	0.507***	0.445***	0.458***	0.500***
		(0.0216)	(0.0183)	(0.0215)	(0.0213)	(0.0195)
Industrial		0.0898***	0.118***	0.0881***	0.0765***	0.113***
		(0.0128)	(0.0125)	(0.0138)	(0.0129)	(0.0135)
Traffic		0.648***	0.667***	0.647***	0.648***	0.638***
		(0.0124)	(0.0118)	(0.0124)	(0.0124)	(0.0119)
Stone coal				-0.00863		
				(0.0148)		
Brown coal				0.0373		
				(0.0228)		
Distance to coal plant					-0.00145***	
					(0.000181)	
Av. GDP						0.0107
						(0.0150)
Av. Income						0.413***
						(0.0517)
Unemployment share						-1.031***
						(0.189)
N	5575	5575	5575	5575	5575	5489
R^2	0.305	0.755	0.795	0.755	0.759	0.773
Districts	266	266	266	266	266	264
Weather	Yes	Yes	Yes	Yes	Yes	Yes
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Coal plant in district	No	No	No	Yes	No	No
Distance to coal plant	No	No	No	No	Yes	No
State FE	No	No	Yes	No	No	No

Note: The table presents OLS model outcomes of regressing NO_2 pollution on log(population density) with varying combinations of control variables. Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), state & year fixed effects, average district-level GDP, average income, share of green party voters, the unemployment share, a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial), distance to the most proximate coal plant, whether a stone coal or a brown coal plant lies within the same district. Standard errors in parentheses are clustered at labour market region - year level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.4: OLS with varying sets of controls (PM_{10})

	(1)	(2)	(3)	(4)	(5)	(6)
log(pop density)	0.164***	0.0749***	0.0205***	0.0654***	0.0779***	0.117***
	(0.00637)	(0.00691)	(0.00711)	(0.00773)	(0.00755)	(0.00812)
Distance to CBD		0.000947***	-0.000584**	0.000811***	0.000971***	0.000563**
		(0.000297)	(0.000257)	(0.000302)	(0.000299)	(0.000279)
Distance to Street		-0.0405***	-0.0139**	-0.0397***	-0.0409***	-0.0355***
		(0.00648)	(0.00571)	(0.00658)	(0.00656)	(0.00657)
Suburban		0.0761***	0.0868***	0.0755***	0.0757***	0.0929***
		(0.00948)	(0.00930)	(0.00949)	(0.00953)	(0.00967)
Urban		0.140***	0.157***	0.139***	0.139***	0.145***
		(0.0108)	(0.0103)	(0.0109)	(0.0110)	(0.0112)
Industrial		0.136***	0.129***	0.132***	0.137***	0.124***
		(0.0127)	(0.0102)	(0.0128)	(0.0125)	(0.0122)
Traffic		0.255***	0.253***	0.258***	0.255***	0.250***
		(0.00672)	(0.00600)	(0.00682)	(0.00674)	(0.00663)
Stone coal				0.0253***		
				(0.00911)		
Brown coal				0.0142		
				(0.0210)		
Distance to coal plant					0.000113	
					(0.000140)	
Av. GDP						-0.109***
						(0.0104)
Av. Income						0.130***
						(0.0368)
Unemployment share						0.807***
						(0.123)
N	4648	4648	4648	4648	4648	4570
R^2	0.143	0.474	0.592	0.475	0.474	0.502
Districts	242	242	242	242	242	240
Weather	Yes	Yes	Yes	Yes	Yes	Yes
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Coal plant in district	No	No	No	Yes	No	No
Distance to coal plant	No	No	No	No	Yes	No
State FE	No	No	Yes	No	No	No

Note: The table presents OLS model outcomes of regressing PM_{10} pollution on log(population density) with varying combinations of control variables. Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), state & year fixed effects, average district-level GDP, average income, share of green party voters, the unemployment share, a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial), distance to the most proximate coal plant, whether a stone coal or a brown coal plant lies within the same district. Standard errors in parentheses are clustered at labour market region - year level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.5: OLS with varying sets of controls ($PM_{2.5}$)

	(1)	(2)	(3)	(4)	(5)	(6)
log(pop density)	0.0912***	0.0353**	0.00640	0.00683	0.0337*	0.0550***
	(0.0110)	(0.0161)	(0.0144)	(0.0160)	(0.0178)	(0.0206)
Distance to CBD		0.00122*	-0.000295	0.000783	0.00121*	0.000986
		(0.000651)	(0.000465)	(0.000637)	(0.000670)	(0.000615)
Distance to Street		-0.0408**	-0.0116	-0.0391**	-0.0407**	-0.0373**
		(0.0172)	(0.0119)	(0.0179)	(0.0174)	(0.0167)
Suburban		0.124***	0.0879***	0.125***	0.123***	0.135***
		(0.0294)	(0.0230)	(0.0296)	(0.0292)	(0.0297)
Urban		0.161***	0.125***	0.161***	0.161***	0.156***
		(0.0330)	(0.0274)	(0.0329)	(0.0331)	(0.0330)
Industrial		0.0619***	0.0511***	0.0460**	0.0603***	0.0635***
		(0.0210)	(0.0175)	(0.0207)	(0.0211)	(0.0201)
Traffic		0.109***	0.160***	0.115***	0.109***	0.123***
		(0.0173)	(0.0152)	(0.0176)	(0.0171)	(0.0178)
Stone coal				0.0956***		
				(0.0167)		
Brown coal				0.189***		
				(0.0275)		
Distance to coal plant					-0.0000755	
					(0.000285)	
Av. GDP						-0.0468*
						(0.0243)
Av. Income						0.0737
						(0.0852)
Unemployment share						1.288***
						(0.374)
N	795	795	795	795	795	773
R^2	0.069	0.254	0.568	0.285	0.254	0.286
Districts	107	107	107	107	107	105
Weather	Yes	Yes	Yes	Yes	Yes	Yes
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Coal plant in district	No	No	No	Yes	No	No
Distance to coal plant	No	No	No	No	Yes	No
State FE	No	No	Yes	No	No	No

Note: The table presents OLS model outcomes of regressing $PM_{2.5}$ pollution on log(population density) with varying combinations of control variables. Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), state & year fixed effects, average district-level GDP, average income, share of green party voters, the unemployment share, a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial), distance to the most proximate coal plant, whether a stone coal or a brown coal plant lies within the same district. Standard errors in parentheses are clustered at labour market region - year level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.6: OLS with varying sets of controls (O_3)

	(1)	(2)	(3)	(4)	(5)	(6)
log(pop density)	-0.271***	-0.177***	-0.0854***	-0.162***	-0.143***	-0.128***
	(0.00998)	(0.00945)	(0.00936)	(0.00983)	(0.00931)	(0.00900)
Distance to CBD		0.00230***	0.00112***	0.00270***	0.00232***	0.00167***
		(0.000315)	(0.000282)	(0.000320)	(0.000306)	(0.000277)
Distance to Street		0.0522***	0.0305***	0.0496***	0.0503***	0.0532***
		(0.00727)	(0.00702)	(0.00722)	(0.00724)	(0.00707)
Suburban		-0.139***	-0.177***	-0.134***	-0.142***	-0.144***
		(0.0117)	(0.0127)	(0.0117)	(0.0116)	(0.0115)
Urban		-0.193***	-0.254***	-0.190***	-0.204***	-0.242***
		(0.0138)	(0.0137)	(0.0139)	(0.0134)	(0.0128)
Industrial		-0.0754***	-0.0715***	-0.0580***	-0.0699***	-0.116***
		(0.0116)	(0.0128)	(0.0121)	(0.0110)	(0.0124)
Traffic		-0.231***	-0.265***	-0.234***	-0.230***	-0.218***
		(0.0164)	(0.0146)	(0.0165)	(0.0156)	(0.0141)
Stone coal				-0.0319**		
				(0.0126)		
Brown coal				-0.157***		
				(0.0312)		
Distance to coal plant					0.00131***	
					(0.000148)	
Av. GDP						-0.0155
						(0.0138)
Av. Income						-0.247***
						(0.0572)
Unemployment share						1.401***
						(0.157)
N	3776	3776	3776	3776	3776	3722
R^2	0.260	0.445	0.553	0.450	0.460	0.507
Districts	248	248	248	248	248	247
Weather	Yes	Yes	Yes	Yes	Yes	Yes
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Coal plant in district	No	No	No	Yes	No	No
Distance to coal plant	No	No	No	No	Yes	No
State FE	No	No	Yes	No	No	No

Note: The table presents OLS model outcomes of regressing O_3 pollution on log(population density) with varying combinations of control variables. Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), state & year fixed effects, average district-level GDP, average income, share of green party voters, the unemployment share, a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial), distance to the most proximate coal plant, whether a stone coal or a brown coal plant lies within the same district. Standard errors in parentheses are clustered at labour market region - year level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.7: OLS and IV regressions on municipality level

	NO_2		PM_{10}		$PM_{2.5}$		O_3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
log(pop density)	0.142***	0.132***	0.0337***	0.0525***	-0.00285	0.0140	-0.0750***	-0.0664***
	(0.0117)	(0.0135)	(0.00541)	(0.00751)	(0.00826)	(0.0131)	(0.00677)	(0.00807)
Distance to CBD	-0.00284***	-0.00294***	0.00121***	0.00163***	0.000344	0.000344	0.00183***	0.00198***
	(0.000697)	(0.000705)	(0.000363)	(0.000409)	(0.000550)	(0.000608)	(0.000433)	(0.000459)
Distance to Street	-0.126***	-0.128***	-0.0441***	-0.0461***	-0.0150	-0.00222	0.0457***	0.0473***
	(0.0179)	(0.0188)	(0.00751)	(0.00771)	(0.0144)	(0.0141)	(0.0102)	(0.0107)
Suburban	0.180***	0.180***	0.00961	-0.00511	0.120***	0.0980***	-0.0731***	-0.0718***
	(0.0260)	(0.0278)	(0.0127)	(0.0138)	(0.0300)	(0.0307)	(0.0169)	(0.0177)
Urban	0.342***	0.365***	0.0814***	0.0493***	0.183***	0.143***	-0.137***	-0.152***
	(0.0367)	(0.0400)	(0.0153)	(0.0181)	(0.0330)	(0.0376)	(0.0208)	(0.0226)
Industrial	0.101***	0.103***	0.138***	0.131***	0.0455**	0.0473**	-0.0642***	-0.0553***
	(0.0190)	(0.0194)	(0.0160)	(0.0167)	(0.0222)	(0.0226)	(0.0161)	(0.0144)
Traffic	0.696***	0.701***	0.259***	0.259***	0.114***	0.115***	-0.278***	-0.278***
	(0.0167)	(0.0173)	(0.00854)	(0.00879)	(0.0152)	(0.0149)	(0.0215)	(0.0219)
N	3070	2874	2644	2477	702	663	1947	1823
R^2	0.759	0.760	0.469	0.461	0.222	0.188	0.426	0.428
Municipalities	291	269	269	249	107	99	248	231
First-Stage F-Statistic		151.5		150.6		86.45		123.0

Note: The table presents OLS and IV model outcomes of regressing the respective pollutant concentration on municipality level on log(population density). Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), average municipality-level GDP, average income, share of green party voters, the unemployment share, a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial), and whether the station lies within an environmental zone or not. The instruments used are the combination of historical population from 1910 and soil characteristics. Standard errors in parentheses are clustered at labour market region - year (OLS) and labour market region (IV) level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.8: OLS regressions with Station-specific density in 1km and 5km Buffers

	NO_2		PM_{10}		$PM_{2.5}$		O_3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1km	5km	1km	5km	1km	5km	1km	5km
log(pop density)	0.0977***	0.175***	0.0268***	0.0407***	0.0375**	0.0174	-0.0453***	-0.0842***
	(0.0133)	(0.0228)	(0.00710)	(0.0131)	(0.0143)	(0.0155)	(0.0106)	(0.0149)
Distance to CBD	-0.00577***	-0.00192	0.00116	0.00171*	0.00116	0.00120	0.00245**	0.000575
	(0.00180)	(0.00177)	(0.000762)	(0.000942)	(0.00112)	(0.00113)	(0.000958)	(0.00100)
Distance to Street	-0.0850*	-0.127***	-0.0204	-0.0348*	0.00756	-0.0158	0.00370	0.0323
	(0.0436)	(0.0351)	(0.0175)	(0.0195)	(0.0266)	(0.0307)	(0.0230)	(0.0230)
Suburban	0.121**	0.175***	0.00740	0.0395	0.107*	0.146**	-0.0686*	-0.0845**
	(0.0591)	(0.0590)	(0.0365)	(0.0356)	(0.0585)	(0.0627)	(0.0355)	(0.0359)
Urban	0.215***	0.241***	0.0480	0.0741*	0.119*	0.191***	-0.0981**	-0.103**
	(0.0778)	(0.0773)	(0.0458)	(0.0429)	(0.0688)	(0.0654)	(0.0491)	(0.0498)
Industrial	0.153***	0.105**	0.0839**	0.0811*	0.0670	0.0526	-0.0276	-0.0215
	(0.0450)	(0.0424)	(0.0395)	(0.0431)	(0.0453)	(0.0458)	(0.0485)	(0.0425)
Traffic	0.674***	0.669***	0.233***	0.231***	0.102***	0.105***	-0.318***	-0.328***
	(0.0369)	(0.0358)	(0.0214)	(0.0203)	(0.0367)	(0.0374)	(0.112)	(0.0971)
N	382	392	322	331	130	132	243	253
R^2	0.750	0.800	0.456	0.474	0.264	0.269	0.434	0.500
Districts	217	225	202	209	102	104	196	204

Note: The table presents OLS and IV model outcomes of regressing the respective pollutant concentration on log(population density) with 2015 data. Circular buffers of 1 and 5 kilometres around a measuring station were specified to capture population density within that buffer. Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), average municipality-level GDP, average income, share of green party voters, the unemployment share, a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial), and whether the station lies within an environmental zone or not. Standard errors in parentheses are clustered at labour market region - year level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.9: Alternative measures of density

	Alternative population density		Log of population		Log of Employed per area	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
Panel A: NO_2 ***Regressions***						
log(Alt. density)	0.114***	0.100***				
	(0.00596)	(0.0233)				
log(Population)			0.0967***	0.0891**		
			(0.00801)	(0.0369)		
log(Empl. density)					0.213***	0.213***
					(0.00915)	(0.0405)
N	5575	5273	5575	5273	5528	5226
R^2	0.746	0.744	0.728	0.727	0.752	0.753
Districts	266	246	266	246	264	244
Panel B: PM_{10} ***Regressions***						
log(Alt. density)	0.0289***	0.0369***				
	(0.00318)	(0.0103)				
log(Population)			0.0521***	0.0492***		
			(0.00321)	(0.0155)		
log(Empl. density)					0.0377***	0.0597***
					(0.00530)	(0.0202)
N	4648	4379	4648	4379	4601	4332
R^2	0.469	0.462	0.491	0.487	0.460	0.452
Districts	242	224	242	224	240	222
Panel B: $PM_{2.5}$ ***Regressions***						
log(Alt. density)	0.0135*	0.0157				
	(0.00770)	(0.0233)				
log(Population)			0.0519***	0.0440		
			(0.00779)	(0.0319)		
log(Empl. density)					0.00619	-0.00143
					(0.0114)	(0.0374)
N	795	743	795	743	774	722
R^2	0.252	0.239	0.287	0.275	0.252	0.240
Districts	107	100	107	100	105	98
Panel B: O_3 ***Regressions***						
log(Alt. density)	-0.0775***	-0.0572***				
	(0.00432)	(0.0143)				
log(Population)			-0.0688***	-0.0532**		
			(0.00551)	(0.0239)		
log(Empl. density)					-0.132***	-0.120***
					(0.00641)	(0.0250)
N	3776	3568	3776	3568	3751	3543
R^2	0.438	0.430	0.398	0.396	0.439	0.441
Districts	248	231	248	231	247	230

Note: The table presents OLS and IV model outcomes of regressing the respective pollutant concentration on different measures of density. These are (in logs): population divided by district area (Alt. density); population; and employment per square kilometre (Empl. density). Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial). The instruments used are a combination of historical population from 1910 and soil characteristics. Standard errors in parentheses are clustered at labour market region - year (OLS) and labour market region (IV) level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.10: Interacting population density with station type

	NO_2		PM_{10}		$PM_{2.5}$		O_3	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
log(pop density)	0.251***	0.205***	0.0531***	0.0529*	0.0201	-0.0121	-0.165***	-0.124***
	(0.0162)	(0.0678)	(0.00720)	(0.0289)	(0.0171)	(0.0491)	(0.00946)	(0.0343)
Industrial*Density	0.116***	0.134	0.0364*	0.0295	0.0536*	0.0390	-0.164***	-0.189***
	(0.0256)	(0.0822)	(0.0196)	(0.0653)	(0.0289)	(0.0481)	(0.0178)	(0.0479)
Traffic*Density	0.0752***	0.115	0.0701***	0.0764**	0.0448	0.0920	-0.0352	-0.0746
	(0.0228)	(0.0766)	(0.0114)	(0.0353)	(0.0338)	(0.0798)	(0.0324)	(0.0775)
Distance to CBD	-0.00317***	-0.00384***	0.000868***	0.000590	0.00127*	0.000402	0.00228***	0.00283***
	(0.000446)	(0.00147)	(0.000296)	(0.000891)	(0.000662)	(0.00149)	(0.000310)	(0.000967)
Distance to Street	-0.106***	-0.104***	-0.0434***	-0.0363*	-0.0410**	-0.0284	0.0529***	0.0517**
	(0.0111)	(0.0367)	(0.00635)	(0.0187)	(0.0171)	(0.0437)	(0.00727)	(0.0242)
Suburban	0.294***	0.294***	0.0802***	0.0761***	0.132***	0.153**	-0.147***	-0.145***
	(0.0173)	(0.0517)	(0.00969)	(0.0268)	(0.0302)	(0.0660)	(0.0117)	(0.0348)
Urban	0.457***	0.476***	0.147***	0.137***	0.170***	0.198***	-0.197***	-0.210***
	(0.0224)	(0.0746)	(0.0109)	(0.0320)	(0.0334)	(0.0726)	(0.0137)	(0.0443)
Industrial	-0.791***	-0.925	-0.139	-0.0896	-0.339	-0.200	1.129***	1.324***
	(0.197)	(0.631)	(0.144)	(0.475)	(0.218)	(0.376)	(0.131)	(0.364)
Traffic	0.0574	-0.257	-0.297***	-0.343	-0.238	-0.607	0.0416	0.344
	(0.180)	(0.606)	(0.0914)	(0.280)	(0.267)	(0.646)	(0.255)	(0.613)
N	5575	5273	4648	4379	795	743	3776	3568
R^2	0.756	0.755	0.478	0.474	0.257	0.242	0.450	0.444
Districts	266	246	242	224	107	100	248	231

Note: The table presents OLS and IV model outcomes of regressing the respective pollutant concentration on log(population density). Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), a categorical variable for the station area classification (rural [base category], suburban, urban), station type (background [base category], traffic, industrial). The instruments used are a combination of historical population from 1910 and soil characteristics. Standard errors in parentheses are clustered at labour market region - year (OLS) and labour market region (IV) level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.11: Comparison of satellite data outcomes with station data results

	NO_2			$PM_{2.5}$		
	(1) Station means	(2) Sat. grid	(3) Sat. distr.	(4) Station means	(5) Sat. grid	(6) Sat. distr.
log(pop density)	0.559***	0.391***	0.396***	0.0724***	0.0983***	0.127***
	(0.0209)	(0.0211)	(0.0203)	(0.00989)	(0.0108)	(0.00865)
N	3631	3631	2565	779	779	647
R^2	0.343	0.240	0.245	0.261	0.265	0.477
Districts	268	268	268	109	109	109
Fixed Effects	No	No	No	No	No	No

Note: The table shows the effect of log(population density) on the respective pollutant (NO_2 and PM_{10}) using OLS with 1. satellite data and 2. monitoring station data. The samples are aligned such that only districts/grid cells are included that contain at least one monitoring station. Columns 1 and 4 show baseline regressions with observations of monitoring stations demeaned over districts. Columns 2 and 5: Satellite grid level regressions. Columns 3 and 6: Regressions with satellite observations demeaned over district. Standard errors in parentheses are clustered at labour market region - year level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.12: Mechanisms and sample comparison (NO_2)

	Pub. Transport	Cars	Branches		Green Space	LEZ	Green Voters	Houses
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Regressions including mechanism variable								
log(pop density)	0.379***	0.212***	0.398***	0.319***	0.305***	0.268***	0.300***	0.247***
	(0.0337)	(0.0139)	(0.0225)	(0.0146)	(0.0139)	(0.0143)	(0.0149)	(0.0177)
log(Nr. of transit users)	-0.0430***							
	(0.00912)							
log(Nr. of cars)		0.0424***						
		(0.00809)						
Share Empl. Production			0.761**					
			(0.347)					
Share Empl. Manuf.			0.0598					
			(0.343)					
Share Empl. Construction			-0.974***					
			(0.337)					
Share Empl. Agric.			4.937***					
			(0.657)					
Share Empl. Finance				-0.563***				
				(0.120)				
Share Empl. Public				-1.314***				
				(0.0792)				
Share Empl. Trade				0.382***				
				(0.127)				
Green Space					-0.00204***			
					(0.000219)			
Env. Zone (red)						0.134***		
						(0.0242)		
Env. Zone (yellow)						0.0711**		
						(0.0327)		
Env. Zone (greem)						0.0957***		
						(0.0261)		
Green Voters							0.399**	
							(0.170)	
log(Nr. of houses)								0.0388***
								(0.00909)
N	1554	5405	5528	5528	3583	5575	4965	3219
R^2	0.783	0.755	0.769	0.775	0.768	0.757	0.753	0.772
Districts	143	264	264	264	253	266	263	251
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Regression without mechanism variables								
log(pop density)	0.257***	0.258***	0.282***	0.282***	0.271***	0.280***	0.321***	0.275***
	(0.0245)	(0.0144)	(0.0143)	(0.0143)	(0.0169)	(0.0141)	(0.0126)	(0.0175)
N	1554	5405	5528	5528	3583	5575	4965	3219
R^2	0.776	0.753	0.755	0.755	0.764	0.755	0.752	0.770
Districts	143	264	251	264	253	266	263	251
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table shows the effect of log(population density) on NO_2 with varying sets of control variables using OLS. Panel A contains the regression results of main interest. Panel B shows the corresponding outcome with the same sample composition as Panel A, but without adding the respective set of controls. All regressions include the following basic controls: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), a categorical variable for the station area classification (rural [base category], suburban, urban), and station type (background [base category], traffic, industrial). Standard errors in parentheses are clustered at labour market region - year level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.13: Mechanisms and sample comparison (PM_{10})

	Pub. Transport	Cars	Branches		Green Space	LEZ	Green Voters	Houses
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Regressions including mechanism variable								
log(pop density)	0.0545***	0.0213**	0.145***	0.0632***	0.0584***	0.0697***	0.122***	0.0397***
	(0.0203)	(0.00847)	(0.0121)	(0.00949)	(0.00926)	(0.00686)	(0.00930)	(0.00954)
log(Nr. of transit users)	-0.00454							
	(0.00605)							
log(Nr. of cars)		0.0476***						
		(0.00434)						
Share Empl. Production			0.313					
			(0.299)					
Share Empl. Manuf.			-0.440					
			(0.296)					
Share Empl. Construction			2.073***					
			(0.220)					
Share Empl. Agric.			0.647*					
			(0.351)					
Share Empl. Finance				0.0201				
				(0.0880)				
Share Empl. Public				-0.0782				
				(0.0602)				
Share Empl. Trade				0.427***				
				(0.0944)				
Green Space					0.000961***			
					(0.000197)			
Env. Zone (red)						0.0814***		
						(0.0200)		
Env. Zone (yellow)						0.0621***		
						(0.0213)		
Env. Zone (greem)						-0.00318		
						(0.0184)		
Green Voters							-1.374***	
							(0.110)	
log(Nr. of houses)								0.0454***
								(0.00582)
N	1317	4507	4601	4601	3069	4648	4195	2775
R^2	0.489	0.492	0.494	0.477	0.481	0.478	0.484	0.500
Districts	138	242	240	240	239	242	240	236
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Regression without mechanism variables								
log(pop density)	0.0418***	0.0704***	0.0744***	0.0744***	0.0717***	0.0749***	0.0562***	0.0705***
	(0.0145)	(0.00732)	(0.00692)	(0.00692)	(0.00897)	(0.00691)	(0.00822)	(0.00937)
N	1317	4507	4601	4601	3069	4648	4195	2775
R^2	0.488	0.473	0.472	0.472	0.475	0.474	0.456	0.483
Districts	138	242	236	240	239	242	240	236
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table shows the effect of log(population density) on PM_{10} with varying sets of control variables using OLS. Panel A contains the regression results of main interest. Panel B shows the corresponding outcome with the same sample composition as Panel A, but without adding the respective set of controls. All regressions include the following basic controls: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), a categorical variable for the station area classification (rural [base category], suburban, urban), and station type (background [base category], traffic, industrial). Standard errors in parentheses are clustered at labour market region - year level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.14: Mechanisms and sample comparison ($PM_{2.5}$)

	Pub. Transport	Cars	Branches		Green Space	LEZ	Green Voters	Houses
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Regressions including mechanism variable								
log(pop density)	-0.0318	-0.0106	0.0634**	-0.0355*	0.0223	0.0101	0.0839***	0.0228
	(0.0294)	(0.0187)	(0.0267)	(0.0200)	(0.0165)	(0.0163)	(0.0179)	(0.0181)
log(Nr. of transit users)	0.00729							
	(0.00925)							
log(Nr. of cars)		0.0533***						
		(0.00997)						
Share Empl. Production			0.00941					
			(0.868)					
Share Empl. Manuf.			-0.174					
			(0.860)					
Share Empl. Construction			2.070***					
			(0.445)					
Share Empl. Agric.			-1.208*					
			(0.686)					
Share Empl. Finance				0.937***				
				(0.197)				
Share Empl. Public				-0.335***				
				(0.107)				
Share Empl. Trade				0.388**				
				(0.160)				
Green Space					0.00183***			
					(0.000491)			
Env. Zone (red)						0.188***		
						(0.0417)		
Env. Zone (yellow)						0.104***		
						(0.0283)		
Env. Zone (greem)						0.0307		
						(0.0277)		
Green Voters							-1.608***	
							(0.221)	
log(Nr. of houses)								0.0362***
								(0.0107)
N	377	786	774	774	749	795	741	713
R^2	0.330	0.291	0.279	0.300	0.282	0.281	0.292	0.284
Districts	62	107	105	105	107	107	106	107
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Regression without mechanism variables								
log(pop density)	-0.0148	0.0317*	0.0315*	0.0315*	0.0343**	0.0353**	0.0206	0.0421**
	(0.0218)	(0.0162)	(0.0162)	(0.0162)	(0.0163)	(0.0161)	(0.0162)	(0.0167)
N	377	786	774	774	749	795	741	713
R^2	0.328	0.261	0.256	0.256	0.267	0.254	0.244	0.271
Districts	62	107	107	105	107	107	106	107
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table shows the effect of log(population density) on $PM_{2.5}$ with varying sets of control variables using OLS. Panel A contains the regression results of main interest. Panel B shows the corresponding outcome with the same sample composition as Panel A, but without adding the respective set of controls. All regressions include the following basic controls: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), a categorical variable for the station area classification (rural [base category], suburban, urban), and station type (background [base category], traffic, industrial). Standard errors in parentheses are clustered at labour market region - year level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.15: Mechanisms and sample comparison (O_3)

	Pub. Transport	Cars	Branches		Green Space	LEZ	Green Voters	Houses
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Regressions including mechanism variable								
log(pop density)	-0.200***	-0.132***	-0.232***	-0.202***	-0.171***	-0.177***	-0.156***	-0.140***
	(0.0220)	(0.0108)	(0.0156)	(0.00977)	(0.00999)	(0.00957)	(0.0128)	(0.0118)
log(Nr. of transit users)	0.0163***							
	(0.00548)							
log(Nr. of cars)		-0.0351***						
		(0.00678)						
Share Empl. Production			-1.340***					
			(0.396)					
Share Empl. Manuf.			0.601					
			(0.401)					
Share Empl. Construction			1.563***					
			(0.257)					
Share Empl. Agric.			-3.026***					
			(0.406)					
Share Empl. Finance				0.430***				
				(0.108)				
Share Empl. Public				1.174***				
				(0.0707)				
Share Empl. Trade				-0.621***				
				(0.105)				
Green Space					0.00131***			
					(0.000169)			
Env. Zone (red)						-0.0845***		
						(0.0208)		
Env. Zone (yellow)						-0.0350		
						(0.0303)		
Env. Zone (greem)						0.0635**		
						(0.0272)		
Green Voters							-0.715***	
							(0.143)	
log(Nr. of houses)								-0.0256***
								(0.00772)
N	877	3632	3751	3751	2334	3776	3473	2099
R^2	0.473	0.449	0.486	0.512	0.463	0.447	0.452	0.474
Districts	123	247	247	247	232	248	246	232
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Regression without mechanism variables								
log(pop density)	-0.156***	-0.163***	-0.177***	-0.177***	-0.155***	-0.177***	-0.190***	-0.155***
	(0.0176)	(0.00948)	(0.00944)	(0.00944)	(0.0102)	(0.00945)	(0.0107)	(0.0108)
N	877	3632	3751	3751	2334	3776	3473	2099
R^2	0.467	0.443	0.446	0.446	0.456	0.445	0.447	0.470
Districts	123	247	232	247	232	248	246	232
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table shows the effect of log(population density) on O_3 with varying sets of control variables using OLS. Panel A contains the regression results of main interest. Panel B shows the corresponding outcome with the same sample composition as Panel A, but without adding the respective set of controls. All regressions include the following basic controls: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, time dummies (day of week, hour, month), a categorical variable for the station area classification (rural [base category], suburban, urban), and station type (background [base category], traffic, industrial). Standard errors in parentheses are clustered at labour market region - year level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.16: Annual and daily/hourly transgression probabilities (Probit)

	NO_2			PM_{10}			$PM_{2.5}$		
	(1) Probit	(2) Probit Hist. IV	(3) Probit Soil IV	(4) Probit	(5) Probit Hist. IV	(6) Probit Soil IV	(7) Probit	(8) Probit Hist. IV	(9) Probit Soil IV
Panel A: Annual threshold transgressions									
log(pop density)	1.477***	1.228***	1.818***	0.476***	0.644***	0.383*	-0.314	-0.624	-0.652
	(0.218)	(0.252)	(0.281)	(0.104)	(0.125)	(0.213)	(0.243)	(0.435)	(0.616)
N	5663	5383	5635	4812	4560	4786	702	651	695
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Districts	269	269	269	247	247	247	109	109	109
Marginal Effects	0.2006	0.2532	0.1456	0.0948	0.0348	0.1064	-0.0162	-0.0051	-0.0034
Panel B: Transgressing daily/hourly thresholds specific nr. of days									
	NO_2			PM_{10}			$PM_{2.5}$		
	> 17 (days)	> 14	> 9	> 34	> 29	> 24	> 34	> 29	> 24
log(pop density)	1.141***	1.219***	1.183***	0.373***	0.381***	0.280**	-0.124	-0.0140	-0.00827
N	2125	2125	2125	4812	4812	4812	795	791	791
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Districts	269	269	269	247	247	247	109	109	109
Marginal Effects	0.0528	0.0588	0.0788	0.0510	0.0656	0.0587	-0.0312	-0.0031	-0.0014

Note: The table presents outcomes of the effect of log(population density) on two outcomes on district level. In Panel A, the dependent variable is a dummy variable indicating whether the annual threshold value was transgressed. In Panel B the dependent variable is a dummy variable indicating whether the daily threshold value was transgressed on a pre-specified number of days. The respective thresholds for Panel A are: 40 $\mu g/m^3$ for NO_2, 10 $\mu g/m^3$ for $PM_{2.5}$, and 20 $\mu g/m^3$ for PM_{10}. The respective thresholds in Panel B are: 200 $\mu g/m^3$ for NO_2, 25 $\mu g/m^3$ for $PM_{2.5}$, and 50 $\mu g/m^3$ for PM_{10}. The number of days (e.g. 17, 14, and 9 for NO_2) lie just below the number of daily exceedances allowed by the EU, which are 18 for NO_2 and 35 for PM_{10}. The regressions are performed using Probit Models and in the case of annual transgressions also Probit IV. Control variables included are: Distance of the pollutant measurement station to the CBD, distance to a major street, weather variables [precipitation, sunshine, wind speed, cloudiness, air pressure] and their interactions, a categorical variable for the station area classification (rural [base category], suburban, urban), and station type (background [base category], traffic, industrial). The instruments used are a combination of historical population from 1910 and soil characteristics. Standard errors in parentheses are clustered at labour market region level. Statistical significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Urban pollution: A global perspective[1]

Abstract

We use worldwide satellite data to analyse how population size and density affect urban pollution. We find that density significantly increases pollution exposure. Looking only at urban areas, we find that population size affects exposure more than density. Moreover, the effect is driven mostly by population commuting to core cities rather than the core city population itself. We analyse heterogeneity by geography and income levels. By and large, the influence of population on pollution is greatest in Asia and middle-income countries. A counterfactual simulation shows that $PM_{2.5}$ exposure would fall by up to 36% and NO_2 exposure up to 53% if within countries population size were equalized across all cities.

[1]Co-authored with Rainald Borck.

3.1 Introduction

Pollution is an important determinant of urban quality of life. Households have flocked to cities over the last centuries and decades, attracted by various agglomeration economies, such as higher productivity and wages. However, city life has been and still is, to different extents, plagued by agglomeration costs stemming from crime, congestion, and pollution. Besides, urbanization and environmental degradation are not evenly spread throughout the world. While developed countries were already more than 50% urbanized by 1950, this threshold has been reached by less developed countries only in 2020.[2] Accordingly, the bulk of the recent and imminent increase in world urbanization will occur in developing regions. This is also where urban air pollution is most severe. For example, taking the average $PM_{2.5}$ concentration value from the WHO air quality database from 2022,[3] 20 of the 25 dirtiest cities were located in India, China, Bangladesh or Pakistan, with the remaining in Cameroon, Iran, Mongolia, Madagascar and Afghanistan. Hence, the relationship between agglomeration and pollution is also a question of socio-economic development. Reigning in pollution, especially in large cities, will be important as developing countries thrive to improve their citizens' well-being. Yet, while there is an extensive literature on the benefits of agglomeration economies, there is much less research on the corresponding costs (Ahlfeldt and Pietrostefani, 2019).

In this paper, we contribute to filling this gap. We use global gridded data on air pollution and population to analyse how agglomeration, in the form of large and densely populated cities, affects exposure to $PM_{2.5}$ and NO_2 pollution.

In theory, population density might increase or decrease pollution concentration in cities. Borck and Schrauth (2021) present a model where residents of a monocentric city pollute due to commuting and residential energy use for heating, electricity, etc. They show that population density increases pollution concentration. The reason is that larger and more densely populated cities have more aggregate commuting and that residential energy use increases as well, even though residents live in smaller dwellings on average.[4] However,

[2]See United Nations (UN) Urbanization Prospects, https://population.un.org/wup/.

[3]See WHO Air Quality Database from April, 2022, https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database.

[4]In the model, total pollution increases more than urban area, implying higher pollution concentrations.

there are some countervailing forces. For instance, public transit is more viable in large and densely populated cities due to economies of density, and denser housing is more energy efficient. Therefore, the relation between density and pollution is theoretically ambiguous.

Similar opposing forces determine whether cities with larger total population (as opposed to density) are more polluted (see Borck and Pflüger, 2019; Borck and Tabuchi, 2019). Further, the relation between population density and pollution is likely to depend on many factors that vary between regions, such as geography, institutions (environmental policies) etc. Therefore, an interesting question that we look at is how pollution and its relation with density varies between regions with different characteristics.

We use 11-16 years (depending on the pollutant) of gridded satellite data to document the distribution of pollution over space and time. There are several main findings. First, we show that about 3/4 of the world population and about 79 percent of city dwellers live in places with particulate pollution above thresholds as recommended by the WHO. Thus it seems that pollution is especially severe in cities. We go on to estimate the elasticity of pollution with respect to population density for $PM_{2.5}$ and NO_2. Using OLS regressions with country fixed effects, we find elasticities of 0.15-0.16 for NO_2 and 0.02-0.03 for $PM_{2.5}$. To tackle concerns of reverse causality and omitted variables, we also instrument population density using historical populations from different periods in time. Doing so has only a very small effect on the estimated elasticities.

We present our results using both grid cells and cities (Functional Urban Areas, FUAs) as units of observation. Examining cities allows us to explicitly differentiate between the different effects of agglomeration size versus population density on exposure. For cities, we find that population size seems to be more important than density. Furthermore, using the definition of FUAs allows us to differentiate between the core city and its surrounding commuting zone. In fact, it turns out that pollution exposure is not significantly affected by core city population, but does rise significantly with population living in FUAs' commuting zone.

Moreover, we study how the pollution-density relationship varies over continents and by income. For the rasterized global data, we find that the pollution-density relation is strongest in middle income countries and in Asia. For the city data, population/density affect pollution most in upper middle and high income countries as well as in Europe and North America.

We also present outcomes on a more local level by estimating the effect of within-city variations in density. Again, we find positive effects of density on exposure, but the effects are mostly smaller in size. Additionally, we estimate spatial first difference regressions, where the estimated elasticities are based on changes between neighbouring grid cells (Druckenmiller and Hsiang, 2018). The corresponding coefficient estimates turn out to be positive, but again smaller in size.

Lastly, we perform a simple counterfactual simulation. Using the exposure–population elasticity from our city analysis, estimated separately for each country, we ask how each country's total exposure would be affected by an equal redistribution of population across cities. We find that for $PM_{2.5}$, exposure falls by 36.5% for the country with the largest drop (Indonesia), which has a large estimated elasticity. Conversely, there are some countries with negative elasticities, so exposure would rise in this counterfactual by a maximum of 22.5% in Senegal.

The study contributes to a small but growing economic literature on urban pollution generally, and on the relation between agglomeration and pollution in particular. Empirical papers in fields other than economics have largely been confined to cross-sectional studies.[5] However, omitted variables and reverse causality are difficult to tackle in these settings. Among the few serious efforts to identify the causal effect of population density on pollution are Borck and Schrauth (2021) and Carozzi and Roth (2020). Borck and Schrauth (2021) use panel data from German districts, while Carozzi and Roth (2020) use cross-sectional data from US metropolitan areas. Both papers instrument density with a variety of historical and geological instruments. Castells-Quintana et al. (2021) and Aldeco et al. (2019) also study global pollution. Aldeco et al. (2019) focus on studying the effect of various policies using a spatial equilibrium model. Castells-Quintana et al. (2021) is also closely related to our paper, but there are several differences. They study emissions in a global panel of cities, while we analyse exposure in both cities and raster cells, which allows for a truly global analysis and lets us study the urban-rural pollution gradient in addition to cross-city differences. Moreover, we do a variety of heterogeneity analyses, and instead of emissions, we look at pollution exposure which is more tightly linked to local welfare.

[5]See, e.g. Sarzynski (2012) and Lamsal et al. (2013). See also Borck and Schrauth (2021) for further references.

The paper is organized as follows. The next section presents our data, descriptive analyses and empirical approach. Section 3.3 shows the results. In Section 3.4, we simulate how total exposure would change if, within countries, we were to redistribute population equally among all cities. The last section concludes the paper.

3.2 Data and estimation

3.2.1 Data

Most data sets we use are derived from satellites and are provided as a grid of raster cells covering the entire world. Those rasterized grid maps come in different resolutions, mostly between 0.01 and 0.25 decimal degrees. We transform all data to 0.25 degree raster cells, which is a compromise between the different levels of aggregation of the native data and moreover alleviates concerns about auto-correlation at finer scales. At the equator, a quarter degree grid corresponds to 27.8 kilometres into one direction or roughly 775 square kilometres overall.[6] For our analyses, we use the years 2000, 2010 and 2015. In the following, the different data sets are described in more detail.[7]

3.2.1.1 Units of observation

In the analyses we use two types of observational units. The first are grid cells spanning the globe. The second are Functional Urban Areas (FUAs) as defined by Moreno-Monroy et al. (2020). These FUAs are cities and their surrounding areas with strong internal commuting links. We view these two data sets as providing complementary results. Hence, defining cities gets us closer to measuring activity in economically meaningful areas. Conversely, using all grid cells – even very thinly populated ones – allows us to measure an urban-rural gradient of pollution. Thus, our paper differs from and complements other papers that have mostly studied cities only (e.g. Carozzi and Roth, 2020; Castells-Quintana et al., 2021).[8]

[6]Moving away from the equator means that equally sized grids cover smaller areas due to the curvature of the earth. At the 45th degree of latitude for example, which crosses South Dakota, Mongolia, France and Italy, 0.01 decimal degrees are equal to 787.1 meters in one direction. The value approaches zero at the poles. Most of human activity takes place between the 50th parallel south and the 60th parallel north.

[7]The NO_2 data is only available from the years 2000 to 2012, of which we use the years 2000 and 2010. The $PM_{2.5}$ data is available until 2015.

[8]Castells-Quintana et al. (2021) also look at the effect of density and polycentricity on pollution at the country level.

Raster data. The first units of observations in our analysis are grid cells of a rasterized world map. The majority of the data we use is provided as raster maps, which then can be matched to each other geographically. The advantage of looking at grid cells is that we abstract from defining cities or urban areas and that there is an increasing database of worldwide data covering a wide range of topics. In addition, it will allow us to measure an urban-rural pollution gradient since observations take into account any type of inhabited land and do not depend on city definitions. We use grids of 0.25 decimal degrees and aggregate all the other raster maps to this size. This leaves us with more than 240,000 cells that fall on land to which we make some minor adjustments.[9] The chosen grid size is a compromise between data that is available at relatively fine grid scale, and data that is available at coarser levels only. It also mitigates concerns about spatial auto correlations.[10]

Cities. There are several reasons why we want to define city delineations. First, defining cities allows us to distinguish between city size and density. In a grid with equally sized grid cells, density would be strictly proportional to population. While basic urban economic theory also predicts a positive relation between population size and density (e.g. Brueckner (1987)), in practice the two vary independently, for instance, due to differences in zoning policies across cities. Since different agglomeration economies and diseconomies may operate at different spatial scales, population size and density might then affect pollution differently (Ahlfeldt and Pietrostefani, 2019; Cheshire and Magrini, 2008). Second, some of the $PM_{2.5}$ pollution stems from sources not directly attributable to daily human activities such as volcanoes or wildfires (see e.g. NASA Earth Observatory (2015)). While this may be interesting in its own right (if it leads to rural areas being dirtier than they would otherwise be), abstracting from these types of events by focusing on urban areas allows us to concentrate on the effect of human activity in cities on pollution. Third, we can conduct between city analyses to supplement the rural-urban gradient. This type of city size effect helps us connect the empirical analysis with theoretical considerations about optimal city size (see

[9] We drop grid cells which cannot be assigned unambiguously to one single country. As a consequence, about 21,400 grid cells that lie at country borders are dropped from the sample. Furthermore, we harmonize the country composition of our city and grid cell samples. Thus, all countries which do not contain at least one Functional Urban Area are dropped.

[10] For variables available at higher grid resolutions, e.g. 0.1 decimal degrees, we re-project the data to $\frac{1}{4}$ degrees using an appropriate function: For continuous variables, we calculate either the mean of all smaller grids within the respective quarter degree grid (pollution exposure for example is mean exposure within a quarter degree grid) or sum the values of these finely scaled grids, as appropriate. For categorical variables we take the modal value within a quarter degree grid.

e.g. Borck and Tabuchi (2019)). Fourth, we can detect within city differences. Thus, we can study whether there is a core-periphery gradient of pollution exposure within cities and we can compare it to between-city effects or the urban-rural gradient of pollution exposure. Lastly, our historical population instruments consist of geo-coded city locations. Directly instrumenting urban areas rather than grid cells thus seems more adequate.

We define cities as FUAs following Moreno-Monroy et al. (2020).[11] They use population and travel time data in 2015 to determine unique urban centres and their commuting zones. Thus, our city definitions do not vary over time. A FUA consists of an urban core with at least 50,000 inhabitants and the surrounding commuting zone, which is constructed using travel times. FUAs were originally defined by the Organisation for Economic Co-operation and Development (OECD) for OECD countries and Colombia. Moreno-Monroy et al. (2020) use those OECD-defined FUAs to estimate city boundaries for the rest of the world.

Figure 3.1 shows our two main units of analysis and the population distribution as provided by LandScan for the north-eastern USA. More precisely, the figure visualizes the New York, Philadelphia, Baltimore and Washington, D.C. area and the area's FUAs. Grid-cell analyses consider all the non-white grid cells that fall on land. Densely populated city cores are shown in red, and less dense suburbs and rural areas are shown in yellow and blue. The greyish transparent polygons overlaying the population grids depict the resulting FUAs.

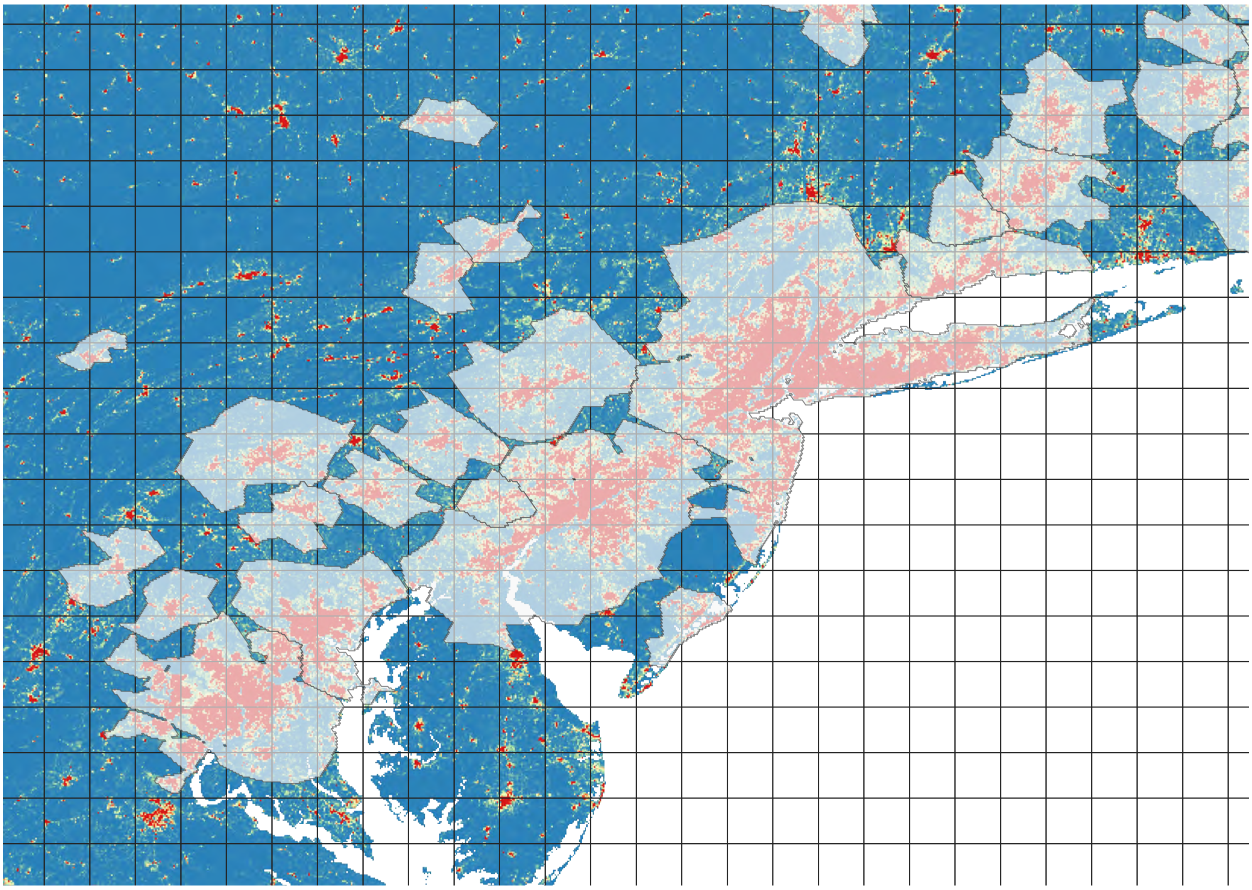
Overall, there are 9,031 FUAs in 188 countries and about 245,000 grid cells in 185 countries. Since the main part of analyses contains within-country effects, we mostly restrict the samples such that very small countries with very few raster cells or cities are dropped.

3.2.1.2 Air pollution data

The most direct measure of ground-level pollution concentration would be measurements from in-situ monitors. However, these are not widely available, especially in many lower income countries. Even among wealthy nations, only selected areas contain monitoring stations. To get coverage of world wide pollution we use satellite data, which captures air

[11] The geo-package with the FUA shapefile is publicly available at https://ghsl.jrc.ec.europa.eu/download.php.

Figure 3.1: Grid-cell units, Functional Urban Areas and LandScan population. New York, Philadelphia, Washington D.C.



Note: The graph depicts a small section of the entire sample. White areas are water surfaces (lakes and oceans) that are excluded from estimations. The overlying grid squares correspond to the raster units, while the greyish transparent polygons show single FUAs. Within both observational units, LandScan population data are shown in different colour gradations. Blue indicates very low population density and red indicates very high population density.

pollution concentration as vertical column densities in the troposphere. For ground-level observations, we resort to data sets that use chemical transport models to translate satellite measures into ground-level pollution.[12] The data products are annual means of dust and sea-salt removed $PM_{2.5}$ from 2000 until 2015 at 0.01x0.01 decimal degree (dd) resolution and annual means of NO_2 without corrections from 2000 until 2010 at 0.1x0.1 dd resolution.[13] For most of our analysis, we use the data that are weighted using Geographically Weighted

[12]See Hammer et al. (2020) and Van Donkelaar et al. (2016) for $PM_{2.5}$ and Lamsal et al. (2008) for NO_2. This data is available online (Atmospheric Composition Analysis Group, 2018).

[13]Satellite measurements of pollution are captured between 8:30 a.m. and 11:00 a.m. local time, depending on the satellite. The cell size means that a raster contains about one square kilometre at the equator compared to the 120 km^2 NO_2 raster cells.

Regression (GWR), but outcomes are not sensitive to using the non-weighted data.[14]

This native data serves to construct our pollution measure of interest, which is pollution exposure (see e.g. Carozzi and Roth (2020) or Aldeco et al. (2019)). Population-weighted pollution exposure E in grid cell G is then given by:

$$E_G = \sum_{i \in G} P_{G,i} \times \frac{N_{G,i}}{\sum_{j \in G} N_{G,j}},$$

where i indexes small grid cells within a large grid cell G.[15] Average pollution concentration in cell i is given by $P_{G,i}$ and population in i is represented by $N_{G,i}$. Average pollution exposure is therefore the sum of grid-specific pollution exposure divided by overall population in a large unit G. We also repeat the analysis with our FUA dataset, where G represents a city instead of a raster cell.

3.2.1.3 Population density

Present population and density. For measures of population and population density, we use LandScan data (LandScan, 2018). Population in this data set is provided on a very fine spatial scale (30 arc-seconds)[16], obtained from censuses and other sources worldwide. The data aims to show where people are located on average over the course of 24 hours. Thus, it includes place of residence and of work in its estimations.[17] However, there is no information about how exactly the information is implemented in the population grid estimates. Henderson et al. (2021) provide a "ground-truthing" exercise and conclude that LandScan data perform well and are suitable for analyses on a global scale.[18]

[14] GWR uses in-situ (on the ground) monitors to detect regional biases of satellite optical depth measurements. This bias is estimated and then corrected using different predictors like land cover or elevation difference (Van Donkelaar et al., 2015). The advantage of GWR is the more accurate representation of ground-level $PM_{2.5}$.

[15] For example, a $PM_{2.5}$ observation in a cell of 0.01x0.01 *dd* within a larger grid cell that spans over 0.25x0.25 *dd*.

[16] 30 arc seconds correspond to a little less than 0.01x0.01 *dd*, which is about 1km at the equator.

[17] By contrast, other data sets distribute population obtained from administrative sources equally in space or use buildings as a proxy for where people live without distinguishing whether those buildings are commercial or residential ones.

[18] This finding is confirmed by Galdo et al. (2019) who use machine learning combined with human judgment to identify urban areas in India and compare their outcomes with LandScan data.

Historical population measures. We also instrument population or density using historical population data, following a large literature in urban and regional economics since Ciccone and Hall (1996).

The data comes from Reba et al. (2016), who provide geo-referenced population data worldwide ranging from 3700 B.C. to 2000 A.D. using historical, archaeological, and census-based estimates. This data set comprises about 1500 settlements worldwide. We use it to construct two different instruments. The main instrument will be the population in 1900. This is the year with most observations in the historic population sample prior to 1914 and represents the population in industrialized times before the two world wars. The second instrument is population in the last year of observation before 1750, and thus in pre-industrialized times. In Section 3.2.3 below, we will come back to the issue of instrument relevance and exogeneity.

In the raster analysis, we assign historical population to single grid cells. Thus, each grid cell is instrumented with the historical population data of the settlement that lies within this grid cell. Some grid cells contain several data points. In this case, we sum up the population over these data points. The drawback of instrumenting with historical population counts is that the estimation sample is drastically reduced and that there is a concentration of settlements in economically developed countries. Since using those instruments therefore deprives us of much valuable information, we mainly present the IV results in order to compare them with OLS outcomes on a harmonized sample. For most of this study, we will concentrate on within-country OLS estimates.

3.2.1.4 Controls

We use a number of variables in order to control for potential observable factors that may be correlated with population density and pollution. Income is an obvious candidate variable that correlates with population density (Combes and Gobillon, 2015) and affects pollution. To control for economic development, we therefore use GDP from the dataset provided by Kummu et al. (2018). It is based on subnational accounts, like states in the U.S. or districts in Germany. In some specifications we control for the presence of coal-fired and other highly polluting power plants in a grid cell.[19] Since power plants may be close to dense areas, this may be one channel through which density affects pollution.

[19]The data is available at https://datasets.wri.org/dataset/globalpowerplantdatabase.

We also control for a variety of topological and climatological variables that may be correlated with density and pollution, such as ruggedness, temperature, wind speed, and precipitation. We compute ruggedness following Nunn and Puga (2012), which is roughly the grid-cell average difference in elevation between a point and the terrain surrounding it. Our ruggedness measure is calculated over land surface only, leaving out water.[20] Temperature and precipitation are taken as long-run averages over the 30 year period between 1960 and 1990 (FAO/IIASA, 2012). Wind data is retrieved from the global wind atlas (Davis et al., 2019).

In addition to weather, we control for variables that might influence pollution through their suitability for trade on the one hand and through their climatological impact on the other. These controls consist of three dummy variables (whether a major river, a large lake, or a coastline is within 25 kilometres of a grid centroid) and a continuous variable (distance of grid centroid to coast). In our city-level analysis, we measure the distance of a city border to the respective first nature characteristic. Data for coastlines, rivers and lakes come from Natural Earth (2018).[21]

Moreover, we control for the agricultural suitability of a raster cell or city (compare Henderson et al., 2018). Locations that are densely populated due to their fertile soil may also suffer from high pollution, since agriculture is a strong producer of particulate matter pollution. We thus take into account land suitability (as continuous variable), and a set of biome indicators. Land suitability for agriculture is based on measures of climate and soil and predicts the probability of land to be cultivated (Ramankutty et al., 2002). Biomes describe the ecological system of an area and its dominant natural vegetation. These categories include for instance “tundras”, “tropical and subtropical dry broadleaf forests”, or “Deserts & Xeric Shrublands”. The 14 biome indicators we use are taken from Olson et al. (2001).[22]

[20]In order to calculate the fraction of water surface we use the world map gridded at 1 km resolution provided by Lloyd et al. (2017), which is based on a global water mask dummy variable gridded at very high resolution (Feng et al., 2016).

[21]We take the “high” resolution datasets from Natural Earth (2018). Rivers are categorized into 10 size ranks, where 1 are the largest and 10 the smallest rivers. We only consider rivers between the ranks 1-6. Large lakes are those with a surface area greater than 5,000 square kilometres, excluding unnatural dams. This leaves us with 29 major lakes.

[22]Just like Henderson et al. (2018) we combine the categories “tropical and subtropical dry broadleaf forests” with “tropical and subtropical coniferous forests” as well as “tropical and subtropical grasslands and savannas and shrublands” with “flooded grasslands and savannas”. Furthermore, we drop areas historically covered by ice or rocks from the analysis. Doing so does not, however, change our results.

We want to assess a variety of country features that may influence the relation between population and pollution. In order to do so, we make use of a large set of national indicators from the World Bank that contains economic performance, energy use or demographic characteristics (World Bank, 2019). This data set ranges from 1960 to 2018, which allows us to calculate for instance means of urbanization rates and their growth, residents of large agglomerations or energy use over different time periods. In our analysis below, we will correlate some of these measures with the density-elasticity of pollution.

3.2.2 Descriptive analysis

Human health is vulnerable to pollution. The WHO has set short- and long-run thresholds to indicate very high pollutant concentration levels that presumably pose major threats to human health. Short-run refers to the average value over 24 hours for PM_{10} and one hour for NO_2, while long-run means the annual mean. In our sample, the long-term threshold of 40 $\mu g/m^3$ for NO_2 is not exceeded in a single raster or city, which would seem to suggest that this pollutant does not constitute a major health threat.[23] However, Borck and Schrauth (2021) show that NO_2 levels for both the short- and the long-run thresholds are transgressed in Germany on a more local level. The respective data is taken from in-situ monitors and therefore provides a more accurate local measure of air pollution. Indeed, the raster size we have available for NO_2 does not allow for a very local consideration of this pollutant. Furthermore, we only have available annual means, so we cannot analyse short-run threshold transgressions.

Things look different for $PM_{2.5}$-pollution. Table 3.1 shows the share of population that lives in raster cells or cities where annual mean $PM_{2.5}$ pollution exceeds the short-run (25 $\mu g/m^3$) or long-run (10 $\mu g/m^3$) WHO thresholds. In 2015, out of the 7.26 billion people in the sample, about 5.52b lived in raster cells with mean $PM_{2.5}$-levels beyond the long-term threshold of 10 $\mu g/m^3$. This corresponds to around 76 percent of the overall world population. Approximately 39 percent were even permanently exposed to concentrations beyond 25 $\mu g/m^3$, which is the WHO 24-hour mean and therefore recommended to be avoided over periods longer than a day. Note, however, that within a raster cell there is

[23]The WHO has recently updated its thresholds (for example, the annual thresholds are now 10 $\mu g/m^3$ for NO_2 and 5 $\mu g/m^3$ for $PM_{2.5}$.), but we use the thresholds that were published during the period we study.

Table 3.1: Descriptive statistics

	Raster		Cities	
	Mean (Std)	Min Max	Mean (Std)	Min Max
$PM_{2.5}$	8.07	0	31.46	.9
	(8.51)	104.94	(24.82)	119.57
NO_2	.38	0	1.70	0.01
	(0.77)	28.7	(2.31)	17.85
Population	29,779	0	435,016	50,079
	(158,719)	11,501,275	(1,384,120)	36,471,787
Overall population	7.26 billion		3.93 billion	
Share population exposed to:				
$PM_{2.5} > 10$ $\mu g/m^3$	76%		79%	
$PM_{2.5} > 25$ $\mu g/m^3$	39%		44%	
Countries	185		188	
Observations	244,649		9031	

Note: Own calculations. The table shows descriptives of unweighted pollution provided by Atmospheric Composition Analysis Group (2018) for the years 2010 (NO_2) and 2015 ($PM_{2.5}$). Population data is shown for the year 2015 and is taken from LandScan (2018). For the calculation of $PM_{2.5}$-exposure with values larger than 10 $\mu g/m^3$ or 25 $\mu g/m^3$ we sum the population in grid cells (FUAs) that match a pollution cell that is higher than the respective value in 2015 and divide it by the overall population in the respective sample.

variation in pollution concentrations such that not everybody is actually exposed to those pollution concentrations. Therefore, the actual number of people permanently exposed to such high concentrations may lie below 75 percent.

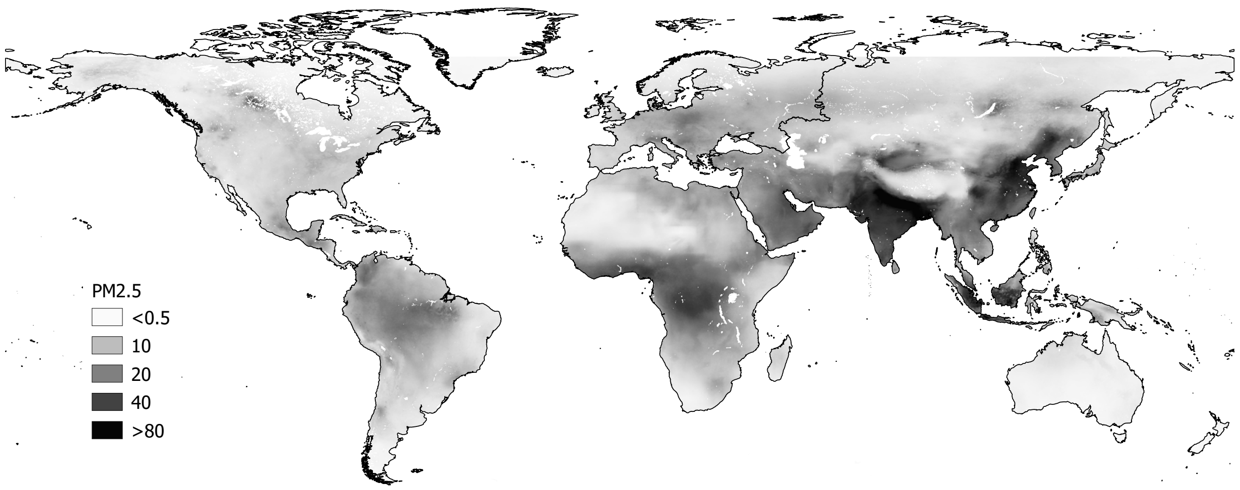
If we only consider FUA, there are about 4 billion people living in urban areas (this is in line with the UN's estimate of worldwide city population) of whom about 79% live in cities with long-term mean $PM_{2.5}$ pollution beyond 10 $\mu g/m^3$. About 44% are in cities where average annual urban pollution even exceeds the short-term threshold of 25 $\mu g/m^3$.

There are marked differences between continents. In Asia, 92 percent of the population face an annual average $PM_{2.5}$-pollution beyond 10 $\mu g/m^3$. In Europe the corresponding number is 67 percent, in Africa 62 percent, in North-America 37 percent, and in South America 32 percent.[24]

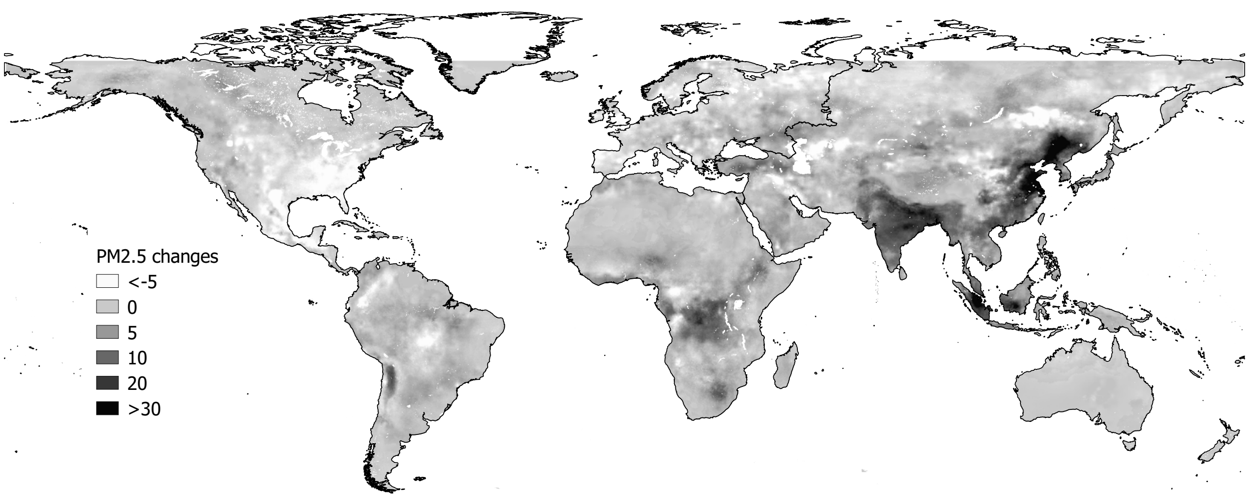
Figure 3.2 shows the geographical distribution of $PM_{2.5}$ in 2015 and its change from 2000 to

[24] These numbers are based on the raster data.

Figure 3.2: $PM_{2.5}$ concentration and development over time



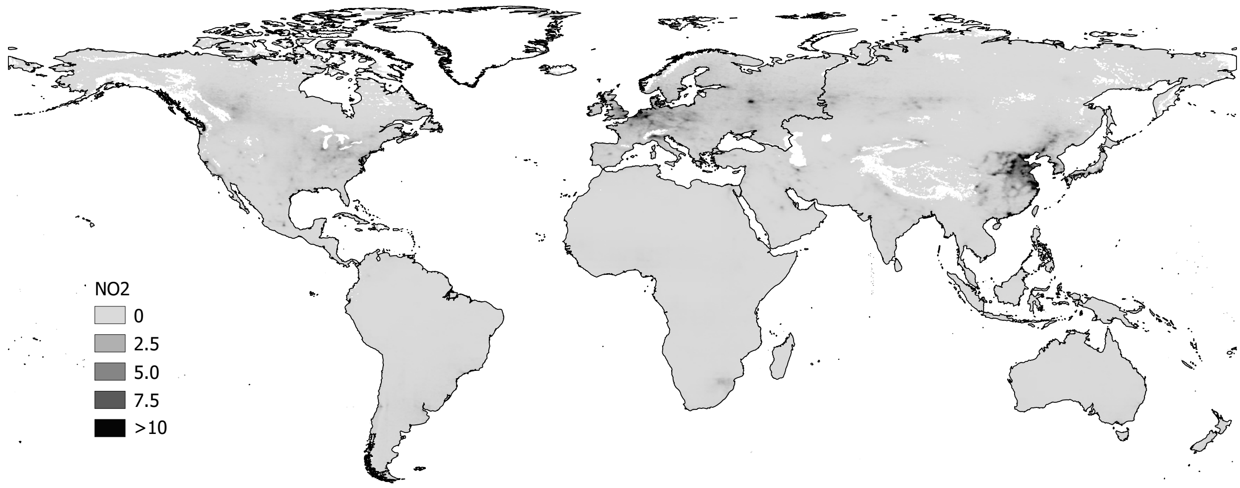
(a) $PM_{2.5}$ concentration in 2015



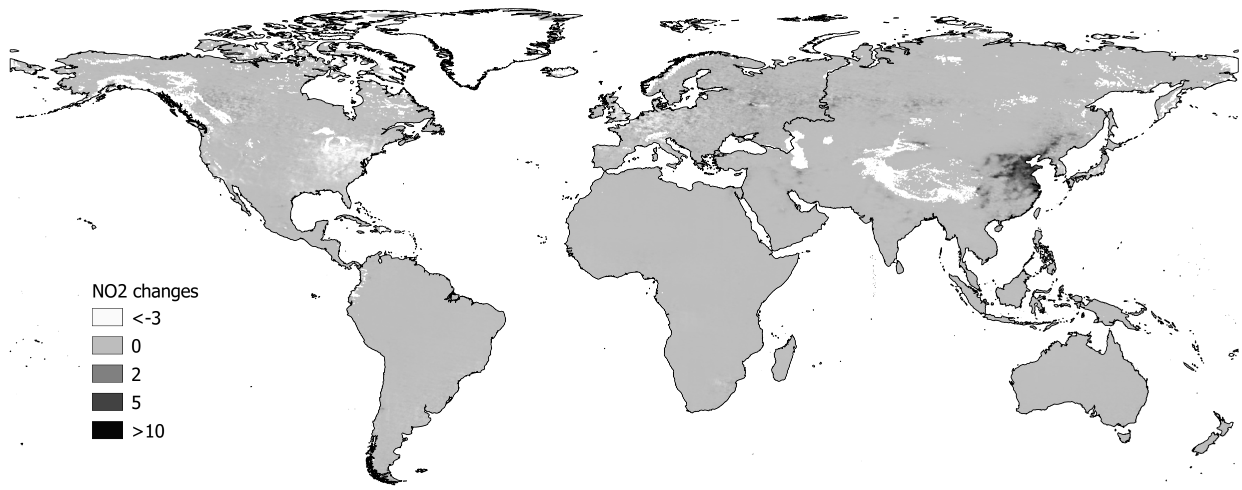
(b) $PM_{2.5}$ change 2000-2015

Note: Figure 3.2a shows the worldwide distribution of $PM_{2.5}$ concentrations in 2015. White/light grey depicts low and dark grey/black high concentration levels. Figure 3.2b shows developments of $PM_{2.5}$ levels between 2000 and 2015. White/light grey areas saw negative, little or no change in pollution concentration; dark grey/black areas have experienced large increases in $PM_{2.5}$ pollution.

Figure 3.3: NO_2 concentration and change over time



(a) NO_2 concentration in 2010



(b) NO_2 change 2000-2010

Note: Figure 3.3a shows the worldwide distribution of NO_2 concentrations in 2010. White/light grey depicts low and dark grey/black high concentration levels. Figure 3.3b shows changes of NO_2 levels between 2000 and 2010. White/light grey areas saw negative, little or no change in pollution concentration; dark grey/black areas have experienced large increases in NO_2 pollution.

2015. Dark grey/black areas in panel (a) are highly polluted with values close to or larger than 25 $\mu g/m^3$ while light grey/white ones may have values close to or below the annual WHO threshold of 10 $\mu g/m^3$. In Panel (b), dark grey/black areas have seen an increase in $PM_{2.5}$-concentrations between 2000 and 2015, while light grey/white areas experienced little change or even a decline. Apparently, many of the highly polluted areas in 2015 either developed into areas with elevated pollution concentrations or became even more polluted over the course of 15 years. The pollution problem has become much more serious especially in India and China, but also in Africa south of the equator (net of dust and sea salt). Many areas in the U.S., especially in the east, have improved their air quality over time, which is also the case in parts of Western Europe.

Pollution of NO_2 is much more concentrated in a few areas as shown in Figure 3.3a. The highest concentration exposure is found in North-eastern China, Middle Europe, parts of the United States and parts of Russia. The range of values is much smaller for NO_2. The maximum value reached is 30 $\mu g/m^3$ in 2010. Figure 3.3b shows the change in NO_2 concentration levels. Again, dark grey/black areas are those where pollution most strongly increased from 2000 until 2010. Some parts of the U.S. and Europe have experienced air quality improvements. In Africa, Australia, and South America, NO_2 concentrations barely changed. Predominantly densely populated metropolitan areas such as Santiago de Chile, Cairo, or São Paulo seem to have higher pollution levels in 2015 compared to 2010.

We now turn to our regression approach to estimate population effects on pollution exposure.

3.2.3 Estimation

In a first step, we run simple Ordinary Least Squares (OLS) regressions of air pollution exposure on population (density) and control variables. To mitigate concerns about spatial autocorrelation, we analyse pollution within relatively large $\frac{1}{4}$ decimal degree grids and we cluster standard errors within three-by-three squares of grid cells times year of analysis (following Henderson et al., 2018). This clustering approach accounts for the potential correlation of pollution in space since particulates for instance disperse spatially with the wind. In a second step, we restrict the sample to cities as defined above.

The OLS regression equation is:[25]

$$\ln(E_{GtS)}) = \beta + \rho \ln(D_{GtS}) + \gamma X_{GtS} + \alpha_t + \epsilon_{GtS}, \tag{3.1}$$

where E_{GtS} is exposure to NO_2 or $PM_{2.5}$ in grid cell/FUA G and year t in country S (in our baseline regressions we only include the last year of observation, i.e. 2010 for NO_2 and 2015 for $PM_{2.5}$). Our parameter of interest, ρ, measures the elasticity of pollution exposure with respect to population density D_{GtS} (or population N_{GtS}, depending on the specification). X_{GtS} is a vector of control variables. As explained above, these contain the log of GDP, several variables about the suitability for trade (whether the raster/city lies on a river, on a lake or on the coast) and agriculture (land suitability, and biome indicators), temperature, precipitation (both as 1960-1990 long term means), wind speed, the presence of dirty power plants, ruggedness and latitude.

We will compare OLS results to within-country estimates, which include a country dummy θ_S. These within-country regressions compare raster cells/cities within a country to each other. The rationale is to control for any countrywide unobserved geographic or political features that may be correlated with both pollution and population (density).

Even though country fixed effects already account for a large portion of unobservables, OLS regressions may still be biased due to reverse causality or omitted variables. Economic theory and empirical evidence suggests that households would want to move to cleaner areas (Chen et al., 2022). Hence, population would be endogenous to pollution exposure. Moreover, within countries, there may be unobservable differences in policies, attitudes, and the like that are correlated with population measures and pollution. We therefore follow the urban economics literature in instrumenting population with historical population levels.[26]

The main assumptions for using historical population data have been widely discussed in the urban economic literature. First, the distribution of population and economic activity tends to be persistent over time (see for instance Davis and Weinstein, 2002). This is intuitive,

[25]Note that in the main analysis, we only look at a cross-section. We keep the time index because we run fixed-effects panel regressions as a robustness check.

[26]See, e.g. Combes et al. (2010) for a classic application and discussion of the issues in estimating agglomeration economies. See Borck and Schrauth (2021) and Carozzi and Roth (2020) for the same approach in estimating density effects on pollution in Germany and the US.

since infrastructure and buildings are durable and thus population changes are sluggish. Therefore, historical population is a good predictor of current agglomerations. Second, the exclusion restriction states that historical population should affect pollution only through its effect on current population. The argument is that, if we go back in time far enough, structural change will have led to a reshaping of local economies such that historic population levels should be exogenous to current pollution concentrations. Suppose, for instance, that a city formed close to a river in pre-industrial times in order to benefit from the trade advantage conferred by the river. It may have grown into a densely populated and highly polluted place nowadays due to industrial and traffic pollution. Then, the exogeneity assumption would be satisfied, since today's agglomeration pattern and its effect on pollution (namely, motorized traffic and industrial production) differs from the historic one (trade).

In constructing historical instruments, there is a trade-off: on the one hand, the exogeneity argument forces us to go back sufficiently long in time. On the other hand, availability constraints force us to use more recent data in order to have a sufficient number of observations.[27] With respect to exogeneity, we believe that population counts prior to 1750 have the stronger arguments compared to more recent population instruments. Before the industrial revolution, which started in the second half of the 18th century, air pollution was probably not a decisive factor for migration decisions, whereas during industrialization, there seems to be already some evidence of sorting with respect to pollution (Heblich et al., 2021).[28] Using population in 1900, however, provides us with at least twice as many observations than more historical population counts. We will use population in 1900 as our main instrument, but outcomes do not differ much using population from pre-industrialized times.[29]

We instrument population (density) as follows:

$$\ln(D_{GtS)}) = \alpha + B_2 X_{GtS} + B_3 Z_{GtS} + \eta_{GtS}, \tag{3.2}$$

[27]We also experimented with soil quality and other natural causes as instruments like Borck and Schrauth (2021) and Combes et al. (2010). However, we were not able to find strong instruments that could explain agglomerations all around the world.

[28]Heblich et al. (2021) argue that more polluted parts of cities in England were poorer as the rich sorted into less polluted areas. The authors find that those sorting patterns have persisted until today.

[29]Borck and Schrauth (2021) analyse German data and show that historically dense places have no more industrial employment than less dense ones. Since industry was a prime polluter following the industrial revolution, this lends some credibility to the exclusion restriction.

where the instrument Z is historical population. The predicted values for population, $\ln(\hat{D}_{GtS})$, are then used in the second stage instead of actual measures of population in equation (3.1).

We also present results from long-difference estimations. These regress the changes in exposure between the last and first year of observation on changes in population/density. The idea here is that there may be some unobserved differences between units that simultaneously affect population density and pollution. For instance, sorting of "green" individuals into large cities might lead to a negative correlation between density and exposure. This kind of heterogeneity is, given that it is time invariant, differenced out in the long-difference estimation.

We also estimate city level regressions, using FUAs as units of observation. This approach compares cities (between city estimates) within countries, but in addition also allows us to look at within-city effects. In within-city grid-level regressions we estimate

$$\ln(E_{GtC}) = \beta + \rho \ln(D_{GtC}) + \alpha_t + \theta_C + \epsilon_{itC}, \tag{3.3}$$

where C is the city index. We drop all the city-specific control variables since we now control for city fixed effects θ_C and only look at within-city differences.

In addition, we estimate Spatial First Differences (SFD) models, following the approach proposed by Druckenmiller and Hsiang (2018). This transfers the idea of first differences in time into physical space. In short, SFD regresses the differences in outcomes between neighbouring grid cells on the differences in controls between these same cells. Since the estimation implicitly accounts for any unobserved factors that are common to neighbouring cells (such as possibly geographic and institutional factors that may be correlated with population density and pollution), this mitigates omitted variable bias. On the downside, some interesting variation is lost by only considering variation between neighbouring cells. Further details are in Appendix 3.C.

3.3 Results

Before presenting our main outcomes using within-country OLS, we briefly compare OLS and IV coefficients first. We estimate IV regressions in order to gauge the magnitude and direction of potential biases. The reason for not using IV results as our favoured outcomes, as described above, is that we are only able to instrument a small subsample of all observed units, which moreover is primarily restricted to a developed world sample.[30]

To compare OLS and IV results, we harmonize the sample to those cities or grid cells for which the corresponding instrument is available. In all regressions we control for the full set of trade, agricultural, and weather variables as well as logged GDP, ruggedness, latitude and an indicator for the presence of a dirty power plant. We show results with population in 1900 as instrument in Table 3.2.

Table 3.2: IV and OLS regressions

	$PM_{2.5}$				NO_2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Raster-level estimations								
log(Sum of population)	0.119***	0.0902***	0.0593***	0.0556***	0.435***	0.424***	0.375***	0.387***
	(0.0219)	(0.0263)	(0.0195)	(0.0168)	(0.0266)	(0.0356)	(0.0328)	(0.0318)
N	1022	1022	1022	1022	1024	1024	1024	1024
R^2	0.588	0.587	0.853	0.853	0.637	0.637	0.793	0.793
Countries	99	99	99	99	99	99	99	99
Country FE	No	No	Yes	Yes	No	No	Yes	Yes
Est. method	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel B: FUA-level estimations								
log(Sum of population)	0.0701***	0.0720***	0.0475***	0.0480***	0.294***	0.321***	0.274***	0.286***
	(0.0167)	(0.0199)	(0.0120)	(0.0123)	(0.0202)	(0.0290)	(0.0215)	(0.0243)
N	836	836	836	836	836	836	836	836
R^2	0.656	0.656	0.884	0.884	0.662	0.661	0.801	0.801
Countries	96	96	96	96	96	96	96	96
Country FE	No	No	Yes	Yes	No	No	Yes	Yes
Est. method	OLS	IV	OLS	IV	OLS	IV	OLS	IV

Note: The table presents coefficients of OLS and IV regressions for both raster-level and city-level (FUA) results. The instrument used is population in 1900. Samples are harmonized such that OLS regressions only include those raster cells/cities for which we have the instrument available. The first two columns of each pollutant show results without country fixed effects, while the latter two of each show results including country FE. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and and indicator for a dirty power plant nearby. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$.

[30]The historical population data by Reba et al. (2016) only covers less than 0.7 percent of all inhabited grid cells in the sample when taking population from 1900 as instrument, and only slightly more than 0.3 percent of all cells when instrumenting with pre-industrial population.

Table 3.3: First stages with Population instruments

	Population 1900				Population 1650-1750			
	$PM_{2.5}$	NO_2	$PM_{2.5}$	NO_2	$PM_{2.5}$	NO_2	$PM_{2.5}$	NO_2
Population 1900	0.642***	0.624***	0.844***	0.830***				
	(0.0266)	(0.0265)	(0.0253)	(0.0251)				
Population 1650-1750					0.479***	0.464***	0.603***	0.584***
					(0.0605)	(0.0608)	(0.0826)	(0.0826)
N	1022	1024	836	836	357	358	322	322
Countries	99	99	96	96	73	73	70	70
First-stage Statistic	583.5	553.2	1114.4	1096.6	62.76	58.20	53.34	49.91

Note: The table presents first stage results for the population instruments for both raster-level and city-level first stage regressions. Columns 1/2 and 5/6 depict raster-level first stage results, while columns 3/4 and 7/8 show FUA-level coefficients. The source of historical population is Reba et al. (2016). Population 1900 is the population count in year 1900, while population 1650-1750 takes the last population count before 1750 observed in the dataset, which we refer to as pre-industrialized population. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The first insight is that the sample size is drastically reduced when considering instrumental variables. With raster cells as units of observation, only slightly more than 1000 grid cells of the roughly 200,000 total cells in the whole sample remain. Regarding FUAs, we have historic population for about 10% of all cities. The comparison of OLS and IV results shows only very small and insignificant differences as soon as we include country fixed effects (columns 3,4,7, and 8). Using population before industrialization as instrument yields similar results (see Appendix Table 3.B.1). The population instruments are exactly identified. Table 3.3 shows the first stage regressions. The F-statistic indicates that the instruments are strong.

Hence, it seems like omitted variable bias or reverse causality does not cause large biases in the estimates.[31] In the remainder of the paper we focus on within-country regressions using the whole sample available for both grid cells and cities.

3.3.1 Raster-level outcomes

We first consider raster-level outcomes. Table 3.4 compares results between simple OLS and within country regressions using the entire sample for both pollutants, NO_2 and $PM_{2.5}$. As this will become important in our city-level results, we differentiate between the sum of population within a grid cell and population density of a cell. All specifications include

[31] Note, however, that this conclusion holds only for the selective sample of cities where we have long-lagged historical population.

Table 3.4: Raster-level OLS regressions

	$PM_{2.5}$				NO_2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Sum of population)	0.102***		0.0294***		0.191***		0.157***	
	(0.00172)		(0.00149)		(0.00233)		(0.00275)	
log(pop density)		0.0943***		0.0180***		0.184***		0.146***
		(0.00169)		(0.00135)		(0.00179)		(0.00175)
log(GDP)	-0.105***	-0.111***	0.0468***	0.0440***	0.217***	0.213***	0.124***	0.121***
	(0.00541)	(0.00543)	(0.00672)	(0.00673)	(0.00526)	(0.00511)	(0.00876)	(0.00877)
Temperature	0.00521***	0.00672***	0.0249***	0.0265***	0.0657***	0.0672***	0.108***	0.110***
	(0.00107)	(0.00107)	(0.00101)	(0.00101)	(0.00118)	(0.00116)	(0.00151)	(0.00147)
Precipitation	-0.000918***	-0.000975***	-0.000286***	-0.000348***	-0.0000865	-0.000195	-0.00155***	-0.00165***
	(0.000116)	(0.000117)	(0.0000978)	(0.0000975)	(0.000135)	(0.000131)	(0.000143)	(0.000136)
Wind speed	-0.189***	-0.194***	-0.118***	-0.120***	-0.0100**	-0.0134***	0.00931**	0.00831**
	(0.00395)	(0.00398)	(0.00324)	(0.00325)	(0.00440)	(0.00414)	(0.00436)	(0.00393)
1(On coast)	-0.799***	-0.859***	-0.688***	-0.694***	-0.265***	-0.348***	-0.300***	-0.345***
	(0.0180)	(0.0183)	(0.0146)	(0.0147)	(0.0199)	(0.0159)	(0.0192)	(0.0143)
1(Close to lake)	-0.360***	-0.371***	-0.260***	-0.254***	0.198***	0.195***	0.131***	0.136***
	(0.0328)	(0.0335)	(0.0335)	(0.0336)	(0.0372)	(0.0301)	(0.0347)	(0.0273)
1(Close to river)	0.0275	0.0330*	-0.0224*	-0.0160	-0.118***	-0.113***	-0.0593***	-0.0551***
	(0.0172)	(0.0173)	(0.0128)	(0.0128)	(0.0206)	(0.0207)	(0.0180)	(0.0181)
1(Close to dirty powerplant)	0.0790***	0.105***	0.103***	0.131***	0.310***	0.328***	0.226***	0.243***
	(0.0190)	(0.0191)	(0.0139)	(0.0139)	(0.0242)	(0.0242)	(0.0214)	(0.0211)
Ruggedness	-0.00183***	-0.00119*	-0.00324***	-0.00309***	-0.0151***	-0.0143***	-0.00680***	-0.00617***
	(0.000702)	(0.000703)	(0.000658)	(0.000660)	(0.00103)	(0.00101)	(0.000965)	(0.000934)
Latitude	0.00288***	0.00261***	-0.0116***	-0.0116***	0.0596***	0.0580***	0.0686***	0.0668***
	(0.000799)	(0.000805)	(0.000957)	(0.000958)	(0.000855)	(0.000836)	(0.00125)	(0.00124)
N	175237	175235	175237	175235	178535	178507	178535	178507
R^2	0.386	0.378	0.607	0.606	0.604	0.652	0.675	0.732
Countries	161	161	161	161	160	160	160	160
Country FE	No	No	Yes	Yes	No	No	Yes	Yes

Note: The table presents coefficients of OLS regressions including all relevant controls. Apart from those that are listed in the table, the estimations additionally account for all biome indicators, and land suitability. Columns 3,4,7 and 8 additionally include country-fixed effects (Country FE). Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

our baseline covariates, i.e. weather, GDP, geographical characteristics, suitability for trade and agriculture, and a dummy for whether there is at least one highly polluting power plant within a grid cell.

Examining the results shows that the coefficients for both total population and density are reduced in magnitude when we include country fixed effects to the $PM_{2.5}$-exposure estimations, while the difference between the estimates is smaller for NO_2. In other words, the within-country effect of density on pollution exposure is much smaller than the overall effect. This suggests, for $PM_{2.5}$, that the effect of density is partly driven by certain highly polluted countries with densely populated grid cells. It might be, for instance, that some countries have policies that both limit migration to large cities and pollution.

Taking the within-country estimates, our main results show an elasticity of pollution exposure with respect to density of 0.02 in the $PM_{2.5}$ regressions and 0.15 in the NO_2 regressions.

This implies that doubling population density would result in a 1.3 percent increase in $PM_{2.5}$ exposure and a 10.7 percent increase in NO_2 exposure.

Looking at the other coefficients, we find that once we control for country fixed effects, grid cells with higher GDP are more polluted.[32] Pollution exposure rises with temperature. By contrast, precipitation and wind speed are negatively correlated with it. Exposure is also strongly affected by dirty power plants.

Using the log of pollution exposure leads to the treatment of all zero observations as missing. To avoid this, we repeat the estimation by replacing all zero values to the minimum non-zero values observed in the data.[33] Interestingly, the pollution-density elasticity for both pollutants becomes significantly higher (0.2 for $PM_{2.5}$ and 0.24 for NO_2 in the regressions with country fixed effects, see Table 3.B.2).

That the elasticity is so much lower when we exclude grid cells with zero outcome points to a significant non-linearity in the effect of (log) density on (log) exposure. To address this question from a slightly different angle, we now present non-linear regressions, where we include categorical variables for large and densely populated areas instead of continuous ones. We thus attempt to more directly measure an urban-rural gap. In order to do so, we categorize grid cells with less than 50,000 inhabitants and those with density below 100 persons per sq. km as 'rural'.[34] The results are shown in Table 3.B.3. The urban-rural gap is clearly evident: Going from rural to urban raster cells significantly increases pollution exposure. We redo the exercise with 4 instead of only 2 categories (see Table 3.B.4). There is some variation in the effects; still, we find that the effect of going from what we call 'rural' to 'urban' is larger than the effect of going from one urban category to the next (e.g. from low to moderate density). In summary, there is an urban-rural gap in pollution exposure which trumps the effect of increasing density within urban areas.

[32] Interestingly, the coefficient on GDP is negative in the $PM_{2.5}$ regression without country fixed effects. This suggests that around the world, grid cells in higher income regions tend to be less exposed to pollution, but this effect is driven by the fact that these grid cells are predominantly located in less polluted high income countries.

[33] This follows Henderson et al. (2018), who use the approach of setting observations with a zero for night lights to the minimum value in the sample in their estimates presented as main results.

[34] This follows the density category 'extremely low' as suggested by https://www.yourarticlelibrary.com/population/population-density-classification-of-the-spatial-distribution-of-population-density/19853.

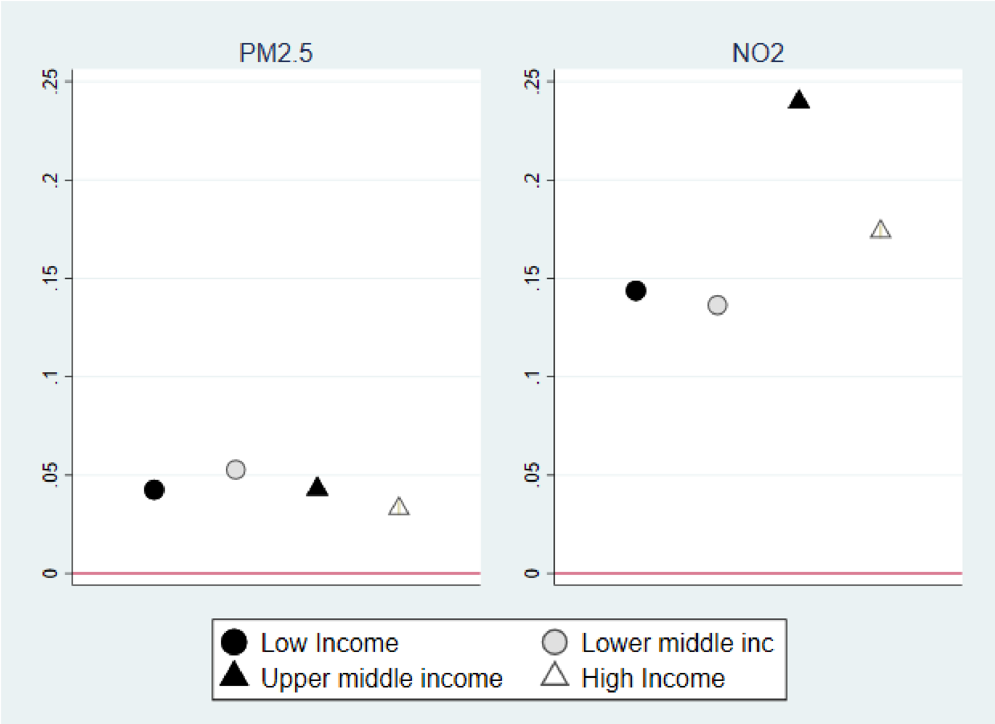
In the remainder of the raster-level results, we will mainly report within-country effects and focus on population density, which makes results more comparable to previous literature. The main estimates in Table 3.4 assume a homogeneous relationship between pollution and population in the entire world. In order to check whether this relationship changes with geography, country income, and the like, we now consider various interactions to analyse the heterogeneity of this effect.

Heterogeneity and robustness. We now turn to analysing heterogeneities in the pollution-density gradient across continents and countries at different stages of development. Figure 3.4 shows the results of running within-country regressions by country income groups and by continents, where income groups follow the World Bank classification into low, lower middle, upper middle, and high income. Population density is a significant determinant of pollution over all income groups and continents, but to a different extent. Figure 3.4a exhibits that the strongest within-country effects of density are found in low middle income countries regarding $PM_{2.5}$-exposure and in upper middle income countries for NO_2. Hence, it seems like the density effect is to some extent non-linear in income, and middle-income countries tend to have a stronger effect of density on pollution than both low and high-income countries. A potential reason is that density is not "dirty" (Carozzi and Roth, 2020) in low income countries, because there is little dirty activity such as driving and heating, whereas in high income cities, cleaner transport modes (e.g. public transportation) and residential energy use (heating and cooking with electricity or "modern fuels") may mitigate the effects of density.[35] In contrast, middle income country agglomerations may be dirtier than low income ones because there is more driving and residential energy use, but technologies for these activities are not as clean as in high income country cities.

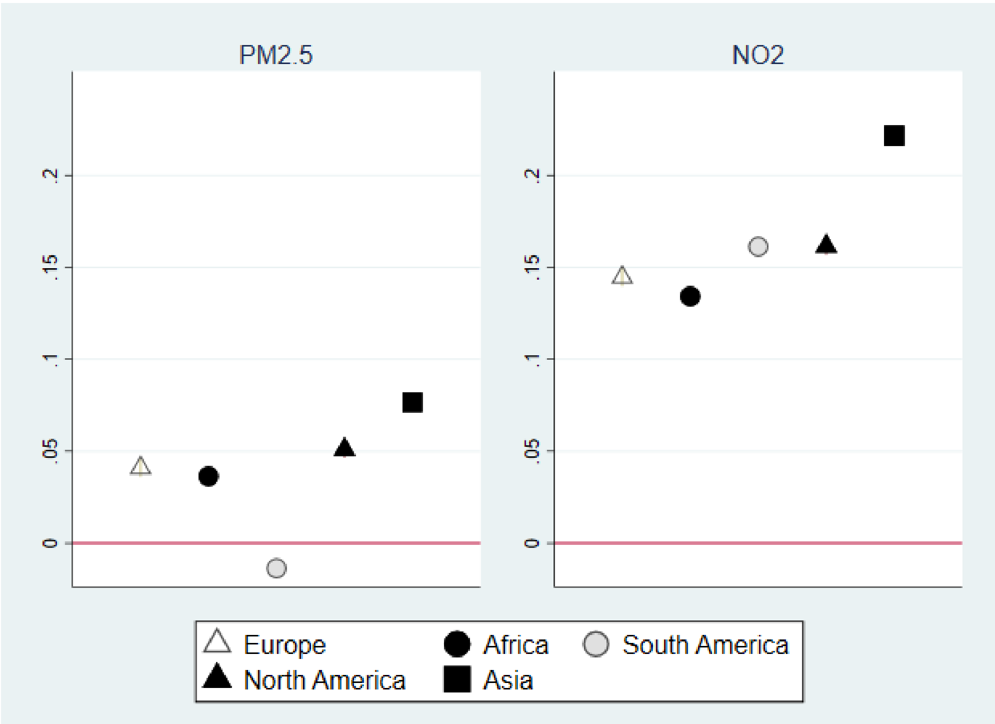
Figure 3.4b shows differences by continent. For NO_2, the density effect is smallest in Africa and largest in Asia. In Figure 3.5 and 3.6, we further show the density coefficients for each country from individual within-country regressions on world maps. The maps show some interesting ramifications. The NO_2 density elasticities in Figure 3.5 suggest that China and India – where most of the biggest and most polluted cities in the world are located – seem to have the strongest effect of density on NO_2 pollution exposure in Asia. In North America,

[35] See Borck and Mulder (2022) for a model with dirty and clean energy use as well as transport modes in developing countries.

Figure 3.4: Heterogeneity of density effect by subgroup



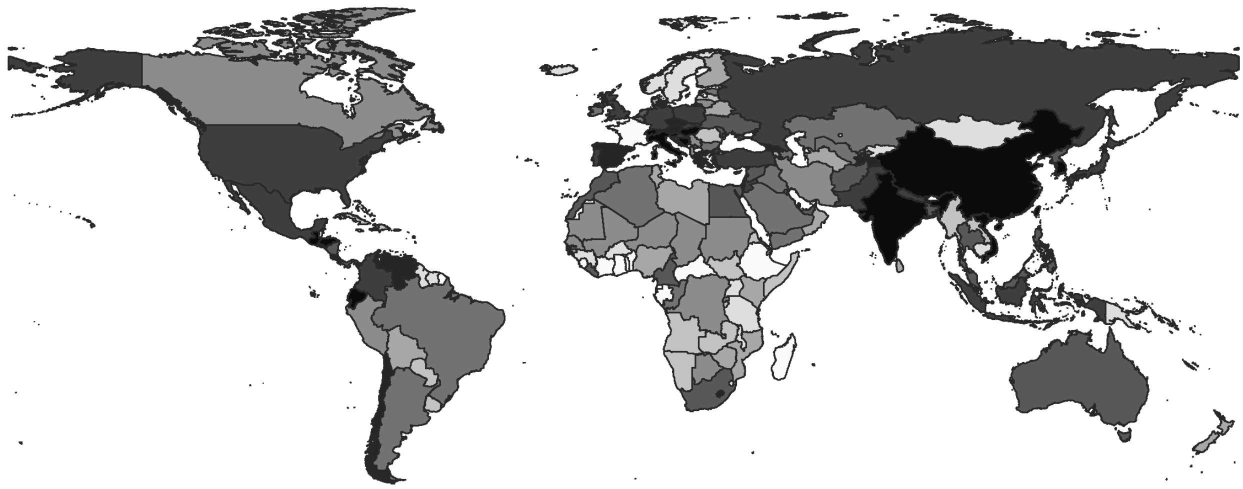
(a) By income group



(b) By continent

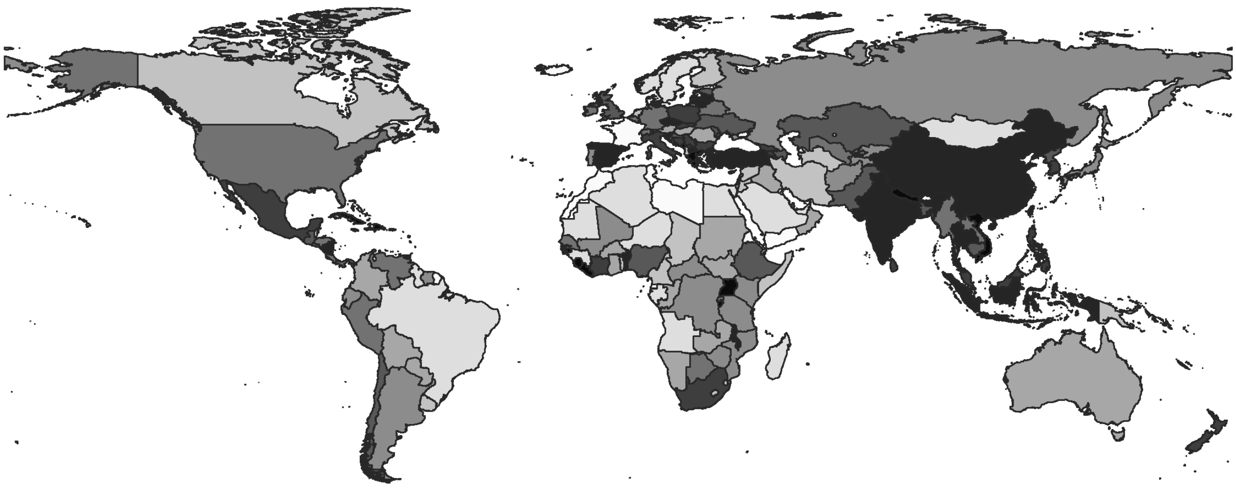
Note: The graphs show coefficients of separate regressions for each of the respective subgroups. The $y-$axes represent the coefficient sizes. All estimations include within-country fixed effects, and all the control variables from our standard estimations (compare Table 3.4). Income group definitions are taken from the World Bank.

Figure 3.5: Density coefficients for NO_2 by country



Note: The map shows a graphical representation of the effect of population density on NO_2, where regressions are performed for each country individually. Dark colours represent coefficients at the upper end of the density-pollution gradient, light colours show those at the lower end. All relevant control variables are included in the estimations.

Figure 3.6: Density coefficients for $PM_{2.5}$ by country



Note: The map shows a graphical representation of the effect of population density on $PM_{2.5}$, where regressions are performed for each country individually. Dark colours represent coefficients at the upper end of the density-pollution gradient, light colours show those at the lower end. All relevant control variables are included in the estimations.

Mexico and the US have larger effects than Canada, while within Europe countries from the south seem to have higher elasticities than countries in the north.

Figure 3.6 shows the country-specific density elasticities for $PM_{2.5}$. In Asia and Europe, most of the countries that have large density elasticities for NO_2 also have large elasticities for $PM_{2.5}$, while the density effect in the Americas seems to be smaller for $PM_{2.5}$ than for NO_2.

The LandScan data distribute population to grid cells using certain grid characteristics, but the exact algorithm is not known. While Henderson et al. (2021) provide a ground-truthing exercise, it might still be the case that the data generation biases the results. As a robustness check, we therefore run the regressions using administrative areas as units of observations[36]. In these data, where population is smoothly distributed among all grid cells within an administrative unit.[37] We present the results in Table 3.B.5. Again, they hardly differ from our previous results.[38]

Lastly, we run long-difference regressions of pollution changes between the last and first year in our sample on population or density changes in the same period. Thus, we control for any time-invariant unobserved heterogeneity between administrative units that might affect both population and pollution. For instance, it might be that within countries, population sorting leads to residents of dense cities being 'greener' on average, which would bias our estimates (downwards in this case). Analysing long differences within rasters differences out these time invariant unobserved heterogeneities. Results are shown in Table 3.B.6.[39] In the long-difference estimates, all the variables that are time-constant drop out, so the only explanatory variable left besides the population data is GDP. As the Table shows, the long-difference estimates again show a positive and significant effect of both population and density on both pollutant-exposure measures. The magnitudes are now reversed, however:

[36] We use gridded population of the world (Gridded Population of the World (version 4) (GPW)) data on administrative (Global Administrative Areas (GADM)) level, see https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11.

[37] This obviously introduces other biases. Nonetheless, it is reassuring to find that the results do not seem to be driven by biases in the computation of grid specific density.

[38] Using the gridded GPW data without aggregating it to GADM level does not change the results either (results not shown here).

[39] These estimates are also based on the GPW data on the GADM level. The reason is that the quality of LandScan data significantly improved over time and therefore comparisons over time should not be made, as stated by the data provider itself (see https://gistlandscan01.ornl.gov/frequently-asked-questions).

it seems that population changes now affect NO_2-exposure more strongly than $PM_{2.5}$. A potential explanation is that the variation in the cross section as well as over time is much lower for NO_2 than for $PM_{2.5}$. This implies that a given change in population over time affects changes in NO_2 less than those in $PM_{2.5}$.

Channels. An interesting question in interpreting the findings is what mechanisms could be responsible for the observed relationship between density and pollution exposure (see also Borck and Schrauth, 2021; Carozzi and Roth, 2020). While a complete investigation is made difficult by the scarcity of available data at a worldwide scale, we nonetheless try to shed some light on these channels here. We follow the analysis in Borck and Schrauth (2021) and leave out some sets of explanatory variables. We then compare the density coefficient with and without these variables. The direction of change of the coefficient then allows us to determine how these variables affect pollution directly and indirectly through their correlation with density.

We report the results of leaving out, one by one, different groups of our explanatory variables in Table 3.B.7. Column (1) shows the baseline regression results, col. (2) leaves out GDP, (3) the weather variables, (4) the trade variables (closeness to river, lake or coast) and (5) the agricultural variables (land suitability and biomes).

As can be seen in the table, the density coefficient rises in all columns compared to the baseline. For both pollutants, we find the largest increase when we leave out weather and agricultural suitability. This seems to imply that the density effect is driven most by the fact that dense areas are located, on average, in areas that have weather that is conducive to high pollution exposure (such as hot, dry areas with little wind). Moreover, dense areas seem, on average, to be located in places with a first nature that is advantageous to agriculture, which tends to increase pollution. Higher income and suitability for trade apparently drive the density effect to a lesser extent.

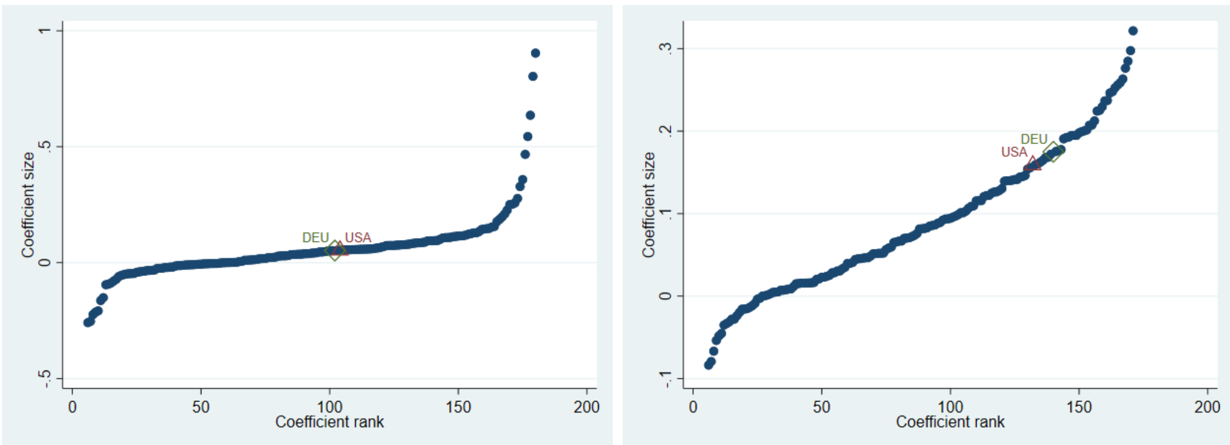
Raster-level outcomes and country characteristics. We now look at how the pollution-density relation changes with country characteristics to get additional insights. Figure 3.7 ranks all country-specific coefficients and plots them by their size.[40] All coeffi-

[40]We excluded the lowest and the highest 2.5% of the sample as outliers, thus removing 5 percent of country coefficients.

cients plotted come from regressions including control variables as presented in Table 3.4. The green hollow diamond and the red hollow triangle show the coefficients for Germany (DEU) and the US, in order to compare them to prior papers in the field (Carozzi and Roth, 2020; Borck and Schrauth, 2021). Both coefficients are somewhat smaller than the ones in those papers. Both Carozzi and Roth (2020) and Borck and Schrauth (2021) run extensive tests to more credibly estimate a causal effect within one country; however, the samples differ from the one used here. In particular, Borck and Schrauth (2021) study German counties and Carozzi and Roth (2020) US CBSA. This difference notwithstanding, we think it is reassuring to see that the magnitude of the coefficients is roughly in line with previously estimated ones from studies that are better able to address causality issues than we are.[41]

About 70% of $PM_{2.5}$ coefficients and 75% of NO_2 coefficients lie in a range between 0 and .3. About 82% of $PM_{2.5}$ coefficients and about 85% of NO_2 coefficients are positive, where about half of the cases with negative coefficients have negative ones for both pollutants. It is interesting to briefly look at the outliers, i.e. countries with the 2.5 percent highest and

Figure 3.7: Distribution of country-specific density coefficients



(a) $PM_{2.5}$ (b) NO_2

Note: The graphs show the distribution of the population density / pollution exposure gradient, when running estimations for each country in the world separately. The coefficients are obtained from regressions for each country separately, controlling for trade variables (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and an indicator for a dirty power plant nearby. The green hollow diamond is the coefficient for Germany (DEU), and the red hollow triangle the one for the United States (USA). These are highlighted in order to compare them with findings by Borck and Schrauth (2021) and Carozzi and Roth (2020) respectively. Outliers (2.5% highest and lowest coefficients) are excluded from the graphical representation.

[41]See also Ahlfeldt and Pietrostefani (2019).

lowest (negative) coefficients. In most instances, these turn out to be small island states such as Jamaica, Malta, East Timor (downward) or Bahamas, Barbados, Cap Verde (upwards).

Figure 3.B.1 in Appendix 3.B shows simple scatter plots between the country-specific density elasticity estimates of our within-country regressions and urbanization patterns (where, again, the density coefficients stem from regressions with all basic controls described above). In general, most of the correlations seem insignificant for $PM_{2.5}$, while we do find some interesting correlations for NO_2. As the figure shows, the more people live in urban areas, the stronger is the density effect on pollution, while the effect of the urbanization rate is also positive but somewhat weaker. This suggests that density is more likely to increase NO_2 pollution when many people live in cities, which underlines the non-linear effects described above. It also links the paper's results to the theory of city systems; indeed it seems like total exposure will be reduced by shifting individuals from denser to less dense regions (Borck and Tabuchi, 2019). Moreover and interestingly, the density coefficients are negatively correlated with renewable energy use (figure not shown). Intuitively, when energy use is relatively clean, packing residents densely together does not produce as much pollution as when countries rely largely on fossil fuels.

In the next subsection, we examine regression results when we explicitly consider cities defined as functional urban areas.

3.3.2 City-level outcomes

We now present results from regressions on the FUA sample. We use different variables to analyse the effect of population on pollution: (i) logged mean population density within a city polygon, and (ii) the logged total population. In addition, we can differentiate between the core city population (log(Pop urban centre)) and that of the surrounding commuting area (log(Pop commuting)). This allows us to move in the direction of considering mechanisms for the relation we study. Sprawling cities with many commuters and single family homes might have different pollution levels compared to dense cities with high-rise buildings and without much long-distance commuting.

Table 3.5 compares the effects of population density with those of total city population and additionally differentiates between population in the urban centre with the amount of people

Table 3.5: Within-country regressions for FUAs

	$PM_{2.5}$			NO_2		
	(1)	(2)	(3)	(4)	(5)	(6)
log(pop density)	-0.00435			0.00227		
	(0.0134)			(0.0102)		
log(Sum of population)		0.0791***			0.0947***	
		(0.00512)			(0.00535)	
log(Pop urban center)			0.00963			0.0402***
			(0.00790)			(0.0109)
log(Pop commuting)			0.0752***			0.112***
			(0.00725)			(0.00815)
log(GDP)	0.0547***	0.0397**	0.0362**	0.331***	0.312***	0.313***
	(0.0163)	(0.0161)	(0.0159)	(0.0242)	(0.0236)	(0.0239)
Temperature	0.0130***	0.0148***	0.0151***	0.000608	0.00355	0.00275
	(0.00278)	(0.00273)	(0.00268)	(0.00389)	(0.00380)	(0.00387)
Precipitation	-0.000334	-0.000294	-0.000274	-0.00208***	-0.00204***	-0.00205***
	(0.000208)	(0.000206)	(0.000199)	(0.000162)	(0.000161)	(0.000169)
Wind speed	-0.0697***	-0.0679***	-0.0649***	-0.00759	-0.00511	-0.00472
	(0.00799)	(0.00788)	(0.00766)	(0.00950)	(0.00935)	(0.00973)
1(On coast)	-0.384***	-0.429***	-0.437***	0.0457**	-0.00609	-0.0175
	(0.0225)	(0.0226)	(0.0221)	(0.0231)	(0.0225)	(0.0225)
1(Close to lake)	-0.265***	-0.282***	-0.259***	0.192***	0.179***	0.179***
	(0.0701)	(0.0693)	(0.0642)	(0.0591)	(0.0570)	(0.0600)
1(Close to river)	0.0560***	0.0362**	0.0363***	0.0621***	0.0371**	0.0277
	(0.0146)	(0.0145)	(0.0140)	(0.0173)	(0.0170)	(0.0178)
1(Close to dirty powerplant)	0.140***	0.0773***	0.0663***	0.502***	0.429***	0.391***
	(0.0138)	(0.0139)	(0.0136)	(0.0207)	(0.0205)	(0.0206)
Ruggedness	-0.000298***	-0.000280***	-0.000286***	-0.000511***	-0.000486***	-0.000532***
	(0.0000287)	(0.0000281)	(0.0000287)	(0.0000354)	(0.0000354)	(0.0000378)
Latitude	0.0126***	0.0134***	0.0138***	0.00918***	0.0109***	0.0103***
	(0.00178)	(0.00176)	(0.00175)	(0.00248)	(0.00245)	(0.00249)
N	8871	8871	8249	8861	8861	8219
R^2	0.688	0.699	0.726	0.786	0.795	0.790
Countries	136	136	136	134	134	134
No. of cities	8871	8871	8249	8861	8861	8219
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table presents coefficients of OLS regressions with FUAs as units of observations. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km), land suitability for agriculture, temperature, precipitation as well as ruggedness, latitude log(GDP), and country-fixed effects (Country FE). Standard errors (in parentheses) are robust. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

living in commuting zones. Interestingly, for both pollutants, the effect of population density is not significant, while total population positively affects pollution. The coefficients indicate that a 1% increase in total population increases $PM_{2.5}$ exposure by 0.08 percent and NO_2 exposure by 0.095 percent. Compared to raster-level results, the population coefficient is larger for $PM_{2.5}$ and smaller for NO_2. Hence, it seems that density *per se* does not drive higher pollution exposure; rather, the relation seems to be driven by the way that large populations are organized spatially within cities.

To elaborate on this theme, in columns (3) and (6), we distinguish between core city and commuting population. For $PM_{2.5}$, we find that the coefficient of core city population is insignificant, while that on commuting population is positive and significant. For NO_2, both are significant, but the coefficient on commuting population is about three times as large. Consequently, it seems that large cities *per se* are not more polluted than smaller ones. Rather, pollution seems to be significantly higher in cities with a large fraction of people commuting into the city from satellite cities. These findings thus shed more light on the link between density, population distribution and pollution exposure. It seems that, when we only look at cities, large and dense development does not need to be bad for the environment. This may be due to the fact that these urban features promote the use of clean public transport and energy efficient buildings, which may partly offset the increased pollution exposure stemming from a high concentration of polluting activities.[42]

Figures 3.B.2 and 3.B.3 again show the distribution of coefficients by continents and income groups. For $PM_{2.5}$, there is no clear trend by income. For NO_2, however, the population effect is strongest among upper middle and high income countries. The upper panel of Figure 3.B.3 shows that for $PM_{2.5}$, there are no pronounced differences in the density effect between continents. For NO_2, the population and density effects are lowest in Africa and Asia. The effect of commuting population is also lowest in Africa.

City level outcomes and country characteristics. Just as for the raster-level results, we repeat the exercise of relating the country-specific population-pollution coefficients to country characteristics, such as urbanization rates and income. We again present scatter

[42]See Brownstone and Golob (2009) for the effect of density on driving, Borck and Brueckner (2018) for the link between density and energy efficiency, and Carozzi and Roth (2020) on the interpretation of higher exposure due to the density of economic activity.

plots, where each point shows the country-specific coefficient of population on pollution exposure. Figure 3.B.4 shows the plots. The correlations for NO_2 again seem to be stronger than for $PM_{2.5}$. We find that agglomeration seems to affect pollution more the higher the urbanization rate and the share of population living in large agglomerations of more than 1 million people.

3.3.3 Local-level outcomes

Within city regressions. Table 3.6 shows the results of within-city regressions, that is, we compare raster cells that lie within the same FUA. Since we control for city fixed effects, all FUA-level variables are absorbed by them, so there are no additional control variables. The table shows that the population effect for $PM_{2.5}$ is about the same as in the baseline raster results. For NO_2, however, the coefficient is much lower. This may be due to the smaller variation within cities, or the fact that densely populated raster cells tend to lie in polluted cities. Beyond that, the smaller variation within cities for NO_2 is somewhat mechanically caused by the larger size of the grid cells.

Table 3.6: Within City regressions

	$PM_{2.5}$	NO_2
	(1)	(2)
log(pop density)	0.0215***	0.0425***
	(0.000351)	(0.000839)
N	66118	66127
R^2	0.995	0.971
Countries	187	175
City FE	Yes	Yes

Note: The table presents coefficients of within-FUA-effects. Each pollution cell is matched to a population cell. Since the within-city-fixed effects capture all relevant control variables we use in other tables, no additional variables are controlled for. Standard errors (in parentheses) are robust. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure 3.B.5 shows differences of results by income groups and by continents. Within cities, the pollution-population gradient steadily increases with income. Hence, different from what we observed before, higher income countries seem to have especially strong pollution in densely populated areas within cities. Looking at differences by continents, American cities exhibit the highest and African cities the lowest pollution-population gradient.

Spatial First Differences. Table 3.C.1 in the Appendix shows the results from Spatial First Differences (SFD) regressions. Again, we find that for both pollutants, the results remain positive and significant. This result is reassuring. The SFD estimate differences out any unobserved heterogeneity that is common to neighbouring cells. The result thus further lends credence to the relationship between density and pollution exposure at a local level.

However, the effect of density on pollution is much smaller when we consider differences between neighbouring cells than when we compare the entire sample, especially for NO_2. Intuitively, the variation in pollution exposure and density is much smaller between neighbouring cells than in the entire sample, which likely explains the smaller effects.

3.4 Counterfactual simulation

In order to quantify the effects of the population exposure relation by country, we present results from a counterfactual simulation in this section, where we compute the country-specific change in exposure from equalizing population across all cities in a country. We do this counterfactual with the FUA sample. We thus assume that a country's population is given by its population living in FUAs. The counterfactual then answers the question: what would be the effect on total exposure if all cities had the same population?

Let country S have M_S cities with population N_i and total population $N_S = \sum_{j=1}^{M_S} n_i$. We want to compute the effect of redistributing population equally among cities within a country. Consequently, all cities have identical counterfactual population $N_i' = N_S/M_S$. From our estimates, we predict current total exposure for city i as $\tilde{E}_i = N_i^{1+\rho}$, where ρ is the estimated exposure-population elasticity. Total exposure is then $E = \sum_j \tilde{E}_j$.[43]

Now consider a counterfactual where city i's population is changed to $N_i' = N_S/M_S$, so all cities are equally large. The counterfactual exposure in city i is $E_i' = N_i'^{1+\rho} = (N_S/M_S)^{1+\rho}$. The proportional counterfactual change in city i is $\hat{E}_i = E_i'/\tilde{E}_i = \hat{N}_i^{1+\mu}$, where $\hat{N}_i = \frac{N/M_S}{N_i}$ is the proportional change in population.

Total counterfactual exposure in the country is $E' = \sum_j E_j'$. We can then compute the

[43]Note that we do not use our usual control variables in these estimates, so their influence is subsumed in the effect of city population. Redoing the counterfactual with controls does not have a large effect on the results.

percentage change in exposure, $\Delta E = \hat{E} - 1 = E'/E - 1$.

We estimate ρ for all countries with at least 15 FUAs in our sample. We report the estimates of ρ along with total population, number of cities and the counterfactual change in exposure, ΔE, for the 10 countries with the smallest and largest change in Table 3.7 for $PM_{2.5}$ and in Table 3.8 for NO_2.

Table 3.7: Counterfactual change in exposure for $PM_{2.5}$

Country	Population (mill.)	# cities	ρ	ΔE	Rank
		A: 10 lowest ranked			
Indonesia	139.491	249	0.240	-36.494	1
Haiti	4.024	21	0.215	-31.816	2
Peru	16.998	41	0.194	-30.862	3
Nigeria	87.799	351	0.182	-30.166	4
Togo	2.953	15	0.205	-27.798	5
Vietnam	38.426	99	0.174	-27.086	6
D.R. Congo	23.493	125	0.159	-26.751	7
Malaysia	19.432	31	0.240	-22.821	8
Sudan	11.219	72	0.131	-22.080	9
Philippines	46.653	67	0.123	-21.646	10
		B: 10 highest ranked			
Colombia	28.753	85	-0.020	2.932	67
Tunisia	4.824	24	-0.035	3.586	68
Ecuador	9.361	29	-0.057	5.551	69
Yemen	6.285	23	-0.073	7.622	70
Venezuela	21.008	65	-0.129	9.436	71
Mozambique	5.143	48	-0.067	9.994	72
Uganda	5.300	21	-0.070	10.360	73
Angola	11.281	46	-0.072	13.524	74
Tanzania	10.234	40	-0.109	14.669	75
Senegal	6.291	29	-0.188	22.535	76

Note: Own calculations. The simulation is computed using the FUA–sample outcomes. ρ is the estimated within-country population elasticity of exposure. ΔE is the percentage change in exposure in the counterfactual relative to the baseline.

Since total exposure is convex in population if and only if $\rho > 0$, all countries with a positive estimated population elasticity would benefit from a reduction in total exposure induced by population smoothing. It is apparent from Tables 3.7 and 3.8 that the countries with the largest percentage drop in total exposure are those with the largest population elasticity. There are, however, some differences in the composition of countries. For $PM_{2.5}$, we see the

Table 3.8: Counterfactual change in exposure for NO_2

Country	Population (mill.)	# cities	ρ	ΔE	Rank
		A: 10 lowest ranked			
Peru	16.438	41	0.365	-53.175	1
China	567.218	1539	0.408	-46.429	2
Thailand	16.893	37	0.292	-45.821	3
Japan	105.882	88	0.298	-44.340	4
Guatemala	5.372	34	0.302	-43.668	5
Indonesia	133.417	249	0.269	-40.361	6
Iran	50.591	165	0.277	-38.800	7
Argentina	27.907	64	0.230	-37.419	8
Chile	12.235	31	0.275	-36.070	9
Australia	15.575	20	0.419	-34.052	10
		B: 10 highest ranked			
Guinea	2.659	17	-0.024	3.059	66
Benin	2.934	21	-0.040	4.122	67
D.R. Congo	18.242	125	-0.023	4.476	68
Togo	1.924	15	-0.045	5.078	69
Burkina Faso	2.689	32	-0.031	5.558	70
Uganda	4.291	21	-0.042	6.011	71
North Korea	10.764	80	-0.073	6.400	72
Tanzania	7.210	40	-0.053	6.429	73
Cote d'Ivoire	6.953	35	-0.076	12.491	74
Ghana	8.090	51	-0.137	26.080	75

Note: Own calculations. The simulation is computed using the FUA–sample outcomes. ρ is the estimated within-country population elasticity of exposure. ΔE is the percentage change in exposure in the counterfactual relative to the baseline.

largest drops in total exposure in countries in East Asia/Pacific and Sub-Saharan Africa, plus two in Latin America/Caribbean. The countries with the largest increase in exposure (where the population elasticity is negative) tend to be lower income countries in Latin America/Caribbean and Sub-Saharan Africa, plus two in Middle East/North Africa.

For NO_2, the countries with the largest percentage drop in total exposure tend to be in East Asia/Pacific and Latin America/Caribbean, while the ones with increases in exposure are almost all in Sub-Saharan Africa, with the exception of North Korea.

3.5 Conclusion

This paper has studied the effect of population and population density on pollution exposure using worldwide gridded data. We find that population density increases exposure. Using city-level data, we find that population size, rather than density, increases exposure. Further, the reason seems not to be a large core city population, but rather a large population commuting into the core city. Lastly, we find positive but smaller effects of density on pollution exposure at more local levels.

We also document heterogeneities of the density effects across countries. Using the entire rasterized data as observational units, the influence of population seems largest in Asia and in middle-income countries. In the FUA sample, population affects pollution most in upper middle and high income countries as well as in Europe and North America.

Finally, we study how reallocating population among cities within countries affects exposure. For most countries, exposure would fall if population were equalized across cities, since total city exposure is convex in city population. This allows us to connect the paper to the literature on optimal city size (Borck and Tabuchi, 2019). Using country specific exposure-population elasticities, we could in principle study how the distribution of optimal city size is determined by the trade-off between agglomeration benefits and costs, stemming from the increase in exposure.

Some avenues for future research suggest themselves. First, it would obviously be of interest to study how differences in the population-pollution elasticities between countries are shaped by institutional determinants, urban structure, and other factors. Second, although we have tried to come close to estimating causal effects, we think it would be fruitful to merge this more descriptive global evidence with the more causal national evidence as in Borck and Schrauth (2021) and Carozzi and Roth (2020). As more data becomes available for more countries and longer time periods, more robust evidence on the agglomeration costs of pollution will certainly be forthcoming.

Appendices

3.A Ground-level pollution data

Pollution data as measured by satellites comes in the form of Aerosol Optical Depth (AOD), which deduct pollution concentration from the intensity of light that is reflected into space. However, this measure is not the same as ground-level pollution concentration, which reflects actual pollution exposure. As we are interested in actual exposure, we use ground-level pollution as described by Hammer et al. (2020) for $PM_{2.5}$. They deduct their estimates by using a GEOS-chem chemical transport model, relating measures from ground-level monitors with satellite measured AOD. The $PM_{2.5}$-data comes in two versions: one that applies Geographically Weighted Regression (GWR) and one that does not. If GWR is applied, then the correlation of ground-monitor measurements with AOD measurements is even higher than without the GWR application (with a slope of up to .97 using GWR compared to one of 0.90 without GWR). We stick to the data product, which is closest to ground-based monitors, which are the GWR products.

3.B Additional tables and figures

Table 3.B.1: IV and OLS regressions with population before 1750 as IV

	$PM_{2.5}$				NO_2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Raster-level estimations								
log(Sum of population)	0.172***	0.251***	0.0822***	0.111***	0.380***	0.498***	0.304***	0.371***
	(0.0262)	(0.0600)	(0.0226)	(0.0356)	(0.0299)	(0.0891)	(0.0378)	(0.0799)
N	357	357	357	357	358	358	358	358
R^2	0.580	0.445	0.879	0.837	0.765	0.489	0.890	0.815
Countries	73	73	73	73	73	73	73	73
Country FE	No	No	Yes	Yes	No	No	Yes	Yes
Est. method	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel B: FUA-level estimations								
log(Sum of population)	0.116***	0.207***	0.0816***	0.0339	0.300***	0.333***	0.298***	0.258***
	(0.0240)	(0.0530)	(0.0234)	(0.0333)	(0.0281)	(0.0607)	(0.0365)	(0.0564)
N	322	322	322	322	322	322	322	322
R^2	0.598	0.578	0.876	0.873	0.768	0.767	0.887	0.886
Countries	70	70	70	70	70	70	70	70
Country FE	No	No	Yes	Yes	No	No	Yes	Yes
Est. method	OLS	IV	OLS	IV	OLS	IV	OLS	IV

Note: The table presents coefficients of OLS and IV regressions for both raster-level and city-level (FUA) results. The instrument used is population between 1650 and 1750. Samples are harmonized such that OLS regressions only include those raster cells/cities for which we have the instrument available. The first two columns of each pollutant show results without country fixed effects, while the latter two of each show results including country FE. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and and indicator for a dirty power plant nearby. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.B.2: Raster-level OLS regressions with non-zero pollution

	$PM_{2.5}$				NO_2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Sum of population)	0.278***		0.242***		0.253***		0.258***	
	(0.00348)		(0.00401)		(0.00698)		(0.00730)	
log(pop density)		0.251***		0.206***		0.238***		0.239***
		(0.00345)		(0.00390)		(0.00672)		(0.00691)
log(GDP)	-0.0581***	-0.0767***	0.0340**	0.0228	0.322***	0.312***	0.229***	0.220***
	(0.00767)	(0.00773)	(0.0148)	(0.0149)	(0.0179)	(0.0178)	(0.0382)	(0.0382)
Temperature	-0.00297*	0.00199	0.0220***	0.0279***	0.0873***	0.0904***	0.173***	0.177***
	(0.00179)	(0.00180)	(0.00233)	(0.00233)	(0.00525)	(0.00524)	(0.00727)	(0.00725)
Precipitation	0.000235	0.0000804	0.000832***	0.000623***	-0.00748***	-0.00765***	-0.00747***	-0.00767***
	(0.000167)	(0.000168)	(0.000186)	(0.000186)	(0.000678)	(0.000678)	(0.000755)	(0.000755)
Wind speed	-0.180***	-0.196***	-0.114***	-0.130***	-0.259***	-0.268***	-0.273***	-0.282***
	(0.00637)	(0.00648)	(0.00680)	(0.00689)	(0.0200)	(0.0199)	(0.0213)	(0.0213)
1(On coast)	-1.425***	-1.593***	-1.329***	-1.461***	-0.998***	-1.115***	-0.756***	-0.867***
	(0.0335)	(0.0347)	(0.0342)	(0.0354)	(0.0932)	(0.0932)	(0.0890)	(0.0885)
1(Close to lake)	-0.934***	-0.969***	-0.869***	-0.888***	0.189*	0.159	0.338***	0.308***
	(0.0836)	(0.0859)	(0.0819)	(0.0839)	(0.0984)	(0.0993)	(0.105)	(0.105)
1(Close to river)	-0.0731**	-0.0562	-0.103***	-0.0845**	-0.00942	0.00141	-0.0592	-0.0493
	(0.0367)	(0.0369)	(0.0364)	(0.0365)	(0.0444)	(0.0444)	(0.0427)	(0.0428)
1(Close to dirty powerplant)	-0.227***	-0.138***	-0.259***	-0.163***	0.0922	0.138**	-0.0389	0.00367
	(0.0258)	(0.0262)	(0.0252)	(0.0253)	(0.0668)	(0.0668)	(0.0582)	(0.0582)
Ruggedness	-0.00243*	-0.000409	0.00137	0.00238*	-0.158***	-0.157***	-0.131***	-0.130***
	(0.00128)	(0.00128)	(0.00138)	(0.00138)	(0.00820)	(0.00820)	(0.00781)	(0.00780)
Latitude	-0.000609	-0.000914	-0.0145***	-0.0157***	0.0551***	0.0537***	0.0918***	0.0896***
	(0.00115)	(0.00116)	(0.00190)	(0.00190)	(0.00353)	(0.00353)	(0.00641)	(0.00641)
N	180489	180453	180489	180453	181663	181629	181663	181629
R^2	0.276	0.260	0.310	0.296	0.196	0.196	0.271	0.271
Countries	161	161	161	161	161	161	161	161
Country FE	No	No	Yes	Yes	No	No	Yes	Yes

Note: The table presents coefficients of OLS regressions including all relevant controls. Apart from those that are listed in the table, the estimations additionally account for all biome indicators, and land suitability. Columns 3,4,7 and 8 additionally include country-fixed effects (Country FE). All pollution observations that are zero are replaced by the minimum pollution value in the sample. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.B.3: Raster-level OLS regressions - urban-rural gradient

	$PM_{2.5}$				NO_2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Urban population	0.710***		0.294***		0.988***		0.761***	
	(0.0109)		(0.00792)		(0.0151)		(0.0153)	
Urban pop density		0.717***		0.265***		1.012***		0.774***
		(0.0117)		(0.00855)		(0.0159)		(0.0156)
N	175390	175388	175390	175388	179484	179466	179484	179466
R^2	0.370	0.366	0.605	0.604	0.211	0.215	0.261	0.266
Countries	177	177	177	177	171	171	171	171
Country FE	No	No	Yes	Yes	No	No	Yes	Yes

Note: The table presents coefficients of estimations using population (density) categories instead of continuous population density or the sum of population. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and and indicator for a dirty power plant nearby. Columns 3,4,7 and 8 include country-fixed effects (Country FE). Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.B.4: Raster-level OLS regressions with population categories

	$PM_{2.5}$				NO_2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
50k to 100k	0.473***		0.211***		0.760***		0.560***	
	(0.0109)		(0.00778)		(0.0144)		(0.0136)	
100k to 500k	0.820***		0.334***		1.083***		0.861***	
	(0.0132)		(0.00972)		(0.0182)		(0.0180)	
500k to 1m	1.144***		0.533***		1.367***		1.219***	
	(0.0252)		(0.0178)		(0.0378)		(0.0305)	
>1m	1.123***		0.569***		1.767***		1.637***	
	(0.0364)		(0.0269)		(0.0614)		(0.0521)	
Low Density		0.553***		0.196***		0.822***		0.608***
		(0.0116)		(0.00853)		(0.0150)		(0.0142)
Moderate Density		0.792***		0.278***		1.084***		0.851***
		(0.0166)		(0.0118)		(0.0215)		(0.0198)
High Density		1.024***		0.436***		1.340***		1.065***
		(0.0216)		(0.0160)		(0.0298)		(0.0244)
Very High Density		1.064***		0.496***		1.535***		1.382***
		(0.0279)		(0.0217)		(0.0425)		(0.0366)
N	175390	175388	175390	175388	179484	179466	179484	179466
R^2	0.376	0.371	0.607	0.605	0.212	0.216	0.262	0.268
Countries	177	177	177	177	171	171	171	171
Country FE	No	No	Yes	Yes	No	No	Yes	Yes

Note: The table presents coefficients of estimations using population (density) categories instead of continuous population density or the sum of population. The population categories are: 0 to 50.000 (50k) [base category], 50k-100k, 100k-500k, 500k-1million or more than 1 million inhabitants. The density categories are: very low density [base category], low density, moderate density, and very high density. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and and indicator for a dirty power plant nearby. Columns 3,4,7 and 8 include country-fixed effects (Country FE). Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.B.5: GADM-level OLS regressions

	$PM_{2.5}$				NO_2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Sum of population)	0.108***		0.0289***		0.143***		0.114***	
	(0.00203)		(0.00162)		(0.00257)		(0.00323)	
log(pop density)		0.112***		0.0569***		0.245***		0.207***
		(0.00172)		(0.00144)		(0.00194)		(0.00214)
log(GDP)	-0.0276***	-0.0374***	0.101***	0.0933***	0.363***	0.357***	0.228***	0.202***
	(0.00403)	(0.00391)	(0.00522)	(0.00516)	(0.00463)	(0.00405)	(0.00679)	(0.00590)
Temperature	-0.00859***	-0.0120***	-0.0163***	-0.0163***	-0.00983***	-0.0232***	-0.0115***	-0.0113***
	(0.000839)	(0.000803)	(0.000920)	(0.000882)	(0.00119)	(0.000982)	(0.00150)	(0.00129)
Precipitation	-0.000105***	-0.000118***	0.0000690***	0.0000710***	-0.000141***	-0.000152***	-0.000141***	-0.000132***
	(0.00000828)	(0.00000828)	(0.00000847)	(0.00000819)	(0.00000881)	(0.00000760)	(0.0000103)	(0.00000893)
Wind Speed	-0.181***	-0.172***	-0.0778***	-0.0683***	0.0270***	0.0588***	-0.0175***	0.0167***
	(0.00314)	(0.00311)	(0.00311)	(0.00305)	(0.00398)	(0.00344)	(0.00388)	(0.00348)
1(On coast)	-0.380***	-0.385***	-0.178***	-0.192***	-0.164***	-0.222***	-0.128***	-0.175***
	(0.0102)	(0.0101)	(0.00760)	(0.00749)	(0.0116)	(0.0101)	(0.0104)	(0.00906)
1(Close to lake)	-0.0936***	-0.00463	0.0964***	0.0928***	-0.327***	-0.223***	0.0621**	0.0536**
	(0.0306)	(0.0297)	(0.0196)	(0.0190)	(0.0336)	(0.0279)	(0.0306)	(0.0268)
1(Close to river)	-0.00559	0.110***	-0.0176**	0.00774	-0.136***	0.0472***	-0.134***	-0.0376***
	(0.0107)	(0.0106)	(0.00690)	(0.00676)	(0.0136)	(0.0110)	(0.0115)	(0.00985)
1(Dirty power plant in district)	0.0452***	0.188***	0.103***	0.0881***	0.00618	0.145***	0.129***	0.0837***
	(0.0127)	(0.0126)	(0.00956)	(0.00928)	(0.0162)	(0.0138)	(0.0147)	(0.0126)
N	44933	44933	44933	44933	44560	44560	44560	44560
R^2	0.397	0.407	0.760	0.769	0.637	0.722	0.775	0.823
Countries	154	154	154	154	152	152	152	152
Country FE	No	No	Yes	Yes	No	No	Yes	Yes

Note: The table presents coefficients of estimations using population data by global administrative areas (GADM) rather than LandScan data. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km), land suitability for agriculture, temperature, precipitation as well as ruggedness, latitude, and log(GDP). Columns 3,4,7 and 8 include country-fixed effects (Country FE). Standard errors (in parentheses) are robust. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.B.6: GADM-level long differences regressions

	$PM_{2.5}$		NO_2	
	(1)	(2)	(3)	(4)
log(Sum of population)		0.146***		0.0591***
		(0.0115)		(0.00759)
log(pop density)	0.147***		0.0597***	
	(0.0115)		(0.00763)	
N	90396	90396	90010	90010
R^2	0.136	0.136	0.087	0.087
Countries	159	159	157	157

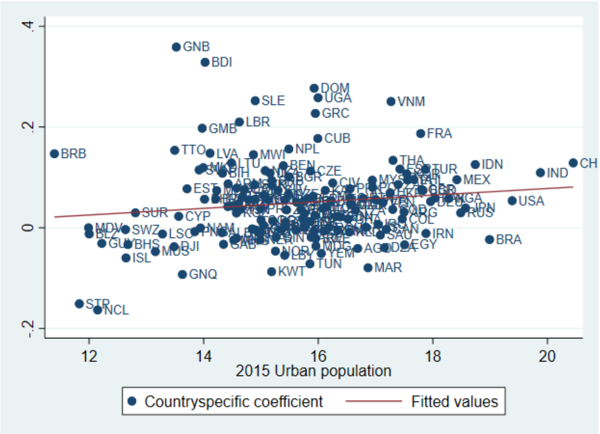
Note: The table presents coefficients of long differences estimations on sub-national administrative level comparable to NUTS-3 regions. This corresponds to districts in Germany. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km), land suitability for agriculture, temperature, precipitation as well as ruggedness, latitude, and log(GDP). Columns 3,4,7 and 8 include country-fixed effects (Country FE). Standard errors (in parentheses) are robust. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.B.7: Raster-level OLS regressions with varying sets of control variables

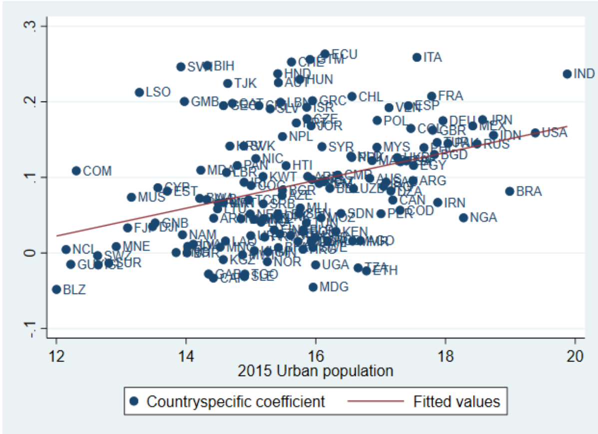
	(1)	(2)	(3)	(4)	(5)
Panel A: $PM_{2.5}$ ***Estimates***					
log(Sum of population)	0.0294***	0.0357***	0.0481***	0.0326***	0.0517***
	(0.00149)	(0.00144)	(0.00148)	(0.00147)	(0.00134)
N	175237	175274	175237	175237	175237
R^2	0.607	0.605	0.589	0.582	0.592
Countries	161	162	161	161	161
Country FE	Yes	Yes	Yes	Yes	Yes
Panel B: NO_2 ***Estimates***					
log(Sum of population)	0.157***	0.172***	0.209***	0.173***	0.222***
	(0.00275)	(0.00262)	(0.00263)	(0.00261)	(0.00244)
N	178535	178579	178535	178535	178535
R^2	0.675	0.666	0.614	0.665	0.606
Countries	160	161	160	160	160
Country FE	Yes	Yes	Yes	Yes	Yes

Note: The table presents coefficients of OLS regressions including different sets of control variables. The first column of each panel is the reference outcome as presented in Table 3.4, that includes all control variables. Column 2 excludes GDP, column 3 excludes weather variables (temperature, wind, precipitation), and column 4 does not take into account trade-specific controls (distance to lake, river, coast). Finally, column 5 excludes the biome indicators. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

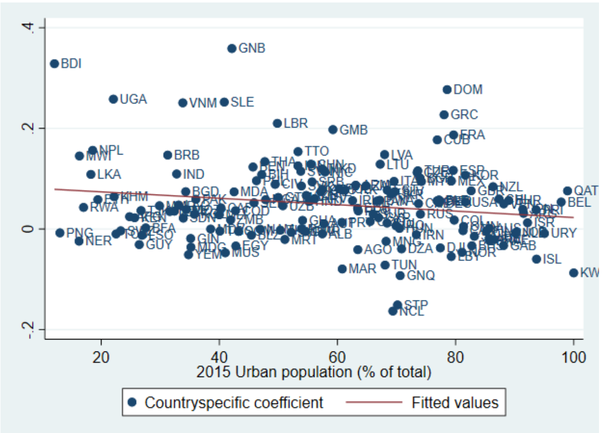
Figure 3.B.1: Country-specific density effects and urbanization



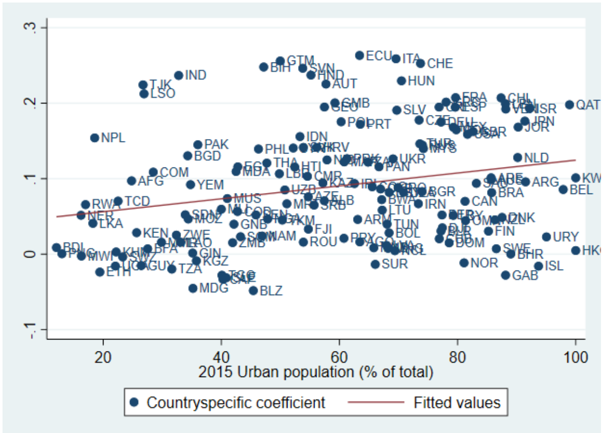
(a) Urban Population total ($PM_{2.5}$)



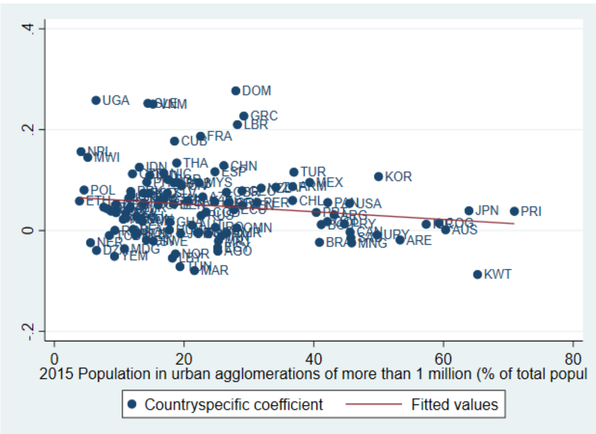
(b) Urban Population total (NO_2)



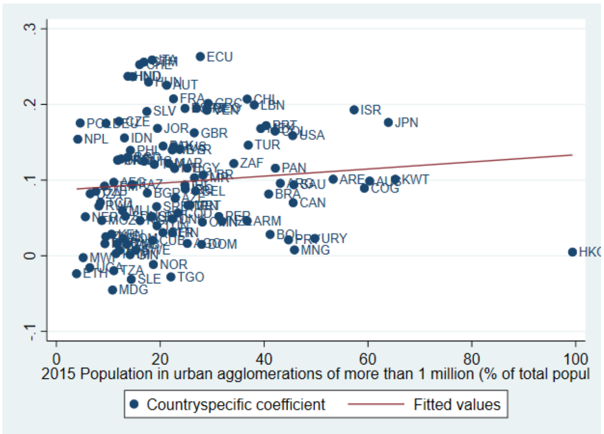
(c) Urbanization rate ($PM_{2.5}$)



(d) Urbanization rate (NO_2)



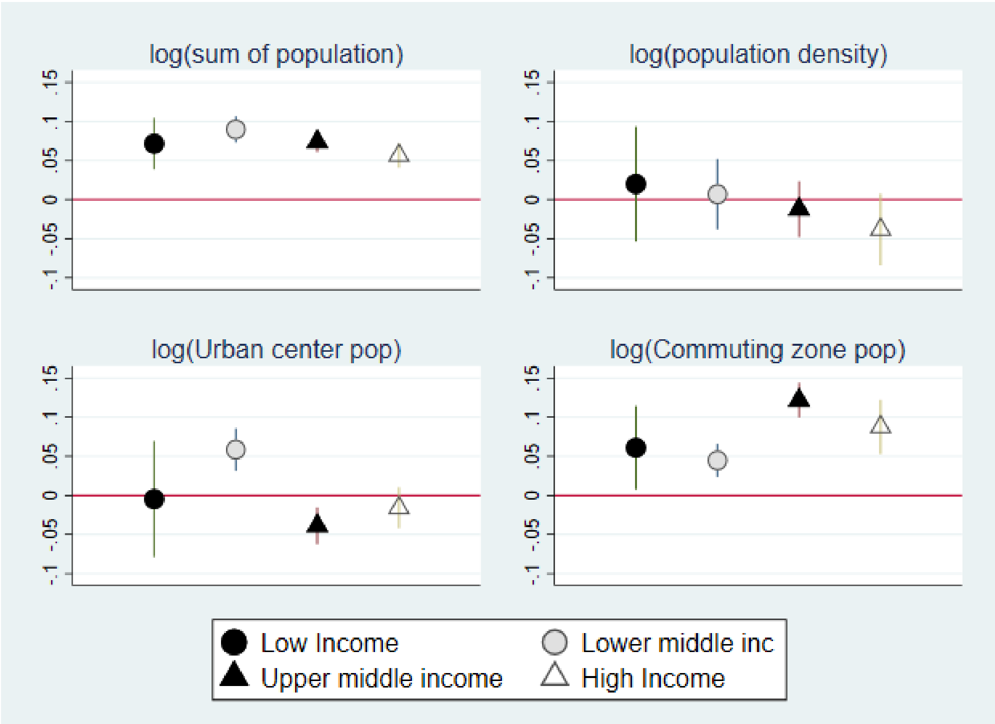
(e) Large Agglomerations ($PM_{2.5}$)



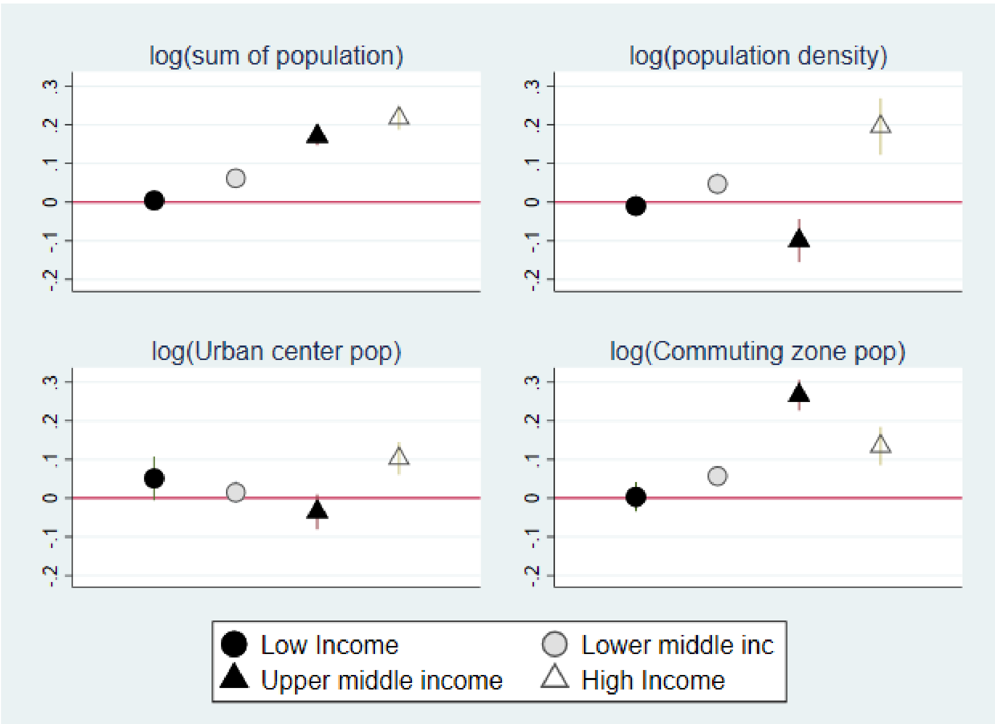
(f) Large Agglomerations (NO_2)

Note: Scatter plots of the country-specific population density effect on pollution exposure correlated with different World Bank indicators as specified by each subtitle. The coefficients are obtained from regressions for each country separately, controlling for trade variables (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and an indicator for a dirty power plant nearby. Outliers (2.5% highest and lowest coefficients) are excluded from the graphical representation.

Figure 3.B.2: Population/density effect by income



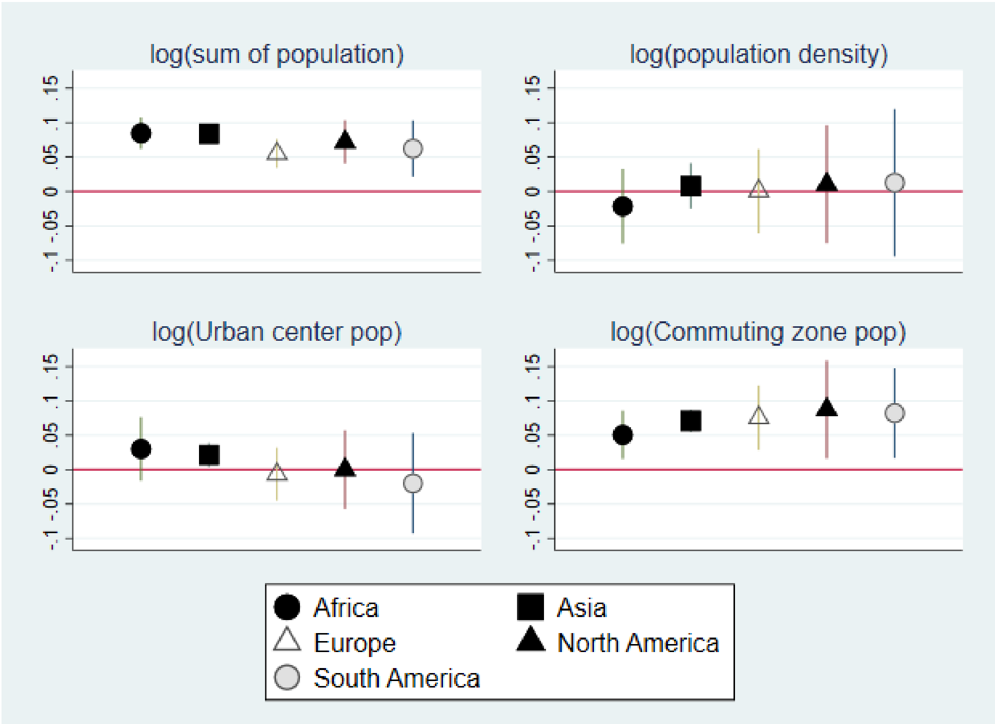
(a) $PM_{2.5}$



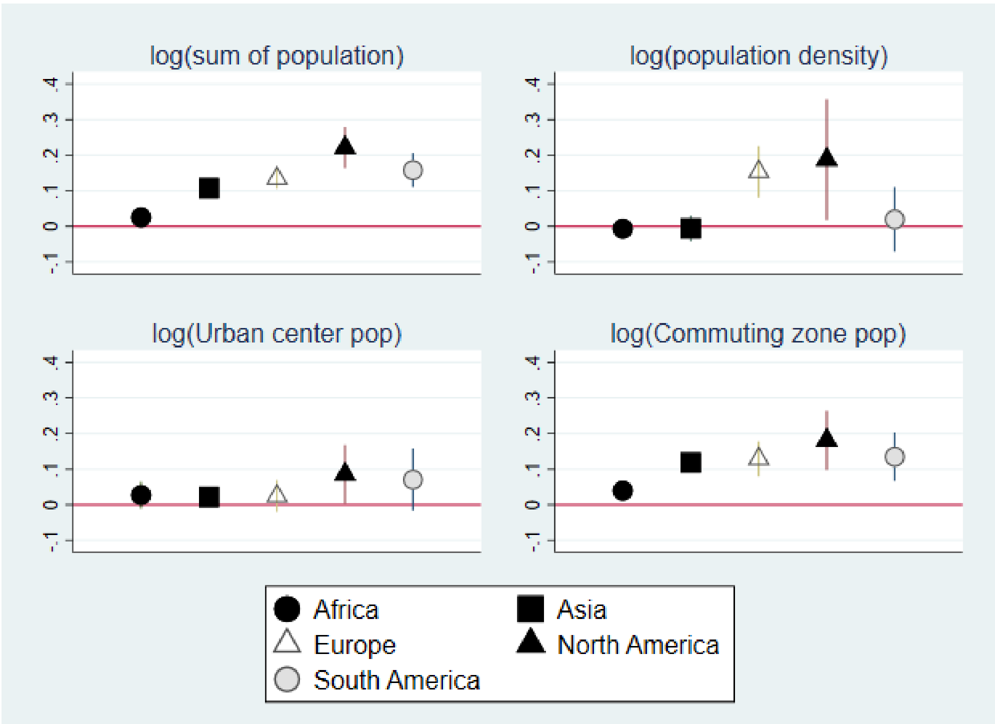
(b) NO_2

Note: Coefficients of different population measures on pollution exposure by subgroups. The coefficients are obtained from regressions for each subgroup separately, controlling for trade variables (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and an indicator for a dirty power plant nearby.

Figure 3.B.3: Population/density effect by continent



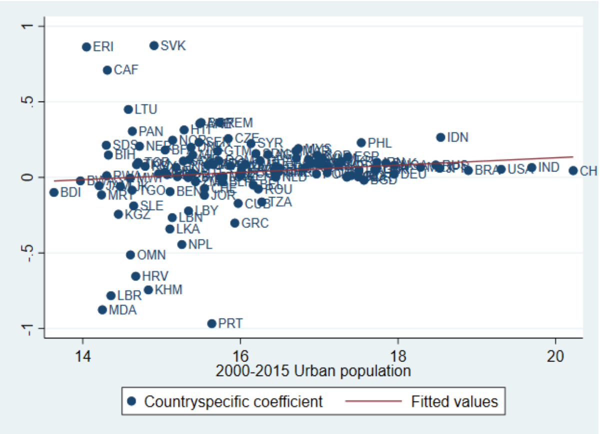
(a) $PM_{2.5}$



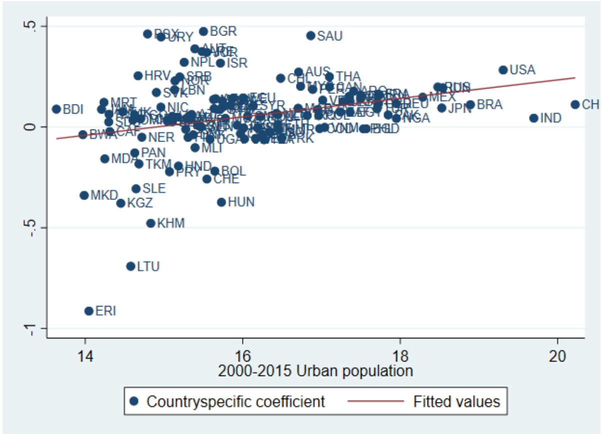
(b) NO_2

Note: Coefficients of different population measures on pollution exposure by subgroups. The coefficients are obtained from regressions for each subgroup separately, controlling for trade variables (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and an indicator for a dirty power plant nearby.

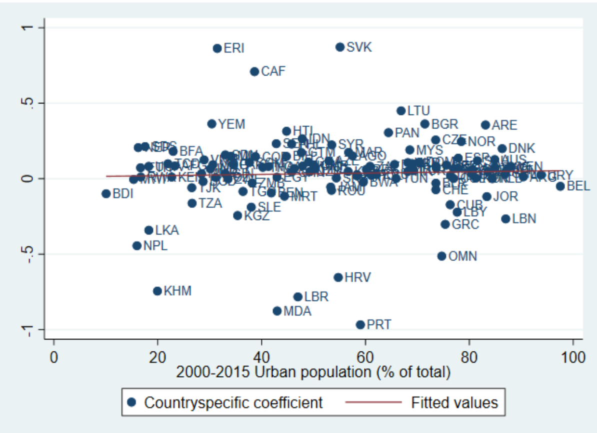
Figure 3.B.4: Country-specific population effects and urbanization



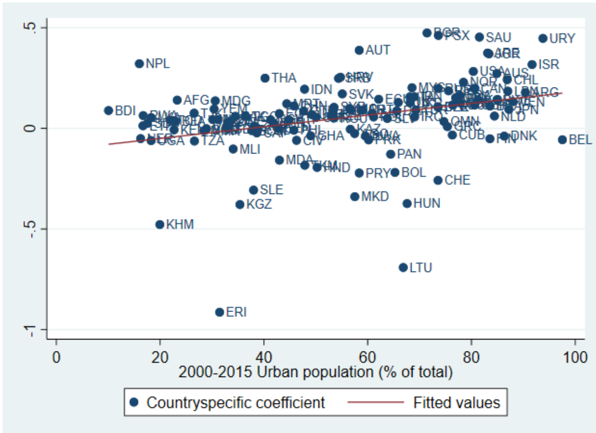
(a) Urban Population total ($PM_{2.5}$)



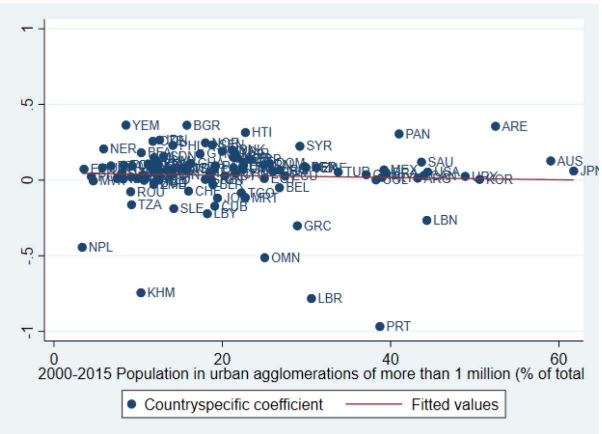
(b) Urban Population total (NO_2)



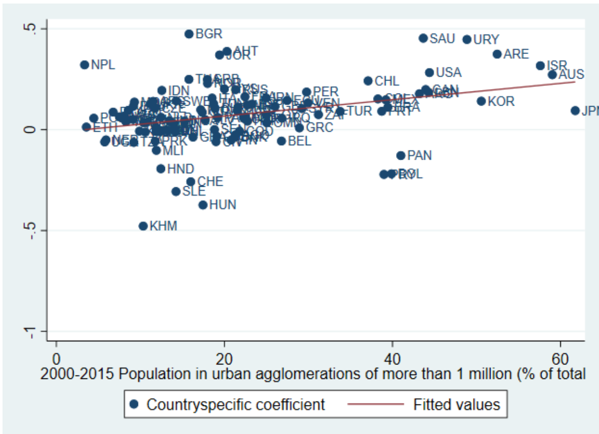
(c) Urbanization rate ($PM_{2.5}$)



(d) Urbanization rate (NO_2)



(e) Large Agglomerations ($PM_{2.5}$)

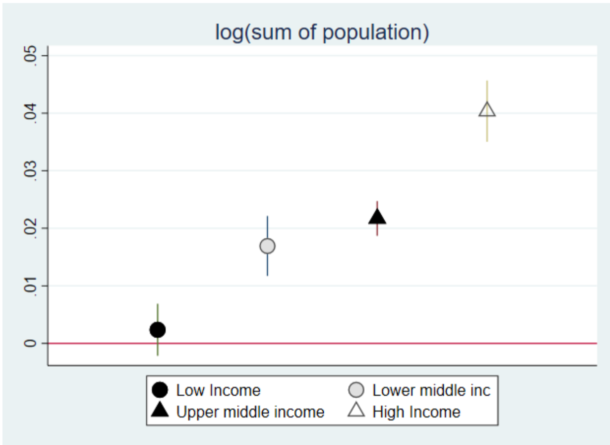


(f) Large Agglomerations (NO_2)

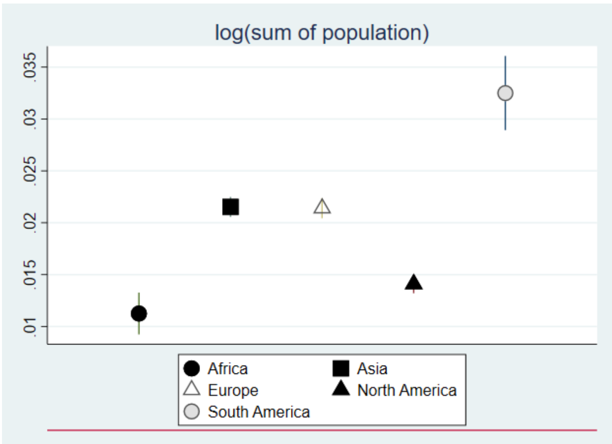
Note: Scatter plots of the country-specific population density effect on pollution exposure correlated with different World Bank indicators as specified by each subtitle. The coefficients are obtained from regressions for each country separately, controlling for trade variables (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and an indicator for a dirty power plant nearby. Outliers (2.5% highest and lowest coefficients) are excluded from the graphical representation.

Figure 3.B.5: Population effect within city by sub-sample

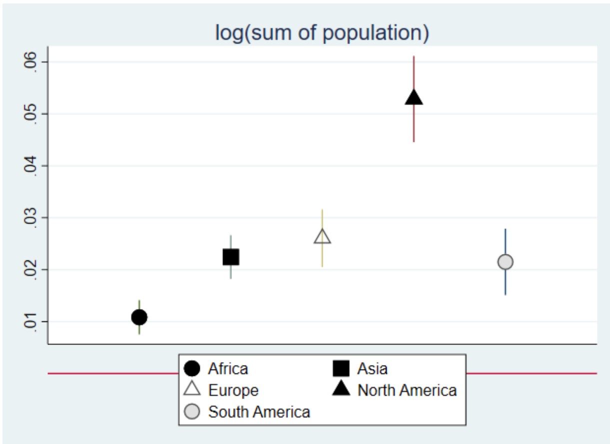
(a) $PM_{2.5}$ by income



(b) NO_2 by income



(c) $PM_{2.5}$ by continent



(d) $NO_2$2 by continent

Note: Coefficients of population on pollution exposure by subgroups within cities. The coefficients are obtained from regressions for each subgroup separately. The results stem from within-city estimations, comparing pollution exposure of small grid cells within cities.

3.C Spatial First Differences

The SFD estimation considers only differences between neighbouring cells. The crucial assumption to hold in this context is that

$$E[y_i|\mathbf{x}_{i-1}] = E[y_{i-1}|\mathbf{x}_{i-1}]\forall \ \{i, i-1\},$$

which states that pollution y in adjacent neighbouring grid cells i and $i-1$ would be equal if they had the same population density x_{i-1}. The authors refer to this as the Local Conditional Independence Assumption (LCIA).

We estimate the following equation:

$$\Delta y_i = \Delta x_i \beta_{SFD} + \theta_c + \Delta \epsilon_i,$$

where Δ is the difference operator. To make sure that the LCIA assumption holds, we add country fixed effects θ_c to our regressions. In a nutshell, we then compare neighbouring cells within countries.

Since there are neighbouring cells in North-South and in East-West direction, we will always run two separate regressions: one for horizontal neighbours (East-West direction) and one for vertical neighbours (North-South). We consider the SFD results as a lower bound on the density effects, since it does not allow to account for many potentially important differences between non-neighbouring cells.

Table 3.C.1: Spatial first differences

	NO_2				$PM_{2.5}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	WE	NS	WE	NS	WE	NS	WE	NS
log(pop density)	0.0313***	0.0613***	0.0313***	0.0368***	0.0266***	0.00986***	0.0203***	0.00596***
	(0.00149)	(0.00225)	(0.00127)	(0.00137)	(0.00179)	(0.00188)	(0.00169)	(0.00182)
log(GDP)			0.0409***	0.277***			-0.00301	0.0829***
			(0.0149)	(0.0180)			(0.0138)	(0.0143)
Temperature			0.0453***	0.0741***			-0.00588***	-0.00111
			(0.00308)	(0.00323)			(0.00186)	(0.00205)
Precipitation			-0.000344*	0.00159***			0.000798***	-0.000262*
			(0.000197)	(0.000265)			(0.000144)	(0.000159)
Wind Speed			-0.0257***	-0.0341***			-0.127***	-0.105***
			(0.00416)	(0.00389)			(0.00403)	(0.00405)
1(Close to river)			0.0199*	-0.00441			-0.00935	0.00519
			(0.0121)	(0.0138)			(0.0130)	(0.0134)
1(Close to lake)			0.0696***	0.0348			-0.107***	-0.0808**
			(0.0233)	(0.0315)			(0.0259)	(0.0368)
1(On coast)			0.00119	-0.0371***			-0.313***	-0.245***
			(0.00976)	(0.00971)			(0.0155)	(0.0156)
Ruggedness			-0.00428***	-0.00130***			0.000767*	-0.000117
			(0.000530)	(0.000459)			(0.000443)	(0.000425)
N	162268	163530	159674	160907	158506	159646	156380	157456
R^2	0.019	0.060	0.177	0.204	0.009	0.004	0.063	0.025
No. of grids	162268	163530	159674	160907	158506	159646	156380	157456

Note: The table presents coefficients of Spatial first differences regressions including different sets of control variables. For all control variables, coefficients are provided in the table. WE stands for West-East meaning that neighbouring raster cells in the West to East direction are subject to within-analyses. NS stands for North-South such that vertical neighbours are compared to each other. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Chapter 4

Ticket to paradise? The effect of a public transport subsidy on air quality.[1]

Abstract

This paper provides novel evidence on the impact of public transport subsidies on air pollution. We obtain causal estimates by leveraging a unique policy intervention in Germany that temporarily reduced nationwide prices for regional public transport to a monthly flat rate price of 9 Euros. Using DiD estimation strategies on air pollutant data, we show that this intervention causally reduced a benchmark air pollution index by more than eight percent. Our results illustrate that public transport subsidies - especially in the context of spatially constrained cities - offer a viable alternative for policymakers and city planers to improve air quality, which has been shown to crucially affect health outcomes.

[1]Co-authored with Niklas Gohl.

4.1 Introduction

The UN's Sustainable Development Goals emphasize the importance of air quality within cities and explicitly list the improvement of urban air quality as a key measure to make agglomerations safer, healthier, more resilient and sustainable.[2] Key contributing factors to air pollution, particularly in cities, are car traffic and congestion. Indeed, research demonstrates that reduced congestion and traffic can lead to better air quality and subsequently improved health outcomes (Margaryan, 2021; Knittel et al., 2016; Currie and Walker, 2011). Frequently proposed measures to reduce automobile use and thereby air pollution include the extension of public transport supply and the reduction of its cost, ultimately encouraging a modal switch towards publicly provided means of transportation. This paper leverages a unique policy intervention in Germany to causally estimate how lowering public transport prices impacts air quality.

In June 2022, Germany introduced the "9-Euro-Ticket" (9ET), temporarily reducing regional public transport fares to a flat rate price of 9 Euros per month from June until August. This policy intervention is unusual in the sense that it temporarily and relatively suddenly reduced public transport fares for a whole country. Further, the implied price drop was substantial - in Berlin, for example, the price of the monthly standard ticket, which normally is 86 Euros, experienced a decrease of about 90%.[3] The aim of the ticket's introduction was twofold. First, mitigating the rising costs of living was a key objective of policymakers, who adopted the measures in the first half of 2022 as part of a relief package. Second, it was considered a potential means to promote environmentally friendly mobility by incentivising a reduction in car use and thereby carbon emissions and pollution (Spiegel Online, 2022).

In order to causally analyse whether this substantial reduction in public transport prices indeed has an impact on air pollution, we adopt a Difference-in-Differences (DiD) approach. More precisely, the empirical setting compares changes in air pollution between months May and June in non-treatment years before 2022 to changes in pollution in the same two months in 2022. We construct a state of the art Air Quality Index (AQI) and show that it decreases by more than eight percent as a consequence of the introduction of the ticket. Further, we document substantial effect heterogeneity and show that the effect is largest in

[2]See Goal 11 at https://sdgs.un.org/goals.

[3]See https://www.bvg.de/de/tickets-und-tarife/alle-tickets/zeitkarten/monatskarte.

urban areas, during work days and in areas with high levels of public transport provision. Ultimately, back-of-the-envelope calculations show that potential health benefits associated with a reduction in air pollution have the potential to exceed the costs of the intervention.

Our results are in line with recent findings that descriptively document a modal switch in response to the introduction of the subsidy. The 9ET itself was sold over 52 million times and survey evidence on individuals' mobility patterns clearly documents a strong increase in the use of public transport and a decrease in other modes of transportation in response to the ticket's introduction (DLR, 2022; VDV, 2022). For example, in a survey of 6,000 individuals by the German Public Transport Association, ten percent of 9ET owners said they avoided at least one of their daily car journeys by using the ticket (VDV, 2022). Further, studies on the 9ET's impact on traffic flows support that car traffic was indeed reduced: following the introduction of the ticket, the level of congestion in many German cities decreased between May and June 2022, while the amount of train travel increased (Süddeutsche Zeitung, 2022; German Federal Office for Statistics, 2022).

Expanding on this descriptive evidence, our paper is, to the best of our knowledge, the first to causally demonstrate that cheaper public transport can mitigate negative externalities of automobile travel such as air pollution. Thus, in the absence of effective first best road pricing systems that internalize the externalities of traffic, large subsidies on ticket prices might be a viable second best solution.

Our research adds to a large body of literature that shows a positive impact of public transit provision on air quality. Anderson (2014) and Bauernschuster et al. (2017) use unexpected public transport strikes as a quasi-natural experiment and find a temporary increase in congestion and in air pollution respectively. Other studies examine the effect of building or extending new subway lines, showing that they improve air quality (Chen and Whalley, 2012; Gendron-Carrier et al., 2022). For Germany, Lalive et al. (2018) find a positive effect of railway expansions on air quality. A second strand of literature develops and calibrates quantitative equilibrium models to quantify welfare effects of public transport investments. For instance, Parry and Small (2009) and Basso and Silva (2014) document significant welfare gains from subsidizing public transport. Borck (2019), in a counterfactual analysis, shows that subsidies reduce air pollution, albeit only modestly due to offsetting long term equilibrium effects such as residential relocation. In contrast, causal reduced form evidence on the effect of public transport pricing on air quality is, to the best of our knowledge, scarce and limited to the study of price increases. Yang and Tang (2018), using a synthetic control group approach paired with DiD analysis, find that an increase in Beijing's public transport fare led to a short run increase in air pollution of approximately 16 percent.

Fully understanding how public transport prices and in particular a price reduction via subsidized tickets impact traffic patterns and consequently air quality is crucial. Price adjustments, as opposed to infrastructural interventions, might offer an easier to implement measure for policymakers to reduce air pollution - especially in spatially constrained cities where extending existing networks or constructing new ones might not be an option.

A priori the effect of a public transport price reduction on air pollution is not clear. If individuals switch from car travel to public transport - as suggested by the aforementioned survey evidence - there might indeed be a fall in air pollution. However, if new passengers are predominantly individuals, who replace walks and bike trips with public transport or seek to avoid crowded trains, there might be no effect at all. Conclusive and causal evidence directly studying the impact of a public transport subsidy does not exist. Our paper closes this gap by providing evidence that a price reduction indeed can reduce air pollution. Further, in contrast to previous research on the relation of public transport and air quality we provide evidence for a whole country, Germany, and not only selected cities.

Overall, our results may have important policy and health implications: they show that subsidizing public transportation might be a viable option to reduce air pollution particularly in cities, thereby contributing to the UN's sustainability goal of creating more resilient, safer and healthier urban agglomerations. For instance, reductions in air pollution have been shown to substantially reduce cardiovascular diagnoses (Margaryan, 2021) and improve infant health (Knittel et al., 2016).

The remainder of the paper is structured as followed. Section 4.2 provides the relevant institutional background. Section 4.3 introduces the different data sources used for estimation and presents the estimation approach. Section 4.4 presents and discusses the results and Section 4.5 concludes.

4.2 Institutional Background

The 2021 general election in Germany resulted in a new coalition government of Social Democrats (SPD), the Greens and the Liberals (FDP). The creation of a sustainable mobility sector was one of the topics highlighted in the new government coalition agreement.[4] Since 2022, and particularly the start of the war in Ukraine on February 24th, living costs - especially in the form of energy prices - increased drastically. This prompted the government to pass a set of relief measures aiming to mitigate such costs and prices. The 9ET was part of the second relief package with the objective, among other things, of maintaining public transport affordable. Additionally, in particular the co-governing Green party frequently emphasised the ticket's potential to reduce carbon emissions and provide a more sustainable alternative for the mobility sector, as stipulated by the coalition agreement (Spiegel Online, 2022).

On February 23rd, 2022, as a consequence of rising energy and fuel prices in previous months, the ruling coalition parties decided upon a first relief package consisting of several measures such as temporary tax cuts. Following the outbreak of war in Ukraine a day later and the associated even stronger rise in gas and oil prices, a second relief package was passed in the German Bundestag. This second set of relief measures included a reduction of the energy tax on fuel in order to support people commuting by cars. Figure 4.A.1 in the Appendix depicts the development of gasoline and diesel prices in Germany in 2022. It shows the steep increase in price of all types of fuel after the start of the Ukraine war. After reaching a peak at the beginning of March, the prices stabilized at moderately lower levels until June 1st. The reduction in energy tax that took effect from then on implied a temporary drop in fuel prices. However, diesel regained pre-tax prices just after a few days. Overall, prices for all types of fuel remained at a relatively high level especially compared to pre-war times. Crucially, in our identification strategy presented in the following section we can control for daily fuel prices to account for these patterns and the tax cut.

In order to additionally compensate users of public transportation, the relief package also included the 9ET. For a total price of nine Euros a month, it allows its holders to use most types of public transportation like buses, subways, and regional trains all over Germany. The

[4] See https://cms.gruene.de/uploads/documents/Koalitionsvertrag-SPD-GRUENE-FDP-2021-2025.pdf.

ticket was available from June 1st, 2022 until the end of August of 2022.[5] One of the aims of subsidizing public transportation, in addition to relieving the financial burden on citizens, was to encourage its utilization and thereby promote sustainable modes of transportation. Just before June 1st, about seven million tickets and by the end of June approximately 21 million tickets had been sold (Handelsblatt, 2022; Süddeutsche Zeitung, 2022). Factoring in the roughly 9 million regular subscribers to monthly or yearly tickets whose fare is automatically reduced to 9 Euros from June to August, more than thirty million 9ETs were in circulation by the end of June.

4.3 Data and Empirical Approach

In order to analyse the 9ET's impact on air pollution, we collect pollution measurements for four key pollutants and data on covariates that have been shown to influence air pollution, such as weather conditions and holidays. Crucially, we also collect data on fuel prices allowing us to fully account for the tax cut in fuel prices that was simultaneously introduced with the 9ET. The remainder of this section firstly presents our different data sources before introducing the empirical approach.

4.3.1 Data

Air pollution data We use air pollution data for months May until September from 2018 to 2022. The data is provided as hourly measures by the Federal Environmental Agency (FEA) (Umweltbundesamt, 2022).[6] We observe whether a measuring station is located close to a street (traffic station) or in residential areas (background station). While the former provides information on air quality in relation to traffic, the latter rather indicates the general quality of the air in an area. We make use of this differentiation in our estimations: in the main specification we exclusively focus on air pollution concentrations measured by traffic stations, in order to test whether it is indeed the reduction in car traffic that reduced air pollution following the 9ET. In a heterogeneity check, we exclusively focus on background

[5] For more information on the relief packages, see https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Themen/Schlaglichter/Entlastungen/schnelle-spuerbare-entlastungen.html.

[6] The FEA API provides the data: https://www.umweltbundesamt.de/daten/luft/luftdaten/doc.

stations, where we expect the potential effect of the ticket to be smaller.[7] Further, the data includes information on whether the station lies in rural, suburban or urban areas.

In order to assess air quality we look at the pollutants nitrogen dioxide (NO_2), particulate matter with diameter less than 10 (PM_{10}) micrometers or smaller than 2.5 ($PM_{2.5}$) micrometers, as well as ozone (O_3). These we use to construct the AQI, commonly used by national and international authorities in order to assess pollution levels and provide information on potential health impacts at local levels. For the construction of the AQI, we follow a benchmark European air quality index (Van den Elshout et al., 2014), which is based on the core pollutants NO_2, PM_{10}, and $PM_{2.5}$ for traffic stations (our key outcome) and additionally O_3 for background stations. The index itself takes on values between 0 and 100, and is then further classified into four categories from very low pollution (index number 0-25) to high pollution (index numbers 75-100).[8]

Fuel prices We collect data on local fuel prices. Since 2013, fuel stations have been obliged to report each and every change of fuel prices (specifically for diesel and gasoline) in real time to the Market Transparency Office for fuels ("Markttransparenzstelle für Kraftstoffe") run by the German Cartel Office.[9] We aggregate all fuel prices to district level by taking the daily mean of all stations within each administrative entity.[10]

Meteorology When analysing air quality data, it is critical to control for current weather conditions. Wind and rain tend to improve air quality, since they clean the air e.g. from pollutants such as particulates. We use weather data aggregated to daily levels to control for mean temperature, mean wind speed and total precipitation levels. In order to do so, we acquired measurements from about 3.000 stations from the German Meteorological Service (DWD Climate Data Center (CDC), 2022). Since measurement stations are independently located from air quality measuring stations, we follow the approach by Auffhammer and

[7]In our data set, 411 ground-level stations measure the concentration of NO_2, 360 stations measure PM_{10} and 273 $PM_{2.5}$. One of the reasons why there are fewer monitoring stations for $PM_{2.5}$ is that the coverage of this pollutant started relatively late compared to the other pollutants and the measuring network is still being expanded.

[8]For a technical description for the construction of the AQI see Table 4 in Van den Elshout et al. (2014).

[9]For more information see https://www.bundeskartellamt.de/EN/Economicsectors/MineralOil/MTU-Fuels/mtufuels_node.html;jsessionid=4E22F5632D11B1D267F456C94321D7C6.2_cid390

[10]The price data is provided on the following website: https://creativecommons.tankerkoenig.de/.

Kellogg (2011) to match each air quality station to the nearest weather station.[11] Through this method, we are able to match about 99 percent of pollution observations to the weather variables of interest.

Further controls We control for a variety of potential economic and traffic-related variables. In particular, the German Federal Office for Statistics has provided us with district level data on the total mileage of regional and local trains as well as on the bus services per capita. We use this information in our heterogeneity analysis to check whether effects differ across districts with low and high levels of public transport provision.

Furthermore, we know whether a day falls on a weekend, a public holiday or on school vacations.[12] In order to be able to differentiate between effects in urban areas including their commuting zones, we resort to OECD definitions of Functional Urban Areas (FUAs).[13] Those are constructed based on commuters' daily movements and they consist of a city centre to which people commute to as well as the surrounding commuting zone (Dijkstra et al., 2022). In Germany, there are 96 FUAs in total.

4.3.2 Empirical Approach

We employ a DiD design using month June as the treatment group and May as the control group. For the pre-treatment period we focus on pollution measurements from years 2018 and 2019. We explicitly exclude 2020 and 2021 because there could be confounding factors that affect air pollution results in those years due to changes in COVID 19 restrictions in May and June 2020 and 2021, respectively.[14] We focus on working days outside of school

[11]The approach conducts the following steps: After calculating Vincenty distances, we identify the ten closest weather stations to each pollution station within a 75 kilometre distance and a maximum elevation difference of 200 meters. Then, the "primary station" is chosen as the closest meteorology station with more than 50 percent of non-missing observations and matched to the pollution data. Following, missing values are filled by regressing non-missing weather observations of primary stations on weather measurements of the nine other closest stations and using the predicted values from these regressions. Missing values still remain in the case that one of the nine stations had a missing observation. In this case the above step is repeated using the eight closest stations. The number of stations is then subsequently lowered. For a more thorough description see Auffhammer and Kellogg (2011).

[12]Information about state-level holidays and vacations are retrieved from http://www.feiertage-api.de/ and https://ferien-api.de/ respectively.

[13]Shapefiles are provided by the OECD at https://www.oecd.org/regional/regional-statistics/functional-urban-areas.htm.

[14]For example, in most federal states in-class teaching only restarted in late May/June in 2021 and mid and late May in 2020

holidays, as unusual traffic patterns are common during these.[15] Further, there is more construction work during school holidays, which can directly affect air quality and traffic.[16] Equation 4.1 describes the approach in more detail.

$$p_{symd} = \alpha_s + \gamma_{yf} + \eta_{mf} + \zeta_d + \beta Post_m \cdot \kappa_{2022} + X'_{symd}\theta + \epsilon_{symd} \tag{4.1}$$

p_{symd} is the logarithm of the AQI measured at station s, in year y for month m, at day of the week d. We regress this outcome on station level fixed effects, α_s, controlling for all potential time invariant station-specific observables and unobservables such as the number of lanes the station is placed at or the distance to rail tracks. Further, we include federal state specific year fixed effects, γ_{yf}, to control for general differences across years in each federal state f and federal state specific month fixed effects η_{mf} to account for differences in treatment and control months within each federal state. We also control for day-of-the-week fixed effects, ζ_d. This allows us to only compare pollutant levels at the same day of the week with one another. This is necessary as there are pollution trends across days of the week and thus comparing pollutant levels on a Wednesday and a Friday might falsely pick up a difference in pollution between years that can simply be attributed to these trends.[17] We then include a dummy variable $Post_m$ that equals one if the 9ET was in effect in the month of observation. Interacting the Post indicator with an indicator variable for the year 2022, i.e. κ_{2022}, then gives the DiD estimate for coefficient β. This setup allows us to test whether there indeed is a change in air quality between May and June in 2022, i.e. the year when the 9ET was introduced, that goes beyond the changes observed between these two months in previous years that were not subject to such a policy intervention. In addition, we control for a matrix of covariates, X', including linear and squared daily weather conditions in the vicinity of each pollution station such as wind speed, precipitation and temperatures, as these have been shown to influence pollution levels (see e.g. Auffhammer and Kellogg, 2011). Further, we control for daily average fuel prices and interact the fuel prices with the year fixed effects in order to account for different effects across years. Controlling for fuel

[15] There are peaks of traffic within and outside of cities at varying days, especially the beginning and the end of the respective holiday.

[16] In addition, the school vacations begin at different times every year and in each state, which makes controlling for the described patterns difficult.

[17] For example, on Fridays there may be a higher tendency to work from home.

prices is crucial as a fuel tax cut was introduced jointly with the 9ET (see Section 4.2). As detailed above, fuel prices fell only moderately in response to the tax cut compared to the yearly trend (see Figure 4.A.1). In the robustness checks of the paper we also present specifications excluding the first days in June and the last days in May to account for the temporary fall in gasoline prices and potential anticipation effects, which could entail that individuals drove less before June in order to wait for a fall in gas and public transport prices. In our main specifications, the error term ϵ_{symd} is clustered at the district level.

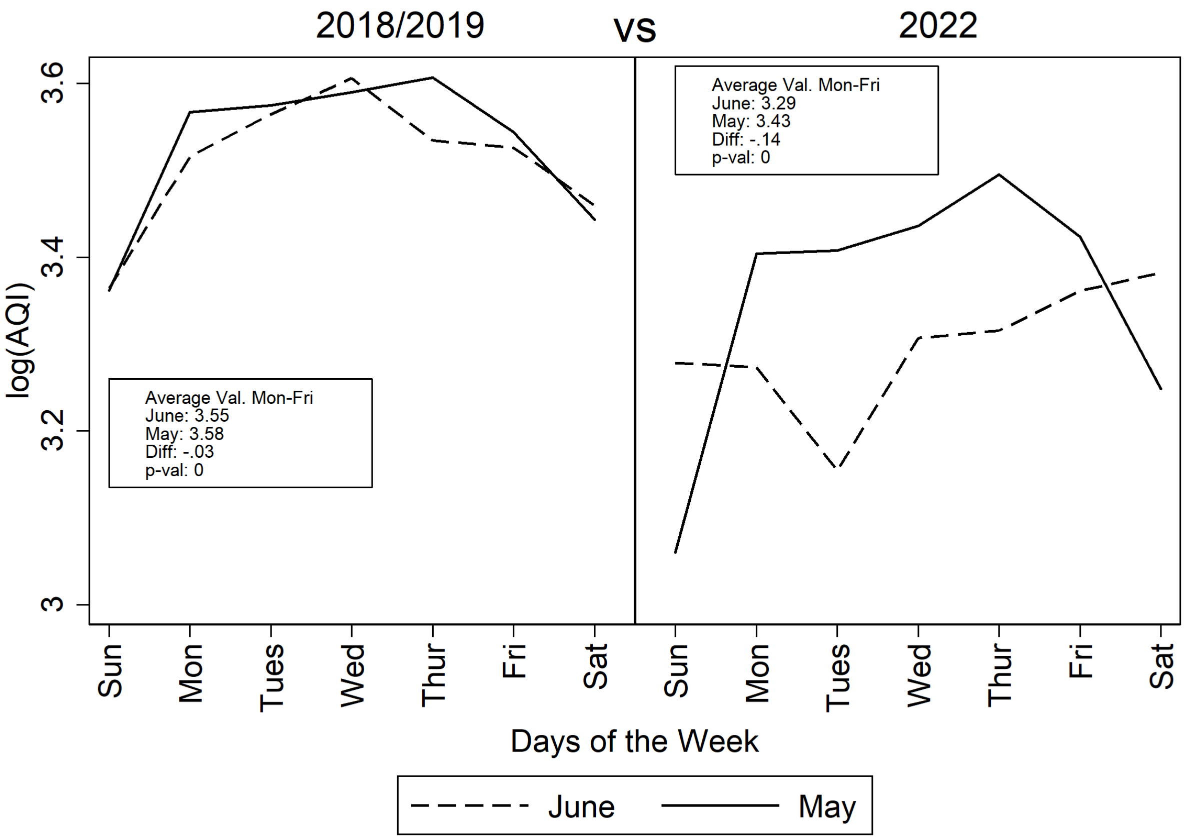
The DiD approach identifies the causal effect of the 9ET on air pollution as long as pollutant levels in control month May and treatment month June would have developed parallelly in the absence of treatment and conditional on the observed covariates and fixed effects. In the following section we check this assumption by carefully examining pre-trends, i.e. average air pollution levels in May and June in the years before 2022. In addition, we provide placebo tests together with our main results. In those, the reform date is shifted forward to 2019, thereby explicitly testing for common pre-trends.

The focus of the main analyses is explicitly on the months May and June, since the common trend assumption is likely to hold only for months close to each other, i.e. months that share similar characteristics. In order to present time patterns of the subsidy's impact, we will subsequently provide estimates including June, July and August as part of the treatment group. However, we consider these estimates mainly of suggestive nature, since days in May and July/August do not necessarily show common trends in air pollution, especially since the bulk of summer vacations in most states fall in July/August and vacation patterns change over the years.

4.4 Results

Descriptive Results We can check the identifying assumption by comparing pre-trends, i.e. air quality before 2022 for treatment (June) and control month (May). More precisely, we compare average values for a given day of the week in May and June before and after the introduction of the 9ET. Figure 4.1 shows the day of the week specific average for the logarithm of the AQI for the pre-policy period, i.e. 2018-2019 (left), and the post policy period (right), i.e. the year 2022, respectively. The graph illustrates that when assessing

Figure 4.1: Day of the week averages for log(AQI).



Note: Own calculations. Day of the week averages for log(AQI) before and after the introduction of the 9-Euro-ticket for treatment (June) vs. control month (May), excluding school and public holidays. Average working day values and differences for May and June before and after the reform are displayed in the graph's textboxes. The p-values are derived from a two-sided t-test on the difference between both values. Corresponding values for weekends are 2018/19: June= 3.411, May= 3.400, Diff= −0.011, p-value= 0.39; 2022: June= 3.331, May= 3.169, Diff= −0.162, p-value= 0.00.

pre-trends it is crucial to only compare pollutant levels at the same day of the week with one another, as there are clear general patterns in the average pollutant levels for a certain day of the week. For example, there generally appears to be a greater level of pollution in the middle of the week and lower levels over weekends.

Figure 4.1 supports our identifying assumption: during the pre-treatment period, AQI values for treatment and control months behaved relatively similarly and followed a parallel trajectory throughout the week, albeit at slightly different levels. The corresponding average values for working days depicted in the textbox show that on average pollution on working days in June is 0.03 log points lower than in May. Looking at the post treatment period, i.e. the year 2022, we can see that average pollution for the control month May follows a roughly similar weekly trajectory as in the pre-treatment period, albeit at an overall lower level. Unlike prior years, average pollution in the treatment month falls substantially compared to the control month. In particular, during the week there appears to be a sharp decline in the log of the AQI in 2022, which suggests an improvement in air quality. This is further confirmed by comparing the average log values of the AQI for working days, i.e. Monday to Friday, in May and June in the post-treatment period, as presented in the textbox: on average, air pollution is 0.14 log points smaller in June than in May 2022 - a remarkable difference compared to the pre-treatment period.

In contrast, weekend average values in June in comparison to May visibly increase across pre- and post-treatment years. In 2018/19 average pollution on weekends was 0.011 log points higher in June. However, average pollution on weekends in June 2022 was 0.162 log points higher than in May 2022. Anecdotal evidence suggests that due to lower public transport prices for regional trains there was a substantial influx in train travel over weekends (Zeit Online, 2022). In particular students and young people are said to have used trains increasingly for weekend trips and holidays, leading to newspaper articles about and television coverage concerning crowded and delayed trains. This potentially had a deterring effect on travellers who might have switched to car use in order to avoid crowded trains, which could possibly explain the observed increase on weekends.

Following the visual inspection of the graphs, we expect a negative effect of the 9ET on air pollution during weekdays and a positive effect during weekends in our main estimation. Further, we revisit our identification assumption by explicitly testing for common pre-trends conditional on covariates and fixed effects using a placebo framework solely focussing on the years 2018/19.

Main Results: Table 4.1 presents the results using log(AQI) as an outcome variable. In all specifications we control for the full set of fixed effects and covariates described above. Specification (1) shows the results for the months of May and June. We find a negative point estimate for the interaction term suggesting that the average AQI fell by more than eight percent. The estimated effect is significant at the one percent level. Since an index value of zero indicates the best possible air quality, the result suggests an overall average improvement in air quality. In specification (2) and (3), we additionally include the months

July and August. As previously pointed out, these estimates are of suggestive nature to analyse how the subsidy's impact developed over time. Point estimates in both specifications remain negative and in case of specification (2) relatively similar to the main specification. When including August, there is a clear decrease in absolute size of the point estimate, suggesting a fall in the 9ET's impact on pollution. One potential explanation here might be a switch back from public transport towards other means of transportations after initially trying public means of transportation in June. However, these results should be interpreted cautiously and as suggestive evidence since a comparison of May and August within our setting is not ideal (see Section 4.3.2).

Table 4.1: Main Results: log(AQI)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interaction	-0.0877***	-0.0833***	-0.0630***	-0.0089	-0.0125	-0.0230	0.0665**
	(0.0207)	(0.0200)	(0.0157)	(0.0172)	(0.0170)	(0.0168)	(0.0314)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observed Years	18,19,22	18,19,22	18,19,22	18,19	18,19	18,19	18,19,22
Observed Months	May-June	May-July	May-August	May-June	May-July	May-August	August-Sep
Observations	21926	26485	30737	14523	17093	20413	15054

Note: Source: own calculations. Standard errors in parentheses and clustered at district level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the log of the AQI as outcome variable. Column (1) shows the results for a basic DiD approach only including months May and June. Columns (2)-(4) augment the estimation with subsequently adding additional months. Columns (4) - (6) implement placebo tests where year 2019 is used as the new policy date. Column (7) shows the effect of abolishing the 9ET. All specifications control for the full set of covariates and fixed effects.

In column (4)-(6) we repeat the estimations for the years 2018 and 2019 only. Crucially, we treat year 2019 as if the 9ET was introduced in June 2019. This allows us to run a placebo test and explicitly examine the common pre-trends assumption. The point estimates are slightly negative and insignificant in each specification, thereby supporting the notion that there is no systematic difference between treatment and control months across years previous to the actual introduction of the ticket.

Lastly, in column (7) we show the outcome of analysing whether the abolishment of the ticket from August to September resulted in an increase in air pollution.[18] Using August

[18] August and September coincide with the end of school holidays in many federal states. Thus there are fewer observations stemming from non-holiday days, which we use in our estimation.

as the treatment group and September as the control group or vice versa might imply that our estimation will also pick up potential spillover effects, undermining our identification strategy. For example, if the introduction of the ticket managed to incentivize lasting behavioural change, some individuals might have permanently switched towards public transport. This in turn would impact air pollution in September as well as August 2022, ultimately violating the Stable Unit Treatment Value Assumption (SUTVA) commonly needed in DiD approaches. Nonetheless, we provide these results as suggestive evidence with the aim of giving additional information on the impact of the end of the 9ET. The positive point estimate in column (7), which uses August as the control and September as the treatment month, confirms that air pollution was significantly lower during the period when the 9ET was in effect than during the months with regular public transit fares.

All in all the results indicate a substantial fall in the AQI of more than eight percent and suggest a decrease in the effect over time.

In comparison to other literature studying the relation between public transit and air pollution our effect size is somewhat smaller. Bauernschuster et al. (2017), for instance, find a 14 percent increase in particle pollution in response to a public transport strike in Germany's five largest cities. Yang and Tang (2018) find a short term increase of approximately 16 percent in a Chinese air pollution index for Beijing in response to an increase in public transport fares. Our estimation in turn focusses on a *fall* in transport prices. Further, our effects are potentially less pronounced, as we focus on all of Germany and not just the largest cities or a single urban agglomeration.

Mechanisms and Heterogeneous Results In a next step, we split our sample along several dimensions to analyse heterogeneous effects and potential mechanisms. Table 4.2 depicts the results. All specifications control for the full set of covariates and fixed effects and use observations from months May and June. First, in columns (1)-(4) we split our sample into the core of a Functional Urban Area (FUA), the total FUA and non-FUA (rural) areas.[19] The effect is driven by a substantial reduction in pollution in core areas. One potential explanation is that most jobs are situated in core areas. If more individuals commute to these core areas by public transport there will be less traffic and congestion

[19]The total FUA includes the core as well as the commuting zone of a FUA.

there, implying a fall in pollution concentration levels. Another possible explanation is that there is more supply of public transport in these areas, which would then allow for an easier switch to public means of transportation.

Table 4.2: Heterogeneous Results

	(1) Core FUA	(2) Non Core	(3) FUA	(4) Non FUA	(5) High PT	(6) Low PT	(7) Weekend	(8) Weekday	(9) Background St.
Interaction	-0.1095***	-0.0062	-0.0985***	-0.0172	-0.1019**	-0.0836***	0.0101	-0.1244***	-0.0241**
	(0.0208)	(0.0441)	(0.0206)	(0.0690)	(0.0418)	(0.0242)	(0.0391)	(0.0218)	(0.0113)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18001	3925	19894	2032	5554	16372	6211	15715	42855

Note: Source: own calculations. Standard errors in parentheses and clustered at district level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the log of the AQI as outcome variable. Column (1) shows the results for stations from the core of functional urban areas, column(2) for stations positioned outside the core. Specification (3) just includes observations from functional urban areas and specification (4) from outside these areas. (5) solely uses measurements from stations in districts in the highest 25 percentile of Public Transport (PT) supply and (6) in districts in the lowest 75 percentiles of public transport kilometres per person. (7) and (8) split the sample into weekends and weekdays and (9) solely includes measurements from background stations. All specifications control for the full set of covariates and fixed effects.

In order to generally assess the role of public transport supply, we split our sample into districts with a relatively high and low level of public transport infrastructure (see columns (5)-(6)). We can do so by using data on public transport kilometres per population for districts in Germany. More precisely, we split our sample into districts in the highest 25 percentiles of public transport kilometres per person and the lowest 75 percentiles. The results show that for both groups we find significant effects. As expected, the effect size is more pronounced for the group with higher levels of public transportation.

Next, in columns (7)-(8) we confirm what the visual examination above indicated: the effect is driven exclusively by weekdays. For weekends we find a positive, albeit insignificant effect when including the full set of covariates and fixed effects.

Lastly, in our data we can differentiate between air quality measured by stations directly next to roads and background measurements in quieter areas. If there indeed is less traffic due to more individuals using public transport, the effect at stations directly exposed to traffic should be stronger. In our main specifications we have so far exclusively focused on traffic stations. In column (9) of Table 4.2 we repeat our estimation for background stations. Only using background stations we would expect a smaller effect size, as these stations are not directly positioned next to streets. The results confirm this notion, as the point estimates remains negative and significant, albeit relatively small in absolute size.

Robustness In the Appendix we repeat our main estimation for different standard error clusters, i.e. at the station specific level and at the station-year level. Tables 4.A.1 and 4.A.2 show the results. Inference and statistical significance do not change across these different clusters.

Further, as detailed in Section 4.2, a tax cut on gasoline prices was implemented from June, 1st in addition to the 9ET. The corresponding law, however, was already passed on May 19th, 2022. Car owners thus might have waited for June 1st in order to purchase new fuel and drive again. As a consequence, pollution levels directly before June 1st might have been slightly dampened and heightened at the beginning of June, leading to a lower estimated decrease in pollutants after the introduction of the 9ET. In order to analyse these potential anticipation effects in the run up to June 1st, we repeat our main estimation for the log(AQI) in a leave-one-out exercise. Here we successively drop the first and last two days in May and June respectively. The results, presented in Table 4.A.3 in the Appendix, indicate no substantial difference in the point estimates for the variable of interest, which remains very similar in size across specifications.

Lastly, we use each pollutant (and its logarithm) that contributes to the AQI as an outcome variable separately in order to check that it is not simply the construction method of the AQI that drives the results. Moreover, the calculation of outcomes for individual pollutants facilitates the interpretation of the results in the following paragraph. Table 4.A.4 in the Appendix shows the results for PM_{10}, $PM_{2.5}$ and NO_2 respectively. All point estimates are significant and negative supporting the results in the main specification using the AQI as the key measure for air quality.

Implications Overall, our results document a decrease in air pollution in response to the 9ET, which might impact further outcomes such as individuals' health. Indeed, the negative relationship between air pollution and health hazards is a well established fact (Anderson, 2009). Generally, AQI values below 50 are categorized as an overall "good" air quality

(Van den Elshout et al., 2014). While the average AQI value in Germany is about 35, and thus an eight percent increase still leaves us within the range of good air quality, there are nevertheless numerous studies that find significant health effects even at moderate levels of air quality deterioration. For instance, the introduction of Low Emission Zones (LEZ) in Germany lowered PM_{10} concentrations by about 3 percent or approximately 0.8 $\mu g/m^3$, which consequently reduced the number of patients with cardiovascular diagnoses by approximately 2-3 percent (Margaryan, 2021). According to back-of-the-envelope calculations, this resulted in economic welfare gains of more than 4.4 billion Euros (Margaryan, 2021). Assuming the relationship between cardiovascular health problems and PM_{10} to be linear, this would imply an even larger effect of the 9ET, as PM_{10} was reduced by more than 1.5 $\mu g/m^3$ following the introduction of the ticket (see Table 4.A.4).

Further research suggests a significant impact of pollution reductions on infant health and mortality, even at levels below common thresholds of concern (Simeonova et al., 2021; Knittel et al., 2016; Currie et al., 2011). For example, PM_{10} reductions of 1 $\mu g/m^3$ have been shown to imply ten saved lives per 100,000 births (Knittel et al., 2016). Taking the value of a statistical life of 1.72 million Euro in Germany, this would imply saved costs of about 200 million Euros per year, given total births of 800,000 and our finding of a 1.5 $\mu g/m^3$ reduction in PM_{10}.

While such back-of-the-envelope calculations are based on highly simplified assumptions, they nevertheless indicate that positive health implications alone could have the potential to amortise the costs associated with the ticket of approximately 2.5 billion Euro.[20] This is especially the case since the exemplary calculations conducted above ignore a range of further health hazards caused by pollution, e.g. pulmonary diseases, lung damages, respiratory distress, or birth defects (for an extensive list of potential health effects only caused by particulate pollution, see Pope III and Dockery, 2006).

4.5 Conclusion

In this paper, we provide novel causal evidence on the effect of a large scale public transport subsidy on air pollution. The policy we study is unique in the sense that it reduced public

[20]The ticket's cost were estimated based on foregone ticket revenues (Die Bundesregierung, 2022).

transport fares for a whole country in some areas by as much as 90 percent. To the best of our knowledge, we are the first to causally study the effects of a public transport subsidy on air pollution. In doing so, we extend the previous literature, which has focussed predominantly on price increases (Yang and Tang, 2018), calibrated quantitative models (Borck, 2019) or leveraged natural experiments such as unexpected public transit strikes (Bauernschuster et al., 2017) to study the general relationship between public transit and air pollution.

As our key finding, we show that pollution levels fall in response to the policy intervention. In particular, the AQI decreases by more than eight percent. Further, we document effect heterogeneity showing that the effects are largest during the week, i.e. when individuals commute to and from work. We also show that the effects are more pronounced in urban areas and regions with a well developed public transport network.

Our results are relevant for policymakers and researchers alike. First, our findings suggest that subsidizing public transportation can indeed incentivize a modal switch, which sparks a decrease in pollution levels and potentially other outcomes not studied in this paper such as carbon emissions. Further, the results echo findings of quantitative equilibrium models, e.g. by Borck (2019), who finds a decrease in pollution in response to lifting public transport fares all together. Our effect sizes lie in between the relatively small effect of public transport on air pollution documented in general equilibrium models by Borck (2019) and relatively large effects found in other papers studying price changes in public transport fares such as Yang and Tang (2018). The differences in effect sizes is plausible, as we estimate the short-term impact of the subsidy and can hence not account for long term equilibrium effects, which might work in offsetting directions.[21] Further, we focus on pollution measurements across Germany, whereas other papers have singled out large urban agglomerations where a stronger relationship between public transport and air pollution seems plausible, and indeed is also shown in our heterogeneity analyses. Additionally, our back-of-the-envelope calculations show that if the 9ET was to be permanent and behavioural adjustments in the modal split were to remain, the costs of the ticket in theory could be amortised by positive health effects and their related economic burden.

Overall, the findings thus indicate that subsidizing public transport might be a viable option

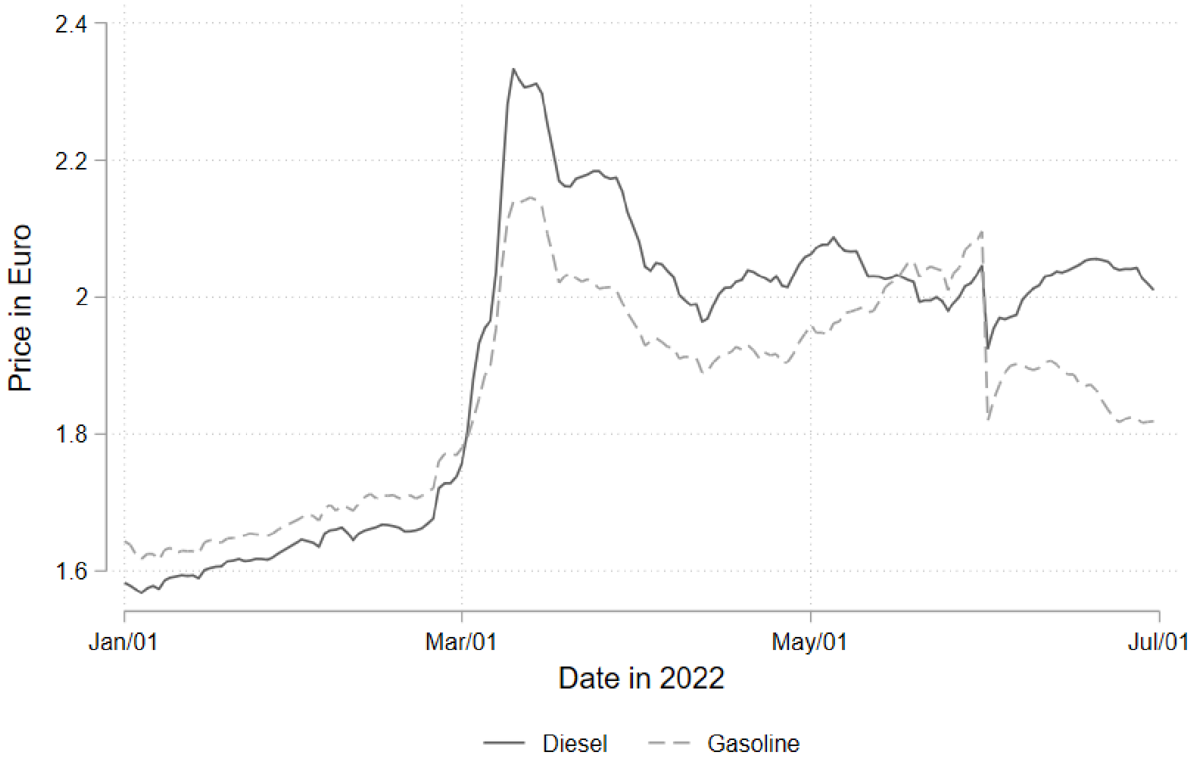
[21]In particular, (Borck, 2019), in his model, emphasizes relocation mechanisms and an increase in housing consumption and residential emissions.

to reduce pollution in the short term and given results by Borck (2019) potentially also in the long term. Thereby, a policy intervention such as the 9-Euro-ticket may indeed contribute to the UN's sustainability goal of creating more resilient, safer and healthier urban agglomerations.

Appendix

4.A Additional tables and figures

Figure 4.A.1: Development of average gasoline and diesel prices in Germany, 2022



Note: Own illustration. The graph shows the development of gasoline and diesel prices between January, 2022 and July, 2022. Gasoline and diesel data are retrieved from https://creativecommons.tankerkoenig.de/. The gasoline price is the average of E5 and E10 (indicating ethanol content in gasoline of 5 and 10 percent respectively) prices.

Table 4.A.1: Results for SE Station Clusters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interaction	-0.0877***	-0.0833***	-0.0630***	-0.0089	-0.0125	-0.0230	0.0665**
	(0.0202)	(0.0197)	(0.0160)	(0.0167)	(0.0165)	(0.0163)	(0.0291)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observed Years	18,19,22	18,19,22	18,19,22	18,19	18,19	18,19	18,19,22
Observed Months	May-June	May-July	May-August	May-June	May-July	May-August	August-Sep
Observations	21926	26485	30737	14523	17093	20413	15054

Note: Source: Own calculations. Standard errors in parentheses and clustered at station level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the log of the AQI as outcome variable. Column (1) shows the results for a basic DiD approach only including months May and June. Columns (2)-(4) augment the estimation with subsequently adding additional months. Columns (4) - (6) implement placebo tests where year 2019 is used as the new policy date. Column (7) shows the effect of abolishing the 9ET. All specifications control for the full set of covariates and fixed effects.

Table 4.A.2: Results for SE Station-Year Clusters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interaction	-0.0877***	-0.0833***	-0.0630***	-0.0089	-0.0125	-0.0230	0.0665**
	(0.0189)	(0.0178)	(0.0152)	(0.0151)	(0.0149)	(0.0146)	(0.0279)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observed Years	18,19,22	18,19,22	18,19,22	18,19	18,19	18,19	18,19,22
Observed Months	May-June	May-July	May-August	May-June	May-July	May-August	August-Sep
Observations	21926	26485	30737	14523	17093	20413	15054

Note: Own calculations. Standard errors in parentheses and clustered at station-year level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the log of the AQI as outcome variable. Column (1) shows the results for a basic DiD approach only including months May and June. Columns (2)-(4) augment the estimation with subsequently adding additional months. Columns (4) - (6) implement placebo tests where year 2019 is used as the new policy date. Column (7) shows the effect of abolishing the 9ET. All specifications control for the full set of covariates and fixed effects.

Table 4.A.3: Leave-One-Out Results

	(1)	(2)	(3)
Interaction	-0.0877***	-0.0876***	-0.0829***
	(0.0207)	(0.0213)	(0.0211)
Covariates	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Station FE	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes
Observations	21926	21660	21395

Note: Own calculations. Standard errors in parentheses and clustered at district level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Outcome variable is the log of the air quality index. (1) leaves out no observations before and after the introduction, (2) the first day on each side and (3) the first two days on each side. All specifications control for the full set of covariates and fixed effects.

Table 4.A.4: Results for Different Air Pollutants

	(1) log(PM_{10})	(2) log($PM_{2.5}$)	(3) log(NO_2)	(4) PM_{10}	(5) $PM_{2.5}$	(6) NO_2
Interaction	-0.1018***	-0.1300***	-0.1066***	-1.5706***	-0.9243**	-2.5090***
	(0.0251)	(0.0400)	(0.0204)	(0.4303)	(0.3513)	(0.6609)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18186	11118	21043	18186	11118	21043

Note: Own calculations. Standard errors in parentheses and clustered at district level, significance levels* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns (1)-(3) show the logged outcomes for different pollutants separately, while columns (4)-(6) repeat the exercise using absolute values. All specifications control for the full set of covariates and fixed effects.

Chapter 5

The causal effect of cycling infrastructure on traffic and accidents: Evidence from pop-up bike lanes in Berlin.[1]

Abstract

This paper analyses the effects of new bike lanes on traffic volume, congestion, and accidents. In order to obtain causal estimates, I exploit the quasi-random timing and location of newly built cycle lanes. Crucially, the new cycling infrastructure replaces existing car lanes thereby reducing available space for motorized traffic. Using a variety of difference-in-differences style methods on geocoded data, I show that the construction of pop-up bike lanes significantly reduced average car speed by 8 to 12 percent and up to 16 percent in peak traffic hours. The results for car volume are insignificant in most specifications. Ultimately, the data do not allow for a conclusive judgment of accidents.

[1] Sole authored.

5.1 Introduction

Traffic has become progressively worse over the past decades.[2] It does not only have a tremendous environmental impact (Schwela and Zali, 1998) but also affects numerous aspects of life quality (Gifford and Steg, 2007). More recently, policy makers have aimed to reduce such negative traffic externalities by building new bicycle paths. In cities around the world, such as Bogotá, New York, Jakarta, Mexico city, London, Paris, and Berlin, authorities have worked on improving the local cycling infrastructure.[3] One common measure in this regard was the conversion of car lanes into bike lanes. Among the expected benefits, as proclaimed by Berlin's bicycle traffic plan[4], are improved local air quality, safer streets, and more efficient use of public space as motorists switch to bicycles. However, little is known about the actual, potentially unintended effects on traffic and road dynamics caused by such infrastructural interventions.

This paper investigates the effects of new bike lanes on traffic volume, congestion and accidents. Specifically, I consider new cycling infrastructure that results from converting car lanes into bicycle lanes within a large metropolitan area. The analysis uses Pop-Up Bike Lanes (PUBLs) in Berlin, which were installed between March and June of 2020, and hence during the COVID-19 induced lockdowns.[5] While the pandemic caused an immediate decrease in traffic volume throughout the entire city, the unexpected and quasi-random installation of the cycling lanes that followed allows for the identification of changes in the outcomes over time and circumvents endogeneity problems such as reverse causality. My findings suggest a decrease in average car speed by between 8 and 12 percent, and thus an increase in congestion on these streets. Absolute accidents are not significantly affected according to my estimations.

My research is of interest for at least three reasons. First, congestion and accidents hinge on traffic volume and are among the most severe agglomeration diseconomies (Ahlfeldt

[2]See for instance https://www.pwc.com/us/en/industrial-products/publications/assets/pwc-mobility-insights-congestion.pdf

[3]See e.g.: Bogotá: https://www.c40.org/case-studies/upgrade-of-the-cycle-network.../,
New York: https://rpa.org/work/reports/the-five-borough-bikeway,
Mexico city & Jakarta: https://www.itdp.org/.../cycling-and-mexico-city-better-than-before/.

[4]https://www.berlin.de/sen/uvk/.../radverkehrsplan/radverkehrsplan.pdf.

[5]Throughout the paper, I will interchangeably refer to the instalment of PUBLs on a street with the terms "treatment", "event", and "intervention".

and Pietrostefani, 2019; Shefer and Rietveld, 1997). Congestion is costly in various ways, especially in the form of time losses, wasted fuel consumption, and as a consequence an increase in CO_2 emissions (Vickrey, 1969; Treiber et al., 2008; Schrank et al., 2015). In Germany, congestion caused an average time loss per driver of about 40 hours in 2021.[6] Additionally, congestion interacts with local pollution, which is a major cost to cities itself and bears adverse health effects e.g. for infants (Borck and Schrauth, 2021; Currie and Walker, 2011; Knittel et al., 2016). There is also evidence that congestion may hinder economic growth in terms of income and employment (Jin and Rafferty, 2017; Hymel, 2009). Accidents are associated with substantial economic costs such as external insurance costs or by causing more congestion (Edlin and Karaca-Mandic, 2006). Second, cycling worldwide is increasingly popular[7] and many policy makers' aim of motivating city dwellers to cycle more requires the construction of better bike lanes (Yang et al., 2019). Due to the limited space in large congested cities, such as Berlin[8], new infrastructure projects require special care in planning them. Accordingly, a study based on such a large congested metropolitan region serves local authorities in their decisions-making process for new infrastructure projects particularly well. Third, PUBLs are explicitly interesting as an object of study. On the one hand, they have already been implemented frequently and proved to be easy-to-install and low-cost interventions. On the other hand, they were placed very suddenly and quasi-randomly. This allows to identify the causal effects of a policy instrument that is a viable possibility for future bicycle infrastructure measures.

While, to my knowledge, this is the first paper to empirically assess the effects of new bicycle infrastructure, and explicitly of PUBLs, on its unintended external effects, the paper connects to a growing amount of research on congestion. Various studies consider different measures that have been implemented in the past to target it. A congestion charge introduced in London, which levied a toll during prime commuting hours, was found to increase traffic speed and reduce total miles driven. A consequence of lower traffic levels was improved air quality and a decrease of the amount as well as the rate of accidents (Green et al., 2016,

[6] https://inrix.com/press-releases/2021-traffic-scorecard-de/. Studying the effects Economic costs are calculated based on values of time as suggested in the same study.

[7] See e.g. https://www.itdp.org/2021/10/26/cycling-is-booming-and-not-just-where-you-think/.

[8] Berlin is among the most severely affected cities regarding time losses due to congestion in Germany. Four out of the ten most congested German streets are situated in the capital city. The calculated costs amounted to more than 800 million Euro in 2021 for this city alone.

2020). Apart from such policies, which directly aim to tackle congestion, there is robust evidence that public transportation and the extension of its network may reduce congestion and lead to significant social benefits, e.g. due to higher air quality and reduced travel times (Anderson, 2014; Bauernschuster et al., 2017). Another related strand of research examines whether an increased road supply could impact traffic. The extension of road infrastructure was found to proportionately increase the amount travelled in the long run, thereby not affecting congestion (Duranton and Turner, 2011).

Despite its increasing importance, the role of bicycles in cities and the accompanying expansion of bicycle infrastructure is still on the fringe of academic research. Among the few exceptions aiming to identify causal effects of cycling on traffic related outcomes, Hamilton and Wichman (2018) found that neighbourhoods with bike-sharing stations had significantly lower congestion levels compared to similar, but untreated neighbourhoods in the Washington D.C. area, which hints towards a supply-driven change in commuting behaviour towards more cycling if more bikes are provided. Buehler and Pucher (2012) and Goodman et al. (2013) also show that a better and safer cycling infrastructure correlates with an increase in the propensity toward bicycle use. Furthermore, whether cycling routes, or "cycle superhighways" in this specific case, are safer or not was found to depend on physical characteristics, e.g. whether cyclists were separated from other forms of travel (Li et al., 2017). Those cycle superhighways were moreover found to reduce traffic volume without affecting average traffic speed (Bhuyan et al., 2021).

This paper contributes to and improves upon the existing literature in various ways. To the best of my knowledge, it is the first study to analyse a reduction of road space for cars in order to make it available for bikes. It is crucial for policy makers to learn more about this type of intervention since space in large metropolitan areas is scarce and the conversion of car lanes into bike lanes is one viable option to accelerate the transformation of cities towards more environmentally friendly places. Additionally, it connects to the much discussed "fundamental law of road congestion" (Duranton and Turner, 2011), which suggests a response of congestion as a consequence of building new infrastructure with an elasticity of one. To the best of my knowledge, there is no evidence on short to midterm effects, and the question of what happens to traffic-related outcomes just after a reduction of lane kilometres. Besides, the substitution of car lanes with bike lanes provides a direct alternative in travelling mode on the respective street rather than being a mere extension of infrastructure. By considering the effects of PUBLs on accidents, I contribute to the strand of literature, which analyses road safety (Edlin and Karaca-Mandic, 2006; Green et al., 2016; Shefer and Rietveld, 1997). Specifically, the study links to the question of how enhanced cycling infrastructure affects the safety of different types of road users (Li et al., 2017).

Furthermore, it is to the best of my knowledge the first paper to use a quasi-experimental design in order to identify causal effects. In general, it is challenging to estimate the causal effects of new bike lanes on city-related outcomes due to the fact that their creation often is meant to be a response to outcomes like road safety or congestion. Thus, city authorities, for example, seek to reduce street accidents by creating safer cycling infrastructure. This paper addresses this type of reverse causality by looking at (quasi) randomly built bike lanes. The choice of roads, except for a few controllable properties, was random and did not depend on prior road dynamics. Additionally, the roll-out of such lanes is normally very slow. Pop-up lanes are a chance to circumvent anticipation effects, because of their very fast and sudden construction. Apart from that, the results of the paper may serve as a valuable contribution to structural models, which consider infrastructure within cities. For policy makers, the results may hint towards potential problems accompanied with the sudden rezoning of car lanes. This allows them to address and tackle relevant problems prior to taking measures. The paper also helps to contribute to the question of costs and benefits of making a city more friendly to bikes in terms of new infrastructure.

In order to identify the causal effects and the potential heterogeneity in the development of the outcomes, I use an event study approach (Clarke and Tapia-Schythe, 2021) and standard two-way fixed effects models. The accident analyses are extended by a synthetic control group design (Abadie et al., 2010), which uses observable characteristics of treated and potential control units in order to determine a suitable comparison group. Besides my main outcome of an 8 to 12 percent decrease in average car speed, I find that the effect size increases to about 16 percent in the main business hours of traffic. Moreover, the results point towards modest changes in car volume. However, these can not be clearly attributed to PUBL installations. I also test for substitution effects on streets close-by, since the measures might merely relocate traffic as drivers seek to avoid affected routes. Accordingly, new cycling lanes increased traffic without affecting congestion on untreated streets in close distance to PUBLs. This suggests a new equilibrium with a more equal distribution of cars on the inner city road network. For accidents, I do not find any significant changes caused by PUBLs, which might suggest an increase in per-cyclist safety due to the rise of cyclists in the city (Kraus and Koch, 2021).

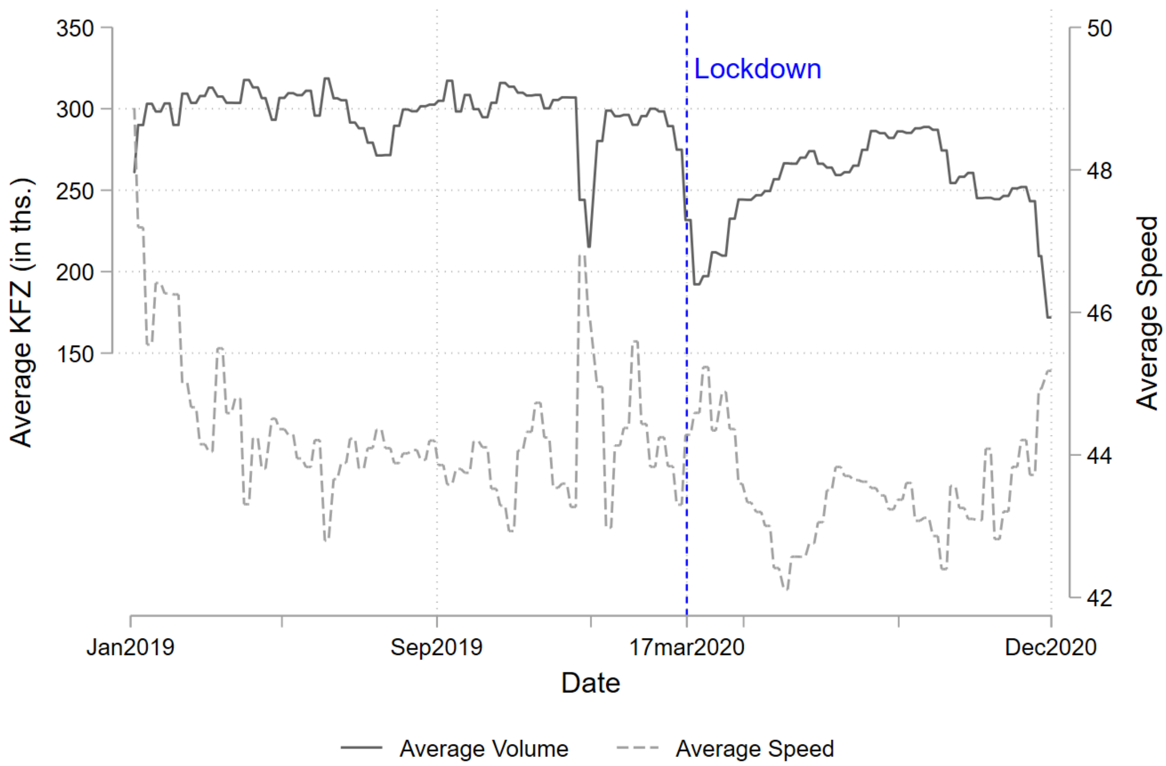
The remainder of this study is structured as follows. I first give an overview about important background knowledge and theoretical considerations in Section 5.2 before describing the data and providing descriptive analyses in Section 5.3. Section 5.4 details the methodological approach, and results are presented in Section 5.5. Section 5.6 discusses the results and puts them into perspective while Section 5.7 concludes.

5.2 Background and Theoretical Considerations

Political premises In Berlin, the coalition government of the three parties SPD, Die Grüne, and Die Linke (Social-Democrats, the Green Party and the Left Party) has set the goal of transforming city life by making it more friendly especially to pedestrians, cyclists, and people using public transport. The goals are written out in the mobility law, which was passed in July of 2018.[9] The law predominately contains plans regarding the city's traffic infrastructure and how they are to be implemented. This includes aspects of organization and funding. For cycle lanes, the aim was to extend the existing infrastructure in order to make cycling more attractive and to increase the share of cycling in the modal split. One major aim was to increase the safety of bicycle users and to reduce, and possibly avoid, cycling accidents. The plan also includes the construction of cycling highways, which primarily are supposed to connect outer parts of the city with the city centre. More goals of the plan include using space more efficiently since bikes require less street capacity compared to cars, to reduce local pollution, and to enhance healthiness by incentivising an increase in physical activity.[10] Since the construction work was planned until 2030, most of the structural measures had not yet been realized in the time frame considered in this paper.

[9]MobG BE - Abschnitt 3: Entwicklung des Radverkehrs (development of bicycle traffic) https://gesetze.berlin.de/perma?d=jlr-MobGBEpG6 is the specific chapter in the law about bicycle traffic.
[10]Compare https://www.berlin.de/sen/uvk/.../radverkehrsplan/radverkehrsplan.pdf.

Figure 5.1: Traffic in Berlin 2019 - 2020

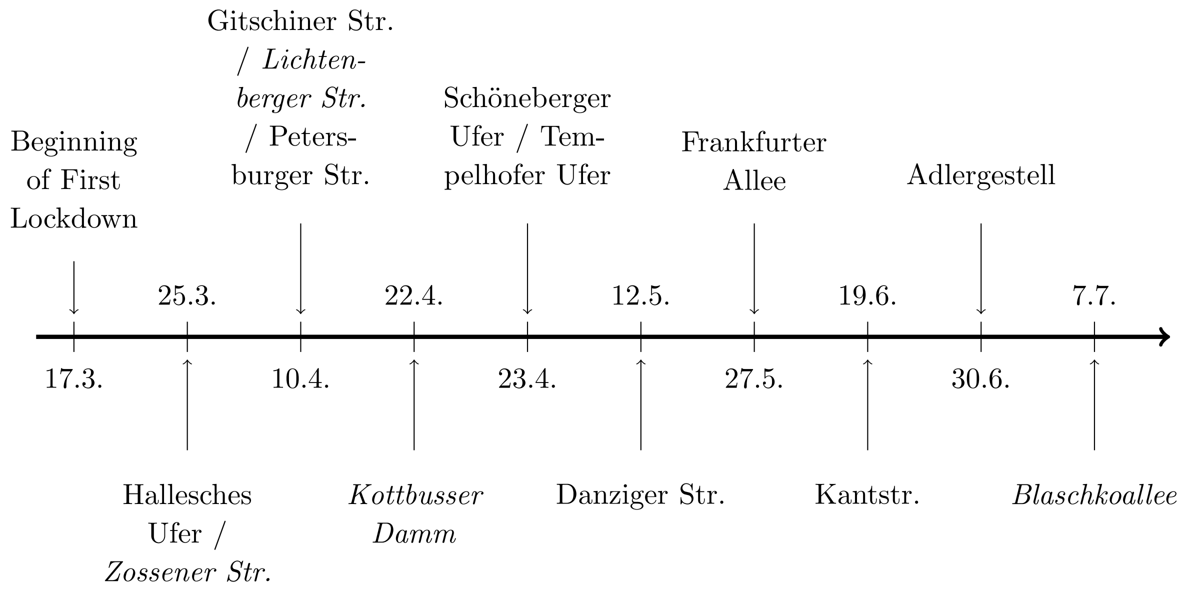


Notes: Own calculation. The graph shows the development of average vehicle volume and average vehicle speed in Berlin from the beginning until the end of the sampling period. The beginning of the COVID-19 induced lockdown on March 17th, 2020 is depicted by the vertical blue, dashed line.

Pop-up bike lanes during the COVID-19 lockdown in Berlin In March of 2020, political measures in form of a lockdown were taken in Germany as a consequence of the COVID-19 pandemic. The lockdown included, among other things, the closure of all stores[11], schools, day care centres for young children, and restaurants. This had an immediate impact on the dynamics on the streets throughout Berlin. Figure 5.1 shows the development of car volume and average speed in the city from 2019 to 2020 with an apparent negative correlation between the total number of cars and speed. Just after the lockdown started on March 17th 2020, there was a sudden decline of overall traffic and congestion as suggested by an increase in average speed. This circumstance was taken advantage of by political decision makers. They justified the establishment of PUBLs in Berlin on the one hand with the fact, that there would be less traffic disturbances due to the reduced automobile

[11]Except for those necessary for daily life, like grocery stores.

Figure 5.2: Time line of PUBL installations



Notes: Own illustration. The figure shows all installations of pop up lanes after the beginning of the COVID-19 induced lockdown in March, 2020. All street names in italic font have no observations in the traffic data set and are therefore not considered in the respective analyses of the effect on volume and speed.

volume and, on the other hand, with the requirement of providing alternatives to public transport that are less risky in terms of a COVID-19 infection (Bezirksamt Friedrichshain-Kreuzberg, 2020a). According to statements by the authorities, the political will to enforce such infrastructural renewals differed between districts. Foremost the local authorities of Friedrichshain-Kreuzberg (FHKB)[12], governed by the Green Party, began setting up PUBLs in their local district.[13] PUBLs are, contrarily to regular bicycle lanes, created spontaneously, circumventing the otherwise relatively long-lasting decision process of where and how to build a bicycle lane. While the implementation of regular bike lanes takes two to ten years, PUBLs are implemented in three to 10 days (Bezirksamt Friedrichshain-Kreuzberg, 2020b). The reduced time necessary for the implementation is also due to the very simple and cheap type of construction of the lanes, which only consist of paint and temporary bollards. Two days after the lockdown came into force, the first PUBL was installed on March 25th in FHKB.[14] Others followed subsequently in different parts of Berlin. Overall, there is a total of 13 affected streets where PUBLs were installed or later extended on nine different dates

[12]Four out of eleven PUBLs were not set up in FHKB, but in the districts Charlottenburg-Wilmersdorf, Neukölln, Pankow and Treptow-Köpenick.

[13]The following information about the setup of PUBLs were gathered through interviews with the authorities of the Berlin Senate, conducted mainly in May 2021.

[14]The segment on Hallesches Ufer into one direction.

between March 25th and July 7th.[15] Almost all streets had bike lanes set up into both directions and mostly of similar length.[16] Figure 5.2 shows a time-line of the exact dates of installation as well as the respective streets. By mid 2021, most of the bike lanes had been made permanent by replacing the movable bollards with fixed demarcations. Until the end of the observation period, there has been no further change in treatment status.

Criteria of installation Following Kraus and Koch (2021) and personal interviews with local authorities, the placement and timing of pop-up bike lanes in Berlin was, conditional on certain characteristics regarding the affected streets, as good as random. The decision on where to locate such a lane was primarily driven by the amount of available street space. Only streets with at least two car lanes were taken into consideration, such that car traffic was not blocked completely on those roads.[17] This characteristic of a minimum lane number was the only real requirement for installing a PUBL.[18] In order to account for the restriction regarding randomness of placement, the main estimation sample will only contain streets with two or more lanes. Only in some robustness checks this sample requirement will be altered. Station fixed effects additionally control for the number of lanes implicitly. The timing of installation was mostly influenced by the availability of construction firms. Due to the fact that those were not instantaneously available for setting up all of the bollards at once, it took some time until all PUBLs had been placed. In addition, anticipating the first PUBL was not possible since there was no initial press release on the project until the day of the first installation (see Bezirksamt Friedrichshain-Kreuzberg, 2020a). The quasi-experimental setup with random timing and placement addresses standard drawbacks when measuring the effects of such policy interventions like reverse causality (e.g. when bike lanes are a reaction to lower demand for cars and increased demand for cycling) or omitted variable bias (e.g. local preferences for more cycling lanes in the city). In Section 5.4, I will go into

[15]Three of the roads cannot be taken into account in parts of the analysis, as they happened on roads, which do not have any traffic volume/speed measuring station nearby. One of those roads is "Blaschkoallee" in Neukölln. This was the last PUBL to be established. Thus, the last date with a change in treatment status in my sample is June 30th, when the street "Adlergestell" in Treptow-Köpenick received a PUBL.

[16]Compare Mobycon (2020) for more information about the implementation process in Berlin.

[17]Effectively, treated streets are a small subset of roads, which were supposed to be equipped with cycling infrastructure in the future as planned in the aforementioned mobility law (compare the *Political premises* paragraph).

[18]One characteristic of treated streets, which is observed in the data, is that most of them had no prior bicycle infrastructure. However, this is not true for all PUBL streets. This specific type of sample selection will be tackled in a robustness check in the results section.

more detail and address potential threats regarding the identification strategy.

Theoretical considerations Generally, the COVID-19 era led to a change in habits due to people diverting from different forms of public transportation, like buses or metropolitan railways, to private means of transport, like cars or bicycles (Tirachini and Cats, 2020). Some European cities experienced declines in public transport ridership of over 90% following lockdowns (Vitrano, 2021). Such disruptive shocks to public transport have been found to increase car utilization and thereby congestion (Bauernschuster et al., 2017; Anderson, 2014). However, since concerns about getting infected with COVID-19 should be evenly distributed among the population within the city, these effects should be felt on all streets throughout a city simultaneously. Had PUBL streets not experienced a structural change, then the lockdown effects should have been the same compared to similar untreated streets. The question remains which explicit effects are then to be expected by the new cycling infrastructure.

The replacement of car lanes with cycle lanes means a loss of space for cars and a gain of space for bikes. Following Duranton and Turner (2011), car traffic increases with the extent of the availability of streets. This means that congestion is unit-elastic with respect to street space. The authors identified the creation of traffic as a main channel and the diversion from other streets as a less important one. In general, this elasticity should apply for both, the creation of new lane kilometres as well as when lane kilometres for cars are reduced. However, in the case of PUBLs in Berlin, the change of car space provision was accompanied by the creation of space for bicycles. Making this alternative way of commuting more attractive may affect the elasticity. Besides, Duranton and Turner (2011) consider long-term developments between cities and regions, while I look at a shorter time horizon within one city.

In terms of volume, there are two potential effects I would expect in the short run. On the one hand, PUBL streets may experience a decline in car volume compared to non-PUBL streets. In order to avoid traffic jams and congestion, there is the incentive to use alternative roads in the surrounding area, which did not experience the installation of a PUBL. This would mean a reduction of volume on the respective street. On the other hand, there may be no will to divert from the accustomed old route, which would be in line with time-inconsistent preferences or a status quo bias (Mattauch et al., 2016). As a consequence, volume would not be affected. In the long run, there also may be behavioural adjustments, e.g. people adapting their preferred route over time and thereby reducing volume on treated streets further.

Volume also directly affects average speed. If there is an increase in volume without street space being changed, then this should lead to more congestion and consequently a decrease in average speed. The reduction of a car lane without a change of car volume should also lead to an increase in congestion as the same number of cars now shares a lower number of lanes. Another possibility is that people adjust the timing of their commute. Thus, the average speed may not change even though the daily/weekly traffic volume changes, simply because people are travelling at different times of the day.

Based on these considerations, I would expect vehicle volume on PUBL streets to decrease slightly or remain unchanged compared to non-treated roads, since car drivers may not be willing to adapt quickly to the new situation, as suggested by the status-quo bias. Consequently, due to the reduction of space for cars, I assume average speed to decrease and thereby congestion to increase, especially if volume is only slightly affected.

Additionally, I want to find out in how far the restructuring of road space towards a more bike-friendly environment affects road safety by looking at the development of accidents. As mentioned before, this was one of the major reasons to create such bike lanes in the first place. If car and bike travel are on separate lanes, which was not given prior to the policy change, collisions between these two modes of travel should decrease. This would mean an overall decrease in the incidence of accidents. In contrast, the separation of lanes could also lead to more accidents if motorists now pay less attention to cyclists and overlook them when taking a turn at an intersection (Summala et al., 1996). Furthermore, pop-up lanes may nudge new cyclists towards using the newly created ways. This could lead to an increase in accidents, especially between cyclists as found by Li et al. (2017). If there are more cyclists on the respective roads, while the total number of accidents is not affected due to sufficient space for cyclists, this would mean a decline in the rate of cycling accidents. Overall, the direction of the effect on accidents cannot be predicted clearly.

5.3 Data and Descriptive Statistics

Traffic measuring stations The outcome measures for traffic volume and average speed stem from 772 measuring stations throughout the city of Berlin. The recordings show hourly values of speed and volume of vehicles passing by one station for the years 2019-2020.[19] This allows to detect within day variations as well as short-term developments of traffic within Berlin. Additionally, it is possible to differentiate between personal cars and lorries. Due to the fact that all PUBLs in my sample were set up between end of March and end of June of 2020[20], I have a post-treatment period of about 40 weeks. The pre-treatment period contains about 70 weeks. Most measuring stations cover the traffic on multiple lanes of which I build the mean over all lanes for speed and for volume.[21] The information about the number of lanes is furthermore used for specifying the control group in most analyses. It allows to compare streets of the same pre-treatment capacity with each other. Figure 5.3 shows a map of Berlin, depicting the network of measuring stations and the PUBLs.

Of the eleven streets, which had a PUBL installation, some lack a traffic measuring station capturing volume and speed. Thus, I am only able to analyse a subset of affected streets regarding this outcome, ignoring the effects on five out of eleven treated streets.[22] Since all of the non-considered streets have similar characteristics to the ones in the sample and there is no correlation of the placement of measuring stations and the creation of PUBLs, I do not consider this to be a major drawback. Overall, I observe 23 measuring stations located at treated streets with two or three lanes.

For the analyses, I exclude all measuring stations, which are situated at highways.[23] The reason is that I assume that inner-city traffic might have developed differently from traffic on the highway, which circles the city. This is because highway traffic may not be substituted as easily by cycling or different forms of public transportation. Furthermore, the daily time

[19] The Senate of Berlin also provided me with (incomplete) data for May-July in 2021. However, due to the fact that many measuring stations drop out and due to gaps in the data between 2020 and 2021, I only consider the mid-term outcomes in a subsection.

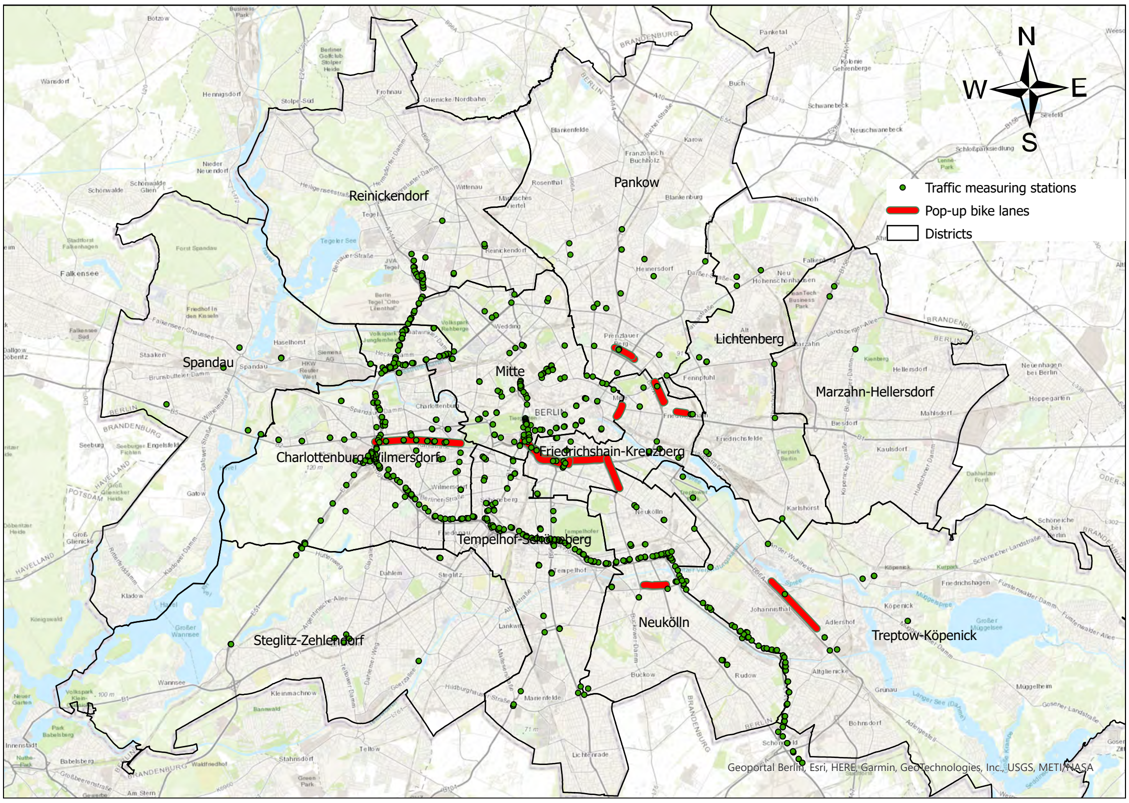
[20] As described in Section 5.2, the last PUBL was installed on July 7th, 2020. However, for this last treated street, there exists no measuring station nearby, such that it cannot be taken into account.

[21] Missing values are ignored in the mean calculation.

[22] Most of these affected streets are dropped due to non-existing measuring stations. One street has missing observations in the post-treatment period (the station in Danziger Str.) and therefore is also deleted from the sample.

[23] In Berlin, these are named: A100, A111, A113, A114, A115.

Figure 5.3: pop-up bike lanes and measuring stations in Berlin

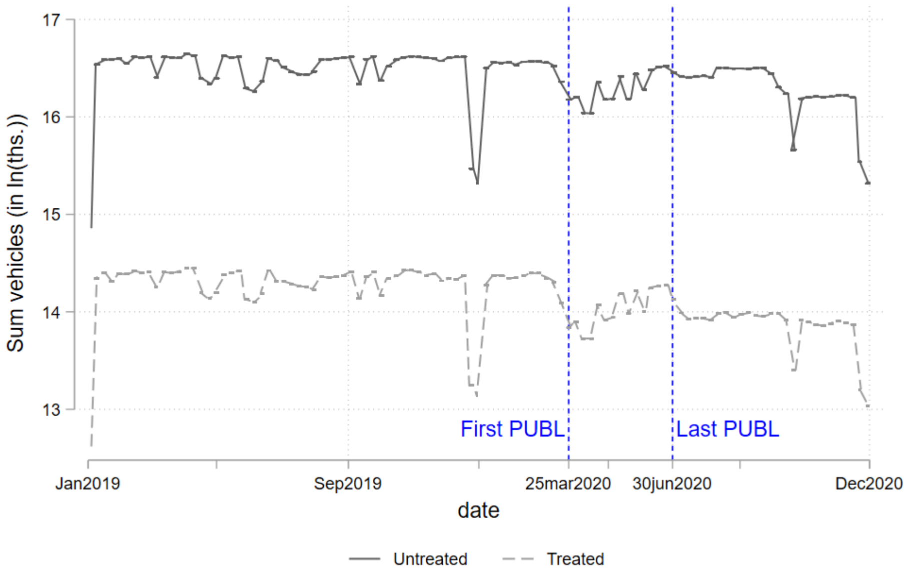


Notes: The figure shows a map of Berlin. Pop-up bike lanes are marked as fat red lines and traffic measuring stations are depicted by green dots. The city is subdivided into 12 districts, which are outlined by fat black lines.

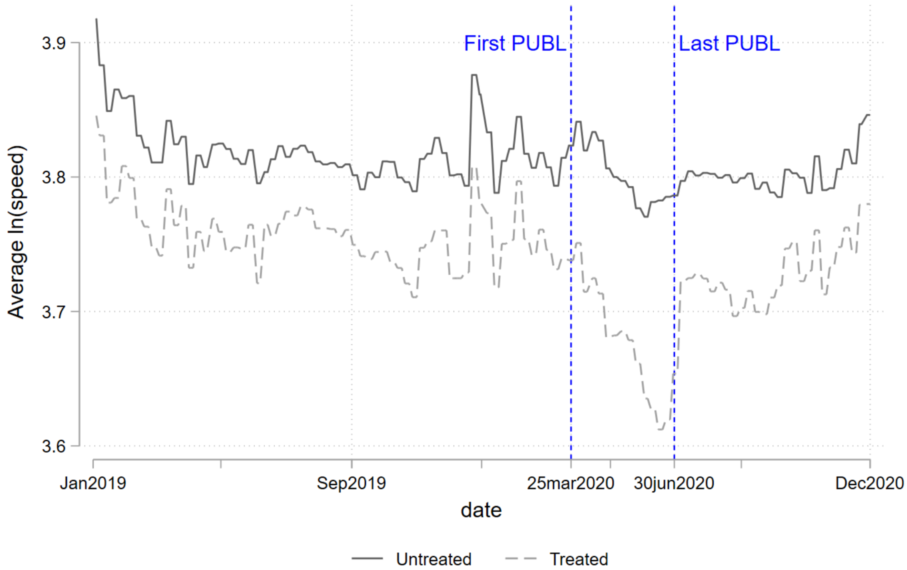
frame is limited from 5 a.m. to 8 p.m. and I exclude Saturdays and Sundays as well as legal holidays from the analyses. I thus concentrate on traffic on workdays and during standard work hours. Some stations are lacking data e.g. for some hours a day or even for entire days.[24] As a consequence, in the event study analyses, which uses data aggregated to weekly values, I restrict the sample to stations, which have observations for at least five hours a day (this results in dropping about 0.04% of all observations) for at least 2 days a week. However, whenever I base the analysis on hourly data, which allows for hour and date fixed effects, I use the entire sample. Considering all these restrictions, the main sample remains with 192 measuring stations in the control group. I address the influence of some of the sample restrictions, like excluding highways from the sample, in the robustness section 5.5.1.4.

[24]There is also partly erroneous data, which I deleted from the sample. Regarding traffic speed for example, there were observations with speed being over 1000. While this is a very extreme case, I cut out observations with speed of either trucks or cars being larger than 100, which is a speed level very unlikely to be reached within city boundaries with its standard speed limit of 50 kilometres per hour. As a result, I deleted ≈ 0.0038 percent of the car speed sample and 0.00015 of the truck speed sample.

Figure 5.4: Development of speed and volume (pooled over cars and trucks) between Treatment and Control group



(a) Volume



(b) Speed

Notes: Figure 5.4a shows the development of vehicles (in logs) from January 2019 until December 2020 by treatment status. Figure 5.4b shows the development of average speed (in logs) by treatment status. In both cases, the grey dashed line represents the development of all treated streets, while the solid line shows the development of untreated streets. The two dashed blue vertical lines in each graph represent the installation of the first and the last pop-up bike lane in the sample. Streets that were treated, but do not contain a traffic measuring station, are not considered here.

Figure 5.4 shows the natural logarithms of total weekly volume and average weekly speed by treatment status. The vertical blue lines mark the first and the last installation of a PUBL in the city. The figures already show to some extent the effect of PUBLs and provide an impression regarding parallel trends between treated and untreated streets prior to the instalment of new cycling infrastructure. As of volume, we can see that the overall development was very stable from 2019 to the beginning of 2020. The lockdown led to a sharp decline of vehicles on the streets of Berlin[25] and had returned to pre-pandemic heights by September of 2020 on untreated streets. Treated streets, however, did not return to pre-pandemic levels with respect to vehicle volume, but remained on a lower level.[26] The large spikes in the graph are the dates around new year's eve and to a lower extent during summer holidays, when traffic is significantly lower all over the city.

While between January of 2019 and March of 2020 overall speed has remained fairly constant with some minor variation on both treated and untreated streets, the lockdown led to two different developments. Treated streets experienced a decrease in speed whereas on untreated streets there initially was a small increase and then a return to normal levels. This is a first indication that treated streets experienced some sort of congestion despite the overall decline of vehicles travelling on them.

In summary, the figures show that spikes in traffic, either downward or upward, affected all streets across the city, even though in parts to different extents. Thus, changes in traffic in general seem to be caused by city-wide events affecting the overall traffic flow.

Accidents Data on accidents stems from the atlas of accidents ("Unfallatlas")[27], which locates every accident in Germany with exact point coordinates. For Berlin, this data exists for the years 2018 to 2020. It is available on a monthly basis and includes information such as the hour of the accident, day of week, whether people were killed or injured, and also which means of transport were involved (bike, pedestrian, car, etc.). It also includes lighting conditions (daylight, dawn, darkness) and road condition (wet, dry, slippery). Due to the fact that I only have monthly recordings, the post-treatment period starts the month after

[25] However, the decline is not as sharp as on new years eve.

[26] The absolute difference in log values stems from the fact that there are much more untreated streets compared to treated ones.

[27] The Unfallatlas is freely downloadable at https://unfallatlas.statistikportal.de/.

the implementation of the PUBLs.

In the accident analyses, I use two different methodological approaches, which require two different levels of spatial aggregation. First, I run an event study design on street level within Berlin. Therefore, I match the accidents to Open Street Maps (OSM) information of Berlin such that every accident is assigned to - if possible - a single street of which I also know whether it contains a PUBL or not. The aim is to have a data set with single streets as units of observations, each of which contains the number of accidents, which occurred there each month.[28] Second, I use the synthetic control group design on municipality level as a more macro-economic approach. For this data set, I take the sum of accidents within a municipality per month for all of Germany between 2018 and 2020.[29] Treated units are then the different Berlin districts in which PUBLs were installed while potential control units are recruited from municipalities in Germany outside of Berlin and for which there is data on the matching variables.[30] These matching variables as well as additional control variables are presented in the following paragraph.[31]

Matching and control variables Since traffic speed and volume are measured locally within a time frame of about two years, most characteristics specific to streets and measuring stations are controlled for by respective fixed effects. Measuring station fixed effects capture e.g. the influence of nearby alternative modes of transport like subways, and date fixed effects account for city-wide shocks on the respective day. However, I am as well able to take into account two time-varying variables on very granular time and spatial scale.

[28]In order to do so, I lay buffers of different magnitudes around the OSM-lanes and successively match those buffers to the accident data. This is necessary since the OSM data contains streets as lines and the accident data contains single coordinates, which hardly ever match exactly. Some accidents cannot be assigned unambiguously to one street, e.g. when the accident happened on a crossing. In these cases, the accident is assigned to both streets. However, more than 90 percent of the accidents (the exact number depends on the year) can be assigned to a single street. In order to combine street segments into entire streets, I combine the OSM data to street segment data as provided by the city of Berlin (Geoportal Berlin, 2021). Otherwise, single streets would be split into several observations.

[29]Observations for three out of 16 states drop out of the data set. The respective states are North Rhine-Westphalia, Mecklenburg-West Pomerania, and Thuringia. The reason is a lack of data for the year 2018.

[30]I exclude municipalities that also received PUBLs in the observation period. This includes municipalities in Munich, Hamburg and Düsseldorf.

[31]In order to lower the computational burden, I limit the control sample to municipalities with less than 25.000 inhabitants and thereby ignore very small and rural regions. However, I assume those not to be comparable to the treated units of interest. Changing the threshold value to 10.000 has no effects on the results.

Firstly, I include georeferenced road construction works.[32] The data contains exact dates when road works took place (mostly a time range of several days or weeks) and is differentiated by status (e.g. approved, finished, in coordination etc.) and limitations to the public caused by them (for example the closure of a traffic lane or even the entire street). I only consider construction works, which are finished, approved or ongoing. In most specifications, I will use a simple dummy variable, which indicates whether some type of road construction took place or not. In additional analyses I will also account for non-binary limitations to traffic and the public.

Secondly, I know about changes in the speed limit regime. Few streets, all of which in the control group, experienced a change in speed limit from a maximum speed of 50 km/h to 30 km/h[33] during the observation period. These changes are mandated by the Berlin Senate and realized by local authorities. Reasons for such measures are noise control, air pollution prevention, or road safety. I account for those changes using a dummy variable switching to one on the date of implementation. There are also temporary changes in the speed limit during a day, e.g. from 6 a.m. until 5 p.m., which mostly happen on streets close to a school or a kindergarten. Controlling for the latter variation in speed limit is only possible when using the entire hourly data-set in the two-way fixed effects estimations. In the case of daily temporary speed limits, a binary variable switching to one in the respective time frame is included in the estimations. Overall, about 12 percent of the sample is affected by speed limit changes over time or throughout the day.

I additionally gather information about traffic measuring stations and their surroundings using OSM and other sources with city-specific data. OSM allows to match the information whether a traffic measuring station lies within a certain radius (I choose a radius of 50 meter) to a tram or rail line. Thereby, it is possible to see whether a street directly "competes" with the rail line. Furthermore, for every measuring station, I match information whether it lies at a bike lane, which existed prior to the installation of a PUBL.[34] While this type of information is already captured by station fixed effects, the data still makes it possible to

[32]This information was provided by the Berlin Senatsverwaltung für Umwelt, Verkehr und Klimaschutz (Senate Office for Environment, Traffic and Climate protection).

[33]This corresponds to a change of about 31 mph to approximately 19 mph.

[34]The bicycle infrastructure data comes from a collection of shapefiles covering different topics in Berlin (https://www.geodaten.tu-berlin.de/menue/downloads/berlin/). I define a station to lie at a cycle lane if it is within a 15 meter reach.

split outcomes by street characteristics. For example, PUBLs were predominantly installed at streets without major cycling infrastructure. I will therefore test whether outcomes are sensitive to varying control units with respect to their bike-friendliness.

In the analyses of accidents, I furthermore make use of administrative annual data for the more than 10.000 municipalities in Germany. This data is used foremost in order to match treatment and control group in the synthetic control group design. I use information about land use patterns (space available for traffic and for settlements), voting behaviour (percentage of Green Party voters, voter turnout), and population. I also retrieve economic data (unemployment rate), as well as data regarding road safety. All this data is publicly available on a website for regional data by the Statistical Office.[35]

Sample adjustments In order to make the control group in my analyses more plausible, I apply some adjustments to the data set for my main estimations. The installation of PUBLs aimed at streets with more than one car lane and was realized on streets with either two or three lanes. This is why in the majority of estimations I exclude one-lane as well as four-lane streets.[36] Furthermore, most treated streets had no prior cycling infrastructure. Thus, in some estimations I will restrict the sample to streets without any sort of cycle lanes prior to the PUBL installation.[37] For the control group, I exclude all measuring stations, which lie within a 1 kilometre radius of a treated unit. This is done in order to account for potential deviation effects of traffic. Thus, if a street is treated, then surrounding streets might be affected as a consequence, because the drivers search for different, now potentially faster routes. In this case, the potential control group would be affected by the treatment itself. In further estimations, I explicitly test for these spillover effects to nearby roads.

[35]https://www.regionalstatistik.de/genesis/online.

[36]Four-lane streets only make up less than 0.3 percent of the overall sample, while about 25 percent of the sample are streets with one lane.

[37]Unfortunately, there was no list of streets (or the senate was not willing to provide me with such a list), which indicated streets eligible for PUBLs but not implementing them. In this case, it would have been possible to take such unselected roads as a control group (compare e.g. Greenstone et al. (2010) for a similar setup).

5.4 Methodology

I aim to identify the causal effect of bike lanes on several outcomes like traffic volume, average traffic speed, and accidents. In my analyses, I exploit the (conditional) random timing and placement of PUBLs in Berlin during the COVID-19 pandemic in 2020. Thus, traffic measuring stations and streets, where new cycling infrastructure was created, are handled as treatment groups, while a large number of other streets of similar size and characteristics are taken into consideration as potential control group. In order to estimate the effects, I resort to various approaches that are versions of the classical Difference-in-Differences (DiD) method. Firstly, I use an event study design to analyse the data aggregated to weekly levels. I primarily use it to test the common trend assumption as well as to identify the development of the outcomes over time in the post-treatment period. Secondly, I conduct two-way fixed effects estimations in order to obtain effects in terms of single coefficients. Most importantly, they allow me to use the entire hourly data set and thereby take into account all relevant control variables. Lastly, I apply the synthetic control group method in my accident analyses as a robustness check.

Event study design To provide evidence that the common trends assumption between untreated and treated units holds, I firstly estimate a flexible event study model, which takes into account the different timing of implementation and the different streets affected (Clarke and Tapia-Schythe, 2021). In order to make this assumption more plausible, I impose a range of sample restrictions as described earlier. For estimation I use the following equation:

$$Y_{it} = \alpha + \sum_{l}^{L} \beta_l (\text{Lead } l)_{it} + \sum_{k}^{K} \gamma_k (\text{Lag } k)_{it} + \mu_i + \delta_t + \mathbf{X}_{it}\phi + \zeta_i(\text{Station} \times \text{LD}) + \epsilon_{it}. \quad (5.1)$$

The outcome variable Y is observed for individual monitoring station or street i at time t (which is either a running week or a running month variable). Station fixed effects are given by μ_i. They control for observable (e.g. public transport stops, topography or the number of lanes) and unobservable factors (e.g. local or political preferences in the area), which are specific to a monitoring station and its surroundings and that do not change over the time frame observed. Time fixed effects, measured by δ_t, account for shocks, which simultaneously affect the whole city, and could potentially influence travel mode and prevalence, e.g.

holidays. The impact of time-varying characteristics $\mathbf{X}$, like construction works, is measured by ϕ, while ζ_i captures the effect of a station-specific lockdown dummy that I control for in the majority of two-way fixed effects specifications. Finally, ϵ_{it} represents an unobserved error term. Leads and Lags in equation 5.1 are dummy variables, representing the number of periods l and k the unit is away from the event.[38] Consequently, the time of opening a PUBL is normalized such that for each case the opening is at $l = 0$. One lead or lag variable is omitted as the baseline difference between treated and untreated units. The maximum number of Leads L (Lags K) included in the regression are then the total number of weeks before (after) the treatment. Streets without the implementation of PUBLs serve as pure control group, such that leads and lags are always zero. These binary variables thus capture the difference between treated and untreated streets in comparison to their difference in the base period, which by definition is zero. Without a significant difference between treatment and control group prior to the base period, the common trend assumption in the respective time frame most likely holds. The implicit assumption here is that without treatment, treated and untreated streets would have maintained differences just like in the base line period. The main advantage compared to a standard two way fixed effects model is that rather than relying on a single coefficient for post-treatment, this model allows to capture the development of treatment effects over time via the lag coefficients and to inspect the common trend assumption.

Two-way fixed effects The main analyses are then conducted with a standard Two-Way Fixed Effects (TWFE) model in which the lags and leads of Equation 5.1[39] are replaced by $\beta \text{PostTreatment}_{it}$, where $\text{PostTreatment}_{it} = \mathbf{1}\left[t \geq \text{Treatment}_i\right]$. In the estimation, all never treated measuring stations have this treatment indicator always set to zero, while it switches to one for PUBL units after the beginning of treatment. This estimation provides me with a single treatment effect pooled over all treated streets. The advantage of using the TWFE model is that it allows me to use the entire data set and therefore to control for date and hour fixed effects. It is now furthermore possible to specifically control for temporary speed limit zones, which are in place e.g. between 6 a.m. and 5 p.m. on certain streets. Using the hourly data is not possible in the event study design since it does not allow for time gaps

[38] Thus, $(\text{Lead } l)_{it} = \mathbf{1}\left[t = \text{Event}_i - l\right]$ for $l \in \{1, ..., L-1\}$, and $(\text{Lag } k)_{it} = \mathbf{1}\left[t = \text{Event}_i + k\right]$ for $k \in \{1, ..., K-1\}$.

[39] $\sum_l^L \beta_l (\text{Lead } l)_{it} + \sum_k^K \gamma_k (\text{Lag } k)_{it}$.

in the observations. This is why I aggregate observations to weekly levels in event study estimates, which solves the problem of gaps.[40]

Synthetic control group design When analysing the effects of PUBLs on accidents, I will furthermore apply the synthetic control group method (Abadie et al., 2010). Due to the fact that I have geolocated accidents for all of Germany, this allows me to use a macro-perspective, comparing treated districts of Berlin[41] with similar administrative units all over the country.[42] The advantage of using the synthetic controls method is that it does not require to explicitly choose the control group, but that it uses a data-driven matching approach. Thus, it determines a weighted combination of those control units in the donor pool (in my case all German municipalities outside of Berlin) that are closest to the treated unit in terms of characteristics and pre-treatment developments of the outcome variable. I use a variety of municipality-specific observables like population, unemployment rate, or the share of land devoted to traffic infrastructure as matching variables to find a data-driven control group.

Main threat to identification Some aspects might influence the assumption of random assignment, which I want to specifically tackle in my analyses. In the following, I describe the potential problems and how I finally aim to solve them.

First, there is the concern of non-random selection of streets by the responsible authorities for installing a PUBL in the first place. If bike lanes were randomly assigned to any street in the city, then selection bias would not pose a problem. However, if streets are chosen based on their characteristics, e.g. that only streets with minor car traffic are chosen, then the estimated treatment effect will be biased. The choice where to locate a PUBL was based on Berlin-wide plans for extending the cycling infrastructure prior to the pandemic. Even though the plans existed for the whole city, all of the PUBLs were created in only a subset of districts[43]. In Section 5.2, I already argue in how far the allocation of PUBLs was random.

[40] In the robustness section I will address potential problems regarding standard two-way fixed effects.

[41] With respect to Berlin, I use the terms "district" and "municipality" interchangeably. However, the main comparison group in the synthetic control group method is composed of municipalities in Germany.

[42] The measures for car volume and speed are only available for Berlin. Using the synthetic control group design on such a local scale is rather problematic due to the lack of observable control variables on street level, which are required for the matching procedure of treatment and synthetic control group. This is why the use of this method is restricted to the accident analyses.

[43] Charlottenburg, Friedrichshain-Kreuzberg, Neukölln, Treptow-Köpenick, and Pankow.

However, some local district governments were more supportive in establishing PUBLs than others. This may raise the concern that there exist systematic differences between these and other districts in the city.

Second, in general the district-specific differences should be captured in the estimations by the measuring station fixed effects. Nonetheless, shocks like the lockdown after the beginning of the COVID pandemic may lead to different behavioural adaptations in districts with PUBLs compared to districts without, e.g. with respect to commuting choices. A potential reason are political preferences and attitudes in certain areas, which may translate to differential behavioural adaptations regarding home-office or the utilization of public transportation. For example, if a district is populated with more blue-collar than white-collar workers, then home-office might be less of an option there compared to other areas of the city with a differently composed workforce. This may systematically bias the results.

Third, some of the streets lie close to a subway line, while others are further away. As utilization of public transport has significantly changed during the Corona-crisis, those streets might have been affected differently to streets further away from public transport.

The variety of fixed effects included in the estimations, like station fixed effects, should account for general local political preferences that drive local authorities to be more supportive of installing PUBLs. In order to control for potential differential developments after the start of the lockdown and address the other aforementioned concerns, I add an interaction term to the estimation, which accounts for *lockdown* $\times$ *measuring station* effects.[44] This interaction term captures effects, which are present on a very local scale (station/street-level) after the beginning of the lockdown. Thus, it captures e.g. differential developments on streets close to public transport compared to streets further away from it. At the same time, it also subsumes district-specific changes due to local commuting preferences. Due to the fact that some PUBLs were installed right after the beginning of the lockdown, the interaction term may capture away parts of the actual pop-up lane effect. This is why I consider estimates, which control for the interaction term, as lower-bound outcomes.

[44]Lockdown then is a dummy variable and it is defined as the time after March 22nd, 2020. The reason being is that measures were lifted from time to time and partly reinstated again. Thus, there was no clear-cut end of the lockdown. In autumn/winter of 2020, measures became increasingly strict again, resulting in another so-called “hard lockdown” in December of 2020.

In all estimations, I use standard errors clustered on a time and a spatial dimension. The former is a running week variable, while the latter consists of 1km $\times$ 1km grid cells spanning the entire city.[45] In this manner, I tackle concerns that treatment may be spatially or temporally correlated, and therefore account for spillover effects. In alternative specifications in the robustness section, I will also test for different clusters.

5.5 Results

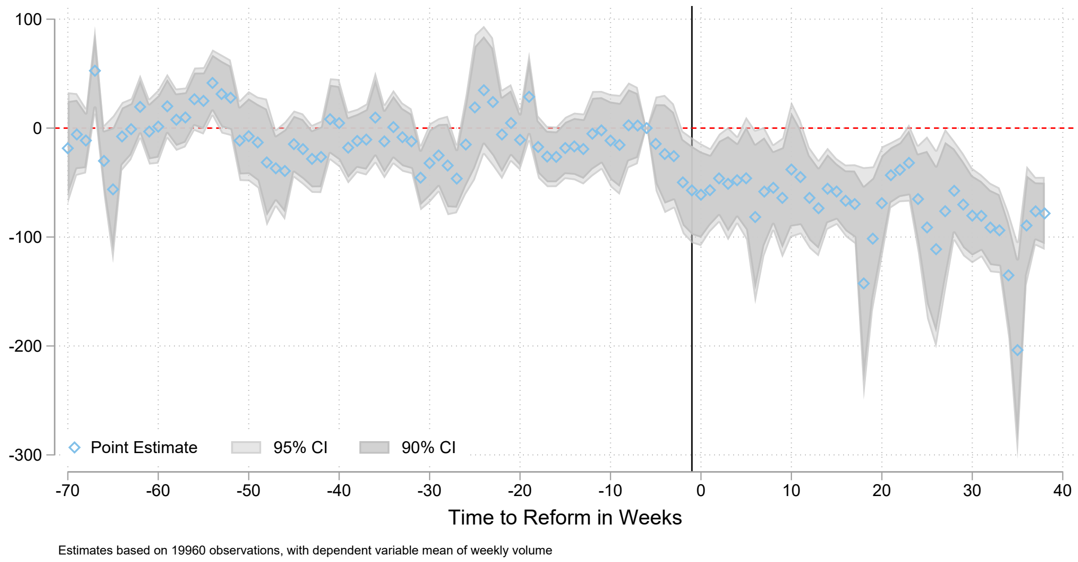
5.5.1 Volume and speed

Figure 5.5 shows the effects of the introduction of a PUBL on traffic volume and on average speed by a graphical representation of an event study design. Blue dots show the main estimation coefficients of leads and lags, while Confidence Interval (CI) are depicted as area shades in light (95% CI) and dark grey (90% CI) respectively. No additional control variables apart from station and week fixed effects are added in these first estimations in order to show the "pure" common trend. Adding construction work and speed limit changes does not change the picture though.[46] Both graphs generally show that the common trend assumption is satisfied, however with some minor drawbacks. Considering average vehicle speed as outcome variable (Panel (b)) suggests a very stable parallel trend between treated and untreated streets prior to the PUBL installations. Only a very small number of observations about one year or more prior to treatment show marginally significant deviations in this case. Regarding the number of vehicles as outcome, more, but still very few, occurrences show significant positive as well as negative differences in the pre-event time frame. However, they do not change the overall picture suggesting common trends between treated and untreated units prior to the lockdown. Additionally, one has to bear in mind that the regressions run here are based on weekly averages of speed and volume, and thereby do not take into account date or hour fixed effects, which may further adjust for unobserved

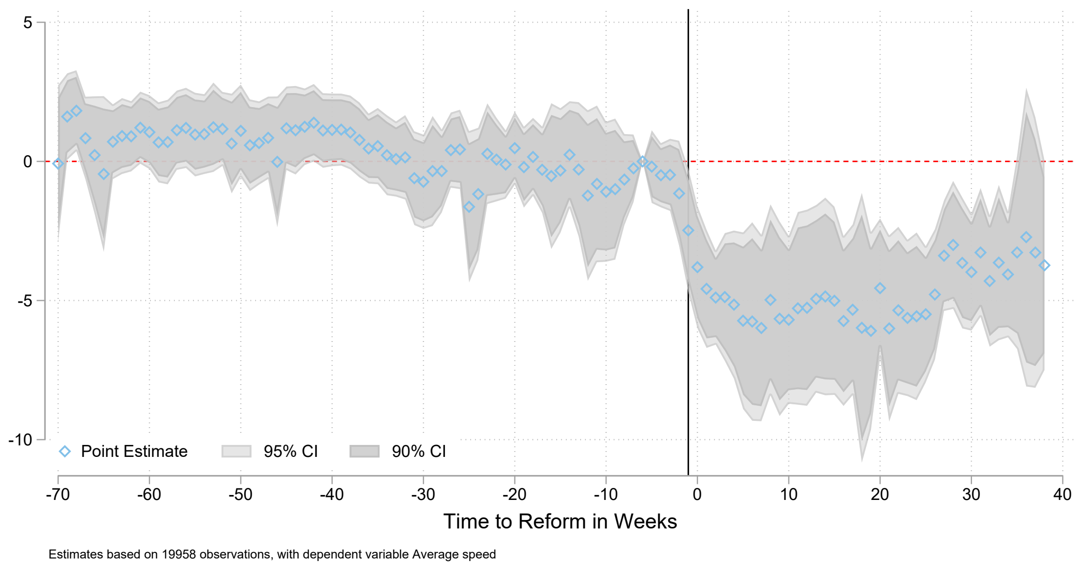
[45] I use the INSPIRE grid for Germany, which is publicly available at (https://gdz.bkg.bund.de/index.php/default/inspire/sonstige-inspire-themen/geographische-gitter-fur-deutschland-in-lambert-projektion-geogitter-inspire.html).

[46] The base period here is chosen to be at five weeks prior to treatment, which is the average number of weeks between the start of the lockdown and the installation of a PUBL. Whenever the *station* $\times$ *post-lockdown* interaction effect is included in the estimation, the base period is at *lead = 1*.

Figure 5.5: Event study outcomes for traffic volume and speed



(a) Number of vehicles



(b) Average vehicle speed

Note: Own calculations. The two graphs show the results of two separate event study estimations. Outcome variable Y in Figure 5.5a is the absolute number of vehicles while it is average vehicle speed in Figure 5.5b. Blue dots represent the main estimation coefficients of leads and lags. Confidence intervals (CI) are depicted as area shades in light (95% CI) and dark grey (90% CI). The vertical solid black line shows the time of treatment, which is anchored at 0. Leads and lags are the time before and after treatment in weeks. The estimations include station and week fixed effects. The sample is restricted to streets with two or three lanes and all streets within a radius of 1km to a treated street are excluded from the estimations. Standard errors are clustered at 1km × 1km grid cell level spanning the city times a running week variable.

differences between treated and untreated streets.[47] The main outcomes presented in this paper moreover include an interaction term, which takes into account station specific developments after the implementation of a lockdown. Figure 5.A.1 shows the event study results considering the interaction effect. It again exhibits the validity of the parallel trend assumption, especially regarding average speed as outcome.

Traffic volume The PUBL introduction exhibits a relatively small, but significant effect on the number of vehicles compared to the development on control streets when ignoring the *station* × *post-lockdown* interaction. The effect is rather modest with a few gaps to the downside and is relatively stable in size. In terms of absolute effect size, about 50-100 vehicles per week less are observed on average on treated streets compared to untreated ones. A negative effect would mean that some drivers do not want to keep using the old way and rather circumvent these high-traffic areas. The pre-trend assumption holds for the vast majority of the pre-treatment weeks, even though the few significant differences may be considered a potential backdrop in the causal interpretation of results. Including the interaction term, as was done in Figure 5.A.1a, causes the PUBL effect on volume to become insignificant. This means that it is not possible to attribute the volume effect directly to the installation of a PUBL, but that the COVID-19 lockdown lead to a slightly differential development on treated streets compared to untreated ones.

Traffic speed Average speed of vehicles significantly decreased directly after the PUBLs were installed, which implies an increase in congestion. In absolute terms, average speed decreased by about 4-5 kilometres per hour (km/h)[48], which is about 10% of the maximum speed allowed on these streets.[49] This result is not surprising due to the fact that car utilization behaviour hardly changed in the beginning, while the number of lanes was reduced (partly from two lanes to one). The effect size is very stable over time, which is certified after including the interaction term as presented in Figure 5.A.1b. This suggests that within the post-treatment time frame observed, there were no major behavioural adaptations of drivers.

[47]In order to run the event study design, the data requires to be balanced, which is why all event study estimates stem from data aggregated to a weekly level.

[48]This corresponds to about 2.5-3.1 miles per hour.

[49]In Germany, the standard speed limit within cities is 50 km/h (or 31 mph) with some exceptions.

Single-coefficient model While the graphical analysis provides an insight into the development of the effect over time and allows to track and reaffirm the parallel trend assumption, I also want to translate these effects into one overall estimate. The advantage of one single coefficient is that it pins down the results to an overall outcome. As discussed before, the effect on the outcome is relatively stable over time, which eases the concern that two-way fixed effects regressions may conceal important heterogeneities.[50] Moreover, I am now able to use the entire data set and therefore to control for date and hour fixed effects in all estimations. This means that these estimations additionally control for intra-day commuting patterns and city-wide shocks, which may have occurred on single days. For example, the few downward spikes, as observed in Figure 5.A.1a, could be attributable to such daily shocks affecting treated and untreated streets differently to some extent. Additionally, the full data set allows to account for temporary speed limit changes throughout the day. While the event study design presented in Figure 5.5 only includes a subset of control variables in order to show the "raw" common trend, I now control in most specifications for the interaction term between a post-lockdown dummy variable and a measuring station indicator. As discussed in Section 5.4, this is to account for different behavioural adaptations at the local level after the COVID-19 lockdown went into effect.

Point estimates of these single-coefficient models are shown in Table 5.1. It includes results for outcomes in absolute values (*Panel A*) as well as for logged values (*Panel B*). All even columns include the *station* × *post-lockdown* interaction term, while uneven columns do not. As before in the event study with heterogeneous timing, I find a significantly negative effect on speed, and thus more congestion for cars. PUBLs led to an overall decrease of car speed of about 11 to 12 percent as *Panel B* shows. This result holds after controlling for the interaction term. In terms of volume, there is a significant decline of vehicles without the interaction term, which turns insignificant as soon as the interaction term is included in the regression. This hints towards a small decline of vehicles on treated streets compared to untreated ones, which cannot clearly be attributed to the installation of PUBLs. However, since the cycle lanes were installed only shortly after the lockdown went into effect, it is in general hard to completely disentangle these two.

[50] In the robustness section I will account for more concerns about heterogeneous effects in terms of time and treatment unit.

Table 5.1: Effect on Volume and Speed

	Volume		Speed	
	(1)	(2)	(3)	(4)
Panel A: Outcome in absolute values				
1(PU lane)	-53.57***	-7.476	-4.512***	-4.123***
	(4.876)	(7.218)	(0.306)	(0.536)
N	1546526	1546526	1543453	1543453
R^2	0.743	0.755	0.757	0.772
Stations	215	215	215	215
Interaction	No	Yes	No	Yes
Panel B: Log-transformed outcome				
1(PU lane)	-0.0228***	-0.00540	-0.113***	-0.122***
	(0.00724)	(0.0126)	(0.00878)	(0.0144)
N	1543493	1543493	1543453	1543453
R^2	0.628	0.650	0.585	0.607
Stations	215	215	215	215
Interaction	No	Yes	No	Yes

Note: Own calculations. The table presents the coefficients of the treatment effects of separate two-way fixed effects estimations with vehicle volume and vehicle speed as dependent variables. Panel A shows the coefficients of interest with outcomes in absolute terms. Panel B shows the same for logged outcome variables. Even columns include an interaction term between a unique measuring station identifier and a post-lockdown dummy variable while uneven columns do not. All estimations include station, date, and hour fixed effects, a dummy whether construction work takes place, and an indicator for a change in speed limits. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.1 shows outcomes of TWFE estimations using the weekly data set, which was used in the event study model. It shows that coefficients are very similar or even slightly higher compared to the regressions based on hourly data, and that inference is hardly different. In the following, I will mainly show outcomes of single-coefficient models using hourly data in order to be able to account for the full set of controls.

5.5.1.1 Heterogeneity

Next, I will analyse in how far outcomes differ by treated unit, by time of the day, between type of vehicles (cars and lorries), and whether there are differences between different street sizes.

Treated unit Table 5.B.2 presents leave-one-out analyses. This means that I conduct the main analysis repeatedly, always omitting a different treated measuring station. Such analyses may reveal whether results are driven by single streets or stations in the sample. All estimations include the entire set of control variables as well as the *station* $\times$ *post-lockdown* interaction term. With respect to average speed, the table shows that the overall picture holds. Thus, congestion increases significantly over all specifications. However, one treated station appears to be out of line in terms of effect size. While the majority of coefficients lie between .12 and .13, leaving out "Station 15" in the table results in a coefficient of about .09. Hence, this specific station has a significant larger effect on the size of the overall result than others. The street where this measuring station is located is *Kantstrasse* in the western district of Berlin Charlottenburg. It lies very close to a nearby highway circling the city. A possible explanation for the difference in effect size is that this specific segment was reduced to a one-lane street. Narrowing down the space for cars to a single lane may therefore have a larger effect on speed than limiting a three-lane street to a two-lane one.[51] Looking at volume, there is no single station that drives the results. All estimates remain negative and insignificant. In the following, I will present results with and without this outlier station if required.

Time of day Due to the very granular nature of the data, I am able to differentiate traffic volume and speed by hour of the day. I am thus able to analyse whether the reduction by one street lane mostly affects commuters in traffic peak hours or also other types of trips.[52] In order to do so, I run two types of estimations: i) I clear the sample from all hourly observations except for the peak hours and ii) I estimate the coefficients for each hour of the day separately. In both cases, the estimations include the full set of control variables.[53] Results for the first approach are presented in Table 5.B.3. Point estimates correspond to the ones in the main outcomes. This means that in the main traffic peak hours speed and volume on average do not seem to be differently affected compared to other times of the day. The second approach enables an even closer look into intra-day variations. To illustrate those, coefficients of single-coefficient two-way fixed effects regressions for each hour of the day are depicted in Figures 5.A.2 and 5.A.3. Tables 5.B.4 and 5.B.5 additionally present the exact

[51] Another potential explanation would have been a change in speed regulation, e.g. from 50 to 30. Even though this specific street was affected by such a regulation change in 2018, there was no such change on any treated street during the sample period of this analysis.

[52] Traffic peak hours are defined as the time frame between 6 a.m. and 9 a.m. in the morning as well as 4 p.m. and 7 p.m. in the evening.

[53] This includes date fixed effects, measuring station fixed effects, a construction work dummy, and the station times lockdown-dummy interaction.

coefficients. Unlike the main analyses, which are restricted to daytime, I now estimate the model for all 24 hours of the day. Figure 5.A.2 reveals that the difference between treated and untreated units is significant throughout the day. The effect size, however, varies. In the night hours (midnight until 5 a.m.), speed is significantly slower by about five to seven percent. From 6 a.m. onward, when traffic usually starts to pick up, the effect size gradually increases. The maximum difference is reached in the afternoon with a point estimate of about $-.16$. Thus, average car speed is about 16 percent lower to similar streets compared to times before treatment at the main traffic peak time of the day. Until the evening, the effect size gradually decreases. As for the volume of vehicles, it is significantly higher on PUBL streets compared to non-PUBL streets at night hours, but not during the day, considering outcomes in log terms. Absolute values as presented in Figure 5.A.3, however, suggest that this difference at night is driven by a very small absolute effect of about 20 vehicles. Due to the significantly lower number of cars on the streets at night hours, the effects at that time of the day should be handled cautiously. Considering speed, traffic is relatively free-flowing at night, which explains the smaller effect size.

Cars and Trucks In the data it is possible to differentiate traffic volume and speed by type of vehicles. I thus analyse whether there are differences in the outcome when differentiating between cars and trucks.[54] Table 5.B.6 shows that the entire result is driven by cars. A look into the data reveals that only about six percent of vehicles observed on the streets in the unrestricted sample are trucks. As a consequence, also the variation of observed trucks is much lower compared to cars. This reduced variation may therefore lack power to detect significant results.

Street size Apart from differences analysed so far, there may exist heterogeneities regarding different street sizes. For example, streets with two lanes prior to the instalment of a PUBL may show different results compared to those with three lanes ahead of treatment because they may be harder or easier to substitute for commuters. Table 5.B.7 shows results for two- and three-lane streets separately, restricting also the control group to streets with the respective number of lanes. Apparently, volume is not differently affected on the different types of streets. The effect size for average speed, however, shows slightly different

[54]Trucks are defined as vehicles longer than 7.5 meters. This subsumes buses and the majority of delivery trucks.

results. Smaller streets are affected more severely compared to larger ones. From a policy perspective, this speaks for the installation of new cycle lanes on larger streets, while it may be advisable to spare roads with less space. However, the difference of the effect sizes is only marginally significant. Taking all aspects into account, heterogeneities of results are, if existent, rather small.

5.5.1.2 Spillover effects to surrounding streets

Results so far suggest that congestion on treated streets increased while traffic moderately declined - even though the decrease cannot be unambiguously attributed to the installation of a PUBL. Thus, even though there are in total modestly less vehicles on the streets, the remaining ones are slower due to the reduction in space for motorized traffic. Since the utilization of public transport significantly declined during the lockdown, this might have put additional pressure on the streets by people changing their preferred mode of transport from train to car or bike. These people might want to avoid congested streets by choosing alternative roads close-by. Thus, commuters leaving their old accustomed routes as well as new car users might divert to roads in the vicinity of PUBL streets. In order to test this hypothesis, I repeat the main analysis with measuring stations, which lie within different distance ranges of a PUBL. Thereby, I assign the respective starting date of the treatment to each nearby station and delete the actually treated streets from the sample.[55] Table 5.B.8 shows the corresponding estimates. While speed and therefore congestion seems not to be affected on the surroundings of treated streets, there is a significant increase of car volume on streets within a 750 meter and a 1 kilometre radius of PUBL streets. One would expect the effect size to decrease, the larger the radius drawn to the PUBL. The table suggests otherwise with the 1km-coefficient being larger than the 750 meter one. However, the difference in effect size is not significant, which means that streets up to 1km away from a PUBL are equally affected. Moving further away then shows the expected development with the coefficient size tending to zero and being insignificant.[56] This result suggests that drivers are actually nudged away to some extent from their accustomed routes. An alternative interpretation is that former public transit users, who then change to commuting by car, abstain from using

[55] Since I want to know about the effect on all types of lanes, I abstain from the restriction of limiting the sample to streets with a specific number of lanes. This explains the larger overall sample size.

[56] Note that treated stations between the distances are not mutually exclusive. This means that all stations that are handled as treated in the 750m regressions are also treated in the 1km regressions.

a relatively congested PUBL street and rather use roads close-by. In any case, this has no negative affect on the surrounding streets in terms of congestion. Rather, a new equilibrium with a more equal distribution of traffic seems to be established.

5.5.1.3 Medium run effects

Additionally to the data set running until the end of 2020, the Berlin Senate also provided me with additional traffic data for the time period between March and May of 2021. The reason not to include the entire time frame into the main estimations is that many measuring stations drop out of the observation network in 2021, and therefore the sample becomes less balanced. Additionally, there are more than three months missing between my observation period and the data from 2021, which does not allow to track potential developments in between. Results for the entire sample as well as for the sample without the outlier from the leave-one-out analysis are presented in Table 5.B.9. All coefficients slightly increase in size. The volume effect now becomes significant and larger, while the speed effect is only marginally larger, and does not significantly change. Thus, about one year after the installation of the PUBL, drivers are apparently nudged away to a larger extent from treated streets, which are still more congested than control streets. The combination of less volume combined with similar congestion may be explained by an overall city-wide increase in motorized traffic, which more evenly distributed on the street network compared to pre-treatment times. One reason for the overall traffic increase potentially is a larger share of workers returning to offices rather than working at home. Those now re-entering the streets have an incentive to use roads without a new cycle lane that replaced a car lane.

5.5.1.4 Robustness

I run a variety of tests in order to assess the robustness of the results with respect to aspects like sample composition, clustering, placebo treatments and variations of control variables.

Standard Errors In my main estimates, I cluster standard errors on 1km $\times$ 1km city-wide grids and weeks in order to account for errors being spatially and temporally correlated. To test whether statistical significance depends on the definition of standard errors, I alter this cluster specification by using 1) the twelve local districts of Berlin and weeks, 2) a station/week cluster, 3) standard errors only clustered on grid level, and 4) standard errors only clustered on weekly level. These different types of clustering thus take into account different variations of spatial clusters and of the time component. This means that the data then is treated as independent across the respective clusters (Cameron et al., 2011). Results in Table 5.B.10 show that inference is not affected. The effect on speed is still significant, at least at a 5 percent level and mostly at a 1 percent level. Excluding the station with the largest impact on the results ("Station 15" in Table 5.B.2) in *Panel B* also does not alter the results, regardless of cluster specification.

Sample adjustments In the main part of the analysis, I make some restrictions to the sample, e.g. regarding the number of lanes.[57] I vary the sample composition to check if and to what extent different sampling structures may play a role. In the case of significantly different outcomes, estimates are potentially biased due to sample selection. Results with different sample restrictions are shown in Table 5.B.11.

Firstly, columns 1 and 2 show the results using the full sample, including one-lane-streets and four-lane-streets, and the full set of 24 hours. While the volume coefficient is still not significant, the average speed estimate becomes a little smaller in size. The most likely reason is that part of the effect is offset in night hours, where there is not much traffic on the roads in general.

Secondly, I look at the full sample again, but now only for non-night times, i.e. from 5 a.m. until 8 p.m. as shown in columns 3 and 4. The formerly made presumption regarding speed and its effect being slightly offset during night hours is confirmed, since the coefficient in column 4 jumps back to the result found in the main analysis. Taking into account one-lane-streets and four-lane-streets, however, renders the volume coefficient significant and slightly larger. Thus, if we consider the entire road network in Berlin as control group, we see that the volume on treated streets has actually decreased. The effect is driven by one-lane-streets, which account for about 25 percent of the sample size.

Thirdly, I only consider streets without any type of prior cycling infrastructure. Then, treated streets are such with a PUBL as the first type of cycling infrastructure and control streets are such without any type of cycling infrastructure throughout the sampling period.

[57] I exclude all one- and four-lane streets.

The reason to look at this sample restriction is twofold. Firstly, almost all treated stations had no prior cycling infrastructure. Therefore, the control group is even more harmonized in terms of characteristics. Secondly, while treated streets receive a direct alternative as mode of transport, bikes do not have an own lane to use in the comparison group. If the volume effect was now larger compared to the main outcomes, this might indicate that car drivers were nudged away from using their car and potentially switched to using the bike. The sample size is considerably smaller in these specifications. While the volume effect remains insignificant, indicating that the new cycle lanes are not used by former motorists, the coefficient for average speed becomes slightly larger. Hence, restricting the sample to streets with very similar characteristics shows a slightly stronger congestion-effect than allowing for a more generalized set of streets in Berlin to be part of the control group.

Lastly, in my main estimations I exclude all stations, which lie on a highway, thereby only comparing inner-city streets with each other. The last two columns check whether results hold in the case of including highways into the control group. Results are very close to the full-sample outcomes in columns (1) and (2). The size of the sample is now about twice as large compared to the standard sample and coefficient sizes are similar. Though economically small, the effect on volume is now positive and significant. One potential reason is that home office regulations in times of the COVID-19 lockdown had a larger effect on the highways circling the city and this effect is not entirely subsumed by the *post-lockdown* $\times$ *stations* interaction. That would mean that highway travel was significantly reduced compared to inner-city commuting. However, since I would not consider highways to be part of an adequate control group, this difference should not be over-interpreted. Overall, different sample specifications suggest that results are not prone to selection bias.

Placebo tests A common approach in DiD models to assess the robustness of results is the performance of placebo tests. Table 5.B.12 presents the results of such tests with respect to treatment timing. Therefore, I deleted all observations in the actual treatment period starting in mid March, 2020. Then, treatment status was assigned to all treated measuring stations for three different placebo-treatment dates. The dates were chosen at intervals of about four months in the time frame between April of 2019 and January, 2020. The table shows that all placebo treatment effects are insignificant except for volume in the case of treatment starting in April of 2019, which was about one year prior to actual treatment. In this case, the coefficient is marginally significant with an economically very small effect. Due to the large time interval between actual and placebo treatment as well as the fact that the pre-trend assumption was more fuzzy in the case of volume, I conclude that placebo tests generally support the main results.

Construction work control In my main estimations, I have used a construction work dummy as control variable that switched to one, whenever there was any type of road construction during the observation period. This means that the dummy does not take into account what kind of restriction was imposed on road users. However, there are many different types of construction works and accompanying restrictions. It is likely that imposing a temporary stopping restriction on a street differently affects the average speed of vehicles compared to blocking an entire lane. The reason why I use a dummy variable is that it is not straight-forward to interpret a categorical construction work variable, since there does not exist a natural order, which tells about the severity of restriction. Table 5.B.13 compares the results between using a construction work dummy and allowing the type of restriction to vary. The effect on volume remains insignificant and very small in absolute size. Car speed is significantly affected with about the same coefficient size. The table also reveals that it is important to control for construction work, which significantly affects both, volume and speed.

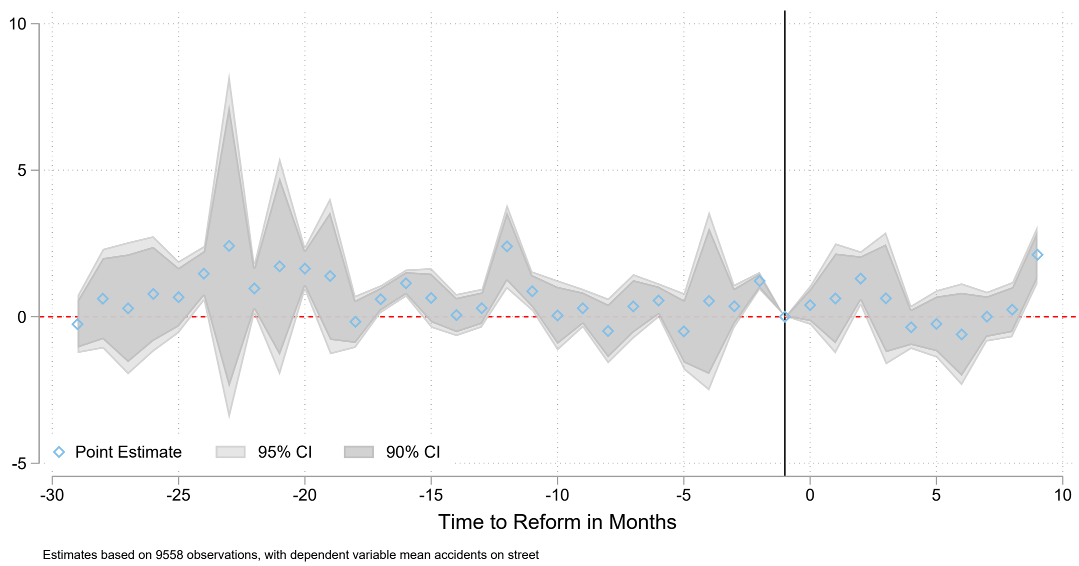
Concerns about Standard TWFE Standard TWFE estimations assume a constant treatment effect, which causes a potential bias, when treatment varies over affected unit and time. This bias is caused by potentially negative weights assigned to the treatment effect of single treated units, which then compose a weighted sum over all DiD estimations. This becomes a problem if the average treatment effects are heterogeneous across groups or periods (De Chaisemartin and d'Haultfoeuille, 2020). With respect to treated units, I show in the leave-one-out analyses, as presented in Table 5.B.2, that the results are homogeneous except for one station. Moreover, the initial event study design exhibits not much variation of the effect over time after treatment. Therefore, when accounting for the outlier, the respective bias does not seem to weigh heavy. However, I still compute the weights as suggested by De Chaisemartin and d'Haultfoeuille (2020). All weights in my regressions are positive, which suggests that this bias is of no relevance in this setting. Apart from checking the weights, I also repeat my main analyses without heterogeneous timing, which again addresses the potential problem when streets are treated at different points in time. Now, all treated roads are assigned the same pre-treatment and post-treatment period. As a consequence, all observations between March, 25th and June, 30th are deleted and the treatment indicator switches to one after that period for all treated measuring stations. Table 5.B.14 shows results for which I delete all observations between the first and the last treatment date. *Panel A* includes all treated traffic monitoring stations, while *Panel B* excludes "Station 15" from Table 5.B.2 in order to show further results of a more homogeneous set of treated roads. The effect on volume and speed are very similar to my main outcomes, which suggests that the staggered timing of treatment in the main analyses does not pose a major problem.

5.5.2 Accidents

To get a first visual impression of the accidents data within Berlin, I plot the monthly mean development of total accidents by treatment status in Figure 5.A.4a. Since about 40 percent of overall observations are zeroes and the maximum number of total accidents per street and month is 20, the monthly mean by treatment status is relatively low and ranges between approximately .5 and 2. This means that the time variation of accidents is limited. The development of accident occurrence on treated and non-treated streets seems to be similar before and after treatment. Panels (b) and (c) of Figure 5.A.4 show that this observation is independent from the means of transport involved in an accident. Apparently, the occurrence of accidents is cyclical with a higher incidence in summer months.

Event study design Now I estimate the event study model with accidents, where the outcome variable includes either all accidents, only those involving bicycles, or those involving cars. Accidents are aggregated at the street level and are only available on a monthly basis. Thus, the results depict the differences between PUBL streets and non-PUBL streets per month. Figure 5.6 shows that the variation of overall accidents between treated and untreated streets is relatively small in absolute terms and that there appear to be no treatment-induced changes. For accidents involving bikes, the variation is even smaller, as Figure 5.A.5 exhibits. Moreover, I fail to detect differences if I take into account the severity of injuries. This means that there is no change in cases of accidents with fatalities or serious injuries, nor in accidents without such victims (results not shown here). Thus, at least in the short term that observations are available, PUBLs did not lead to a significant change of accidents on treated streets, regardless of the mode involved or accident severity. However, this development does not take into account the number of cyclists on these streets due to a lack of data available. If there was a significant increase in cyclists, then the possibility of a decline in per-cyclist accidents is very likely. Repeating the analysis separately for each treated street or conducting the leave-one-out analysis (not shown here) does not alter these results.

Figure 5.6: Event study outcomes for accidents



(a) Overall accidents

Note: Own calculations. The graph shows the results of an event study estimation as described in Equation 5.1. The dependent variable is overall accidents. Blue dots represent the main estimation coefficients of leads and lags. Confidence Interval (CI) are depicted as area shades in light (95% CI) and dark grey (90% CI). The vertical solid black line shows the time of treatment, which is anchored at 0. Leads and lags are the time before and after treatment in months. The estimations include street and month fixed effects as well as controls for road condition at the time of the accident and the type of street (primary, secondary etc.). Standard errors are clustered at 1km × 1km grid cell level spanning the city times a running month variable.

Synthetic control group design Green et al. (2016) analysed the effects of a congestion charge in London on traffic accidents and used the synthetic control group method (Abadie et al., 2010) to identify the effects of interest. Their unit of treatment was the city of London while other cities in the country served as synthetic control units. In order to make my results more robust, I follow their approach using municipality level data for Germany as a whole.[58] Now, the outcome variable is the number of accidents (and subgroups of such) per month and municipality.[59] Berlin is subdivided into its 12 municipalities and treatment status is assigned to the month in which the first street within the administrative unit receives a PUBL. Since I exploit the variation of accidents between municipalities over time, the outcome of this approach also subsumes potential spillover effects on other streets rather than only actually treated ones. Due to the fact that treatment months differ between Berlin's municipalities, I run the synthetic control group method separately for each treated entity rather than pooling them into one treatment group.

The respective treatment unit and synthetic control units are matched on various economic and socio-demographic characteristics, which might influence the decision to use certain modes of traffic and could influence the amount of accidents in a region. Among those are the population of the municipality, the share of Green Party voters, the share of space used for settlements and for traffic respectively, the number of unemployed in the region, and the annual number of accidents from 2018 to 2020.[60] Tables 5.B.15 to 5.B.18 show the predictor balances between each treated and the corresponding synthetic control group. For all five municipalities the predictor matches are very close. Thus, in terms of predictors, the synthetic control groups resemble the treated municipalities.

Actual results on overall accidents by means of a graph are shown in Figure 5.A.6. All subgraphs support the conjecture that the synthetic control groups are good matches for the treated units since the pre-treatment outcomes of both follow parallel paths. Just like in the event study design, I fail to find significant differences between treated units and controls after the installation of PUBLs. Figure 5.A.7 complements this finding by showing accidents involving cars and bikes separately. This means that taking a more macro-economic view by considering entire treated municipalities leads to the same outcome as the within city street-level evaluation. This strengthens the finding of PUBLs not having an effect on total accidents. However, as noted earlier, this ignores the number of cyclists on the streets and

[58] I only exclude municipalities, which received a PUBL themselves in the observed time frame. This includes e.g. the cities of Hamburg and Munich.

[59] The reason not to run a synthetic control group design with traffic data is because I do not have available traffic data for the whole of Germany, but only for Berlin.

[60] In Germany, there are more than 10.500 municipalities in about 400 districts. While most variables, like population, voting behaviour, and land use designation, are available on municipality level, some information like the unemployment rate, is only publicly available on district level. Since municipalities are administrative sub-divisions of districts, I assign the numbers of the district to the respective municipality.

potentially implied decreases (in case of an increase of bicycle users) of per-cyclist accidents.

5.6 Interpretation and discussion

Accidents The results indicate that accidents in absolute terms did not change as a consequence of installing PUBLs. The only comparable paper to identify the effect of bike lanes on accidents by Li et al. (2017) found a contradictory result with a total increase of collisions of about 40% in the aftermath of new cycling ways in London. Due to the fact that Li et al. (2017) have data about the number of cyclists on the treated routes, they are able to directly estimate an effect on the accident rate, which is not possible in my case. As a consequence of an increase of total cyclists, the authors do not find a significant impact of new bike paths on accidents per cyclist. With respect to PUBLs in general, Kraus and Koch (2021) found an increase of cycling in European cities after the installation of PUBLs of about 40% on average. The authors only looked at cycling in entire cities, not taking into account the type of streets specifically affected. If the increase of cyclists transfers to streets with PUBL, that would mean that accident rates on PUBL streets in Berlin actually have decreased. One potential reason for the difference between the London case analysed by Li et al. (2017) and Berlin, that is subject to my study, may lie in the nature of the cycling lanes. While the lanes I consider are separated from car traffic by physical barriers, many bike lanes in London are merely indicated by blue paint on the streets. Taking all aspects into account, my results suggest that cycling has likely become safer on roads with a PUBL. However, additional research on the matter is required due to the rather small sample size and thereby limited variation of the accidents data.

Traffic I find a significant reduction of average speed on PUBL streets, which means higher congestion levels compared to untreated roads. This increase in congestion seems to be primarily driven by the reduction of space available for cars rather than a significant change of total traffic on these streets. If anything, traffic on PUBL streets has slightly declined. In the economic literature, in many cases there is no differentiation between congestion and traffic volume since higher congestion levels in most instances are caused by an increase in traffic rather than a change in infrastructure. In theory, both, the increase in car volume as well as an increase in congestion, may lead to higher levels of local pollution. More cars mean more combustion engines to pollute the air, while higher congestion may increase pollution as a consequence of stop-and-go driving, which leads to higher fuel consumption or increased tire wear (Tu et al., 2022; Sommer et al., 2018). Since the results found in this paper hint toward increased congestion with slightly reduced traffic, conclusions about the effects on air pollution cannot be drawn and the link found e.g. between congestion, traffic pollution, and infant mortality (Knittel et al., 2016; Currie and Walker, 2011) cannot be applied without further ado.

One unambiguous external cost factor of PUBLs, which is borne by car drivers, is the price paid in terms of increased travel times. Based on my estimation coefficients, a driver with an hypothetical one hour commute to work would need about one hour and five to six minutes after the installation of cycle lanes on the routes she uses. Given economic time costs of 9.37€ per hour in Germany, which are based on estimated values of travel time savings (INRIX, 2021), the additional minutes would lead to an increase in time costs to approximately 10.3€ for the hypothetical commuter per one-way trip. This corresponds to a loss of about 2€ per day given that travel times from and to work do not differ. Assuming about 250 days of work a year, this would add up to costs of about 500€ for that specific driver. However, these back-on-the-envelope calculations are very hypothetical, since this would require all streets within the city to be equipped with new cycle lanes, which replace an existing car lane. The longest PUBL, which was installed in Berlin, had a length of about 3.5 kilometres (Kantstr.). If you needed five minutes to pass this specific street before the establishment of the cycle lane, then a 10-percent decrease in average car speed would lose you about 30 seconds. The economic costs in this specific case are therefore limited.[61]

5.7 Conclusion

This paper is among the first to analyse causal effects of bike lanes on multiple outcomes, which determine some of the most important aspects of life-quality in cities. My source of exogenous variation are pop-up bike lanes in Berlin and I analyse their effect on congestion, traffic volume, and accidents. While the number of cars experienced modest but mostly insignificant declines, I find a significant reduction of average speed by between 8 and 12

[61] Actually, repeating the calculation with the hypothetical driver, who then loses 60 seconds a day on Kantstraße, results in yearly costs of only 39€ approximately.

percent. This effect reaches its maximum in peak travel hours with average speed being slower by 16 percent. The absolute number of accidents was not affected by the installation of pop-up bike lanes. However, it is likely that the rate of accidents per cyclist has decreased since prior research suggests an increase in people using bikes in PUBL-cities (Kraus and Koch, 2021). The actual effect on accident rates remains an open question for future research. Determining the effects of newly installed cycle lanes on local air quality is also beyond the scope of this paper, but would be a fruitful amendment to the literature, especially in the light of calculating a more complete cost-benefit analysis. Overall, economic costs of increased travel times are rather modest. The balance between costs and benefits depends on developments in the long run and whether commuters are nudged towards using bikes instead of motorized traffic.

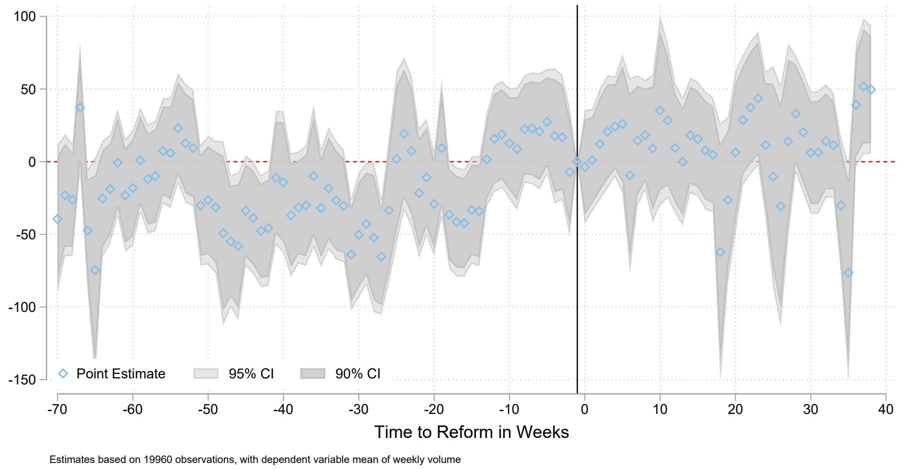
My findings have to be considered in the light of other consequences of the COVID-19 pandemic. Public transport has experienced a significant decline in trust and a decrease in ridership numbers (Vitrano, 2021). While some commuters might have replaced the tram or subway with their bikes, there also may exist the tendency to use the car instead. Street-specific data on bicycle as well as public transport utilization would allow for a thorough analysis of the change in the modal split. Moreover, the development of the modal shift in the long run, after the end of the COVID-19 pandemic, is unclear. Further research should therefore tackle the question of long-term effects of replacing a car with a bike lane.

On the basis of my research I furthermore conclude that the fundamental law of road congestion, which suggests a unitary elastic relationship between lane kilometres available and miles driven, does not necessarily apply in the short run. In the original paper by Duranton and Turner (2011), a change of vehicle lane kilometres does not relieve the streets sufficiently since over the course of several decades congestion remains stable. My findings indicate that infrastructural changes do not directly lead to adaptations in commuting behaviour, but rather require a longer time span to evolve. Consequently, it apparently takes time or different, potentially tougher, measures to disengage motorists from their sticky preferences.

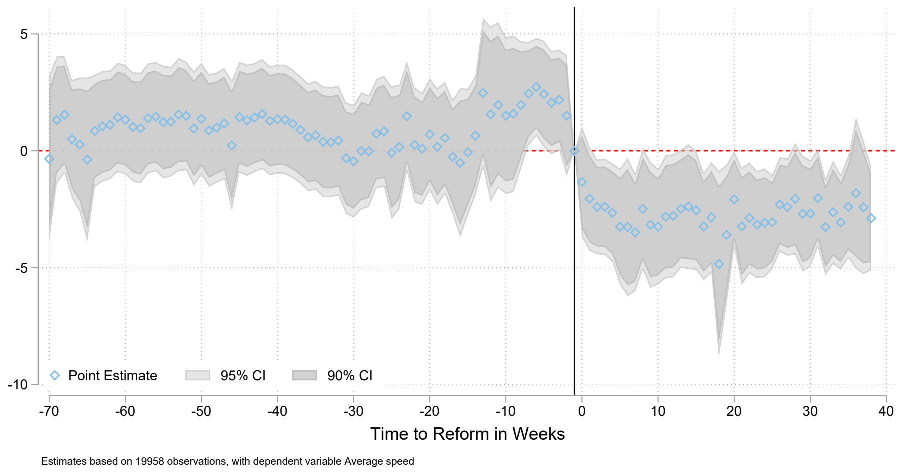
Appendices

5.A Figures

Figure 5.A.1: Main results including an interaction between Station ID and Lock-down dummy



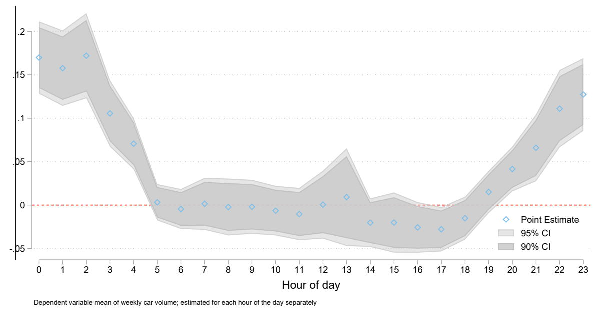
(a) Number of vehicles



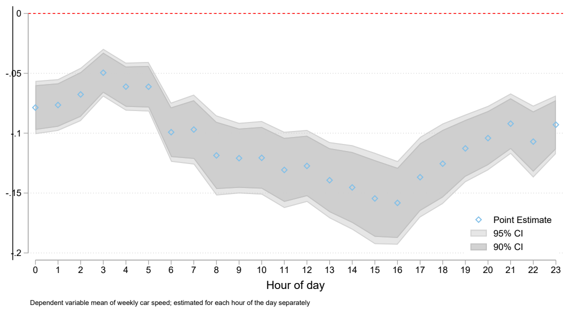
(b) Average speed

Note: Own calculations. The graphs show the results of separate event study estimations for the absolute number of vehicles (5.A.1a) and average vehicle speed (5.A.1b). Blue dots are the main coefficients of leads and lags. Confidence intervals are depicted as area shades in light (95% CI) and dark grey (90% CI). The vertical solid black line shows the time of treatment anchored at 0. Leads and lags are the time before and after treatment in weeks. The estimations include station and week fixed effects as well as an interaction variable between each station and a post-lockdown dummy. The sample is restricted to streets with two or three lanes and all streets within a radius of 1km to a treated street are excluded from the estimations. Standard errors are clustered at 1km × 1km grid cell level spanning the city times a running week variable.

Figure 5.A.2: Effects separated by hour of the day - logged outcomes



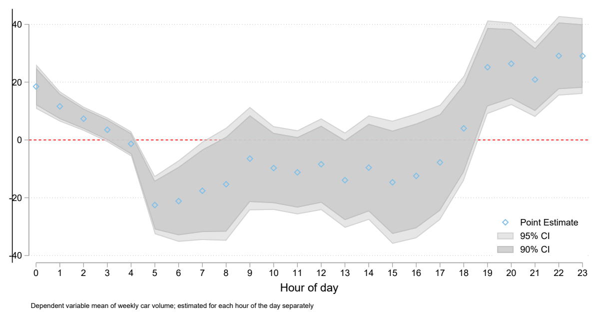
(a) Number of vehicles



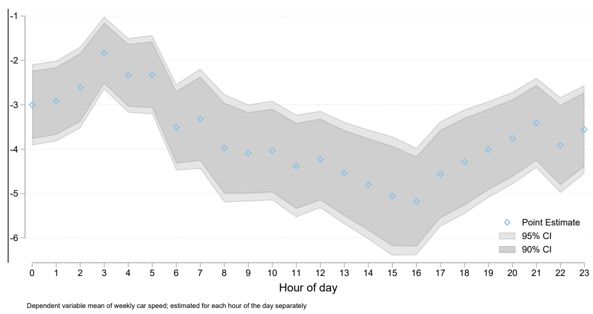
(b) Average speed

Note: Own calculations. The graphs show the results of separate TWFE estimations for each hour of the day for the logged number of vehicles (5.A.2a) and logged average vehicle speed (5.A.2b). Blue dots represent the treatment effect of each estimation. Confidence intervals are depicted as area shades in light (95% CI) and dark grey (90% CI). The estimations include station, and date fixed effects, a construction dummy, a control for changes in speed regulations, and a station $\times$ post-lockdown interaction. The sample is restricted to streets with two or three lanes and all streets within a radius of 1km to a treated street are excluded from the estimations. Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running week variable.

Figure 5.A.3: Effects separated by hour of the day - absolute outcomes



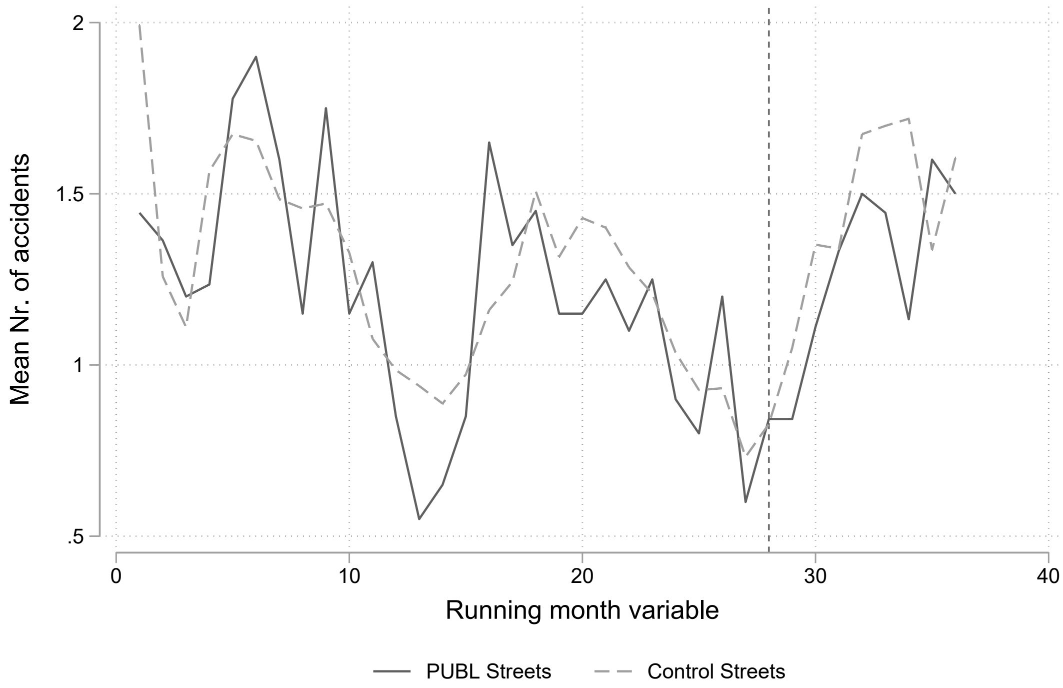
(a) Number of vehicles



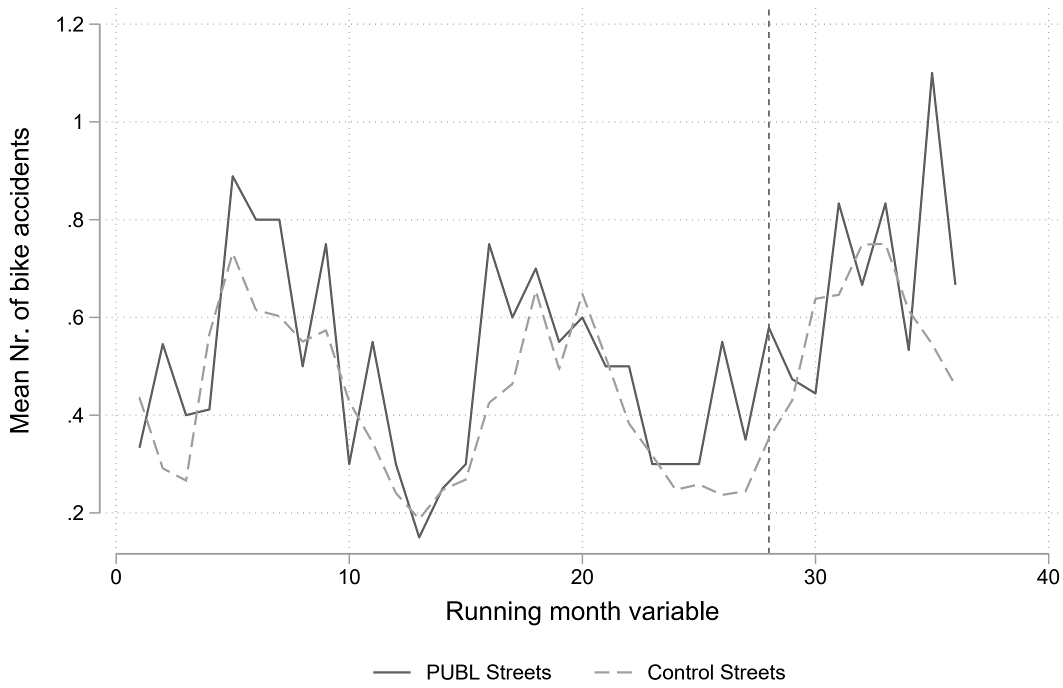
(b) Average speed

Note: Own calculations. The graphs show the results of separate TWFE estimations for each hour of the day for the absolute number of vehicles (5.A.3a) and average vehicle speed (5.A.3b). Blue dots represent the treatment effect of each estimation. Confidence intervals are depicted as area shades in light (95% CI) and dark grey (90% CI). The estimations include station, and date fixed effects, a construction dummy, a control for changes in speed regulations, and a station $\times$ post-lockdown interaction. The sample is restricted to streets with two or three lanes and all streets within a radius of 1km to a treated street are excluded from the estimations. Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running week variable.

Figure 5.A.4: Development of mean accidents by treatment status

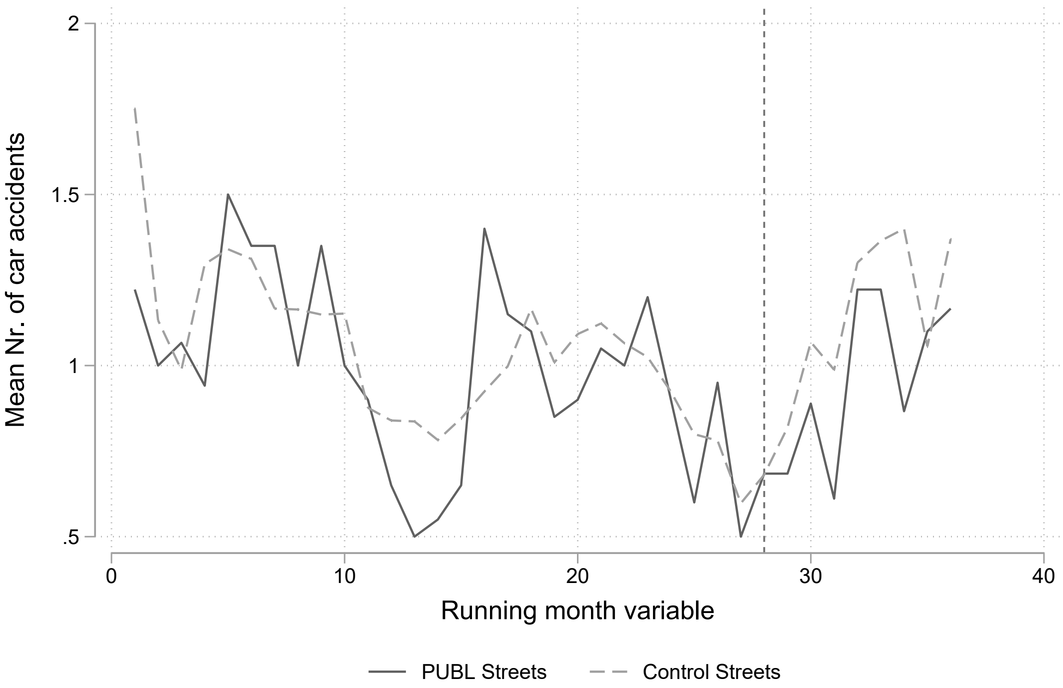


(a) Overall accidents



(b) Bike accidents

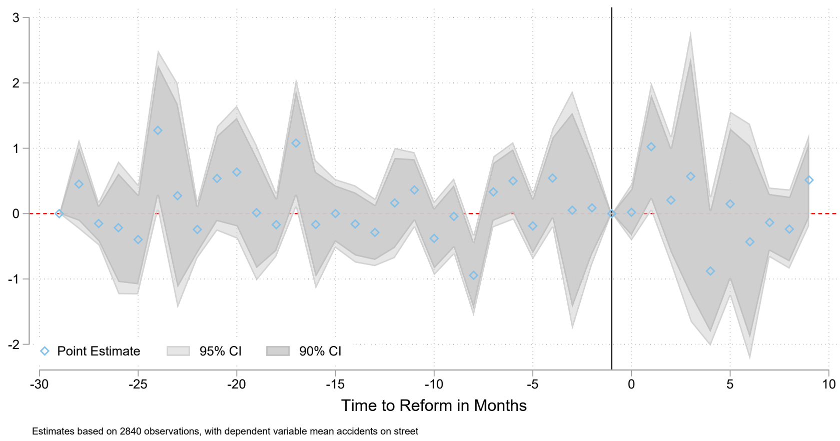
Figure 5.A.4: Development of mean accidents by treatment status



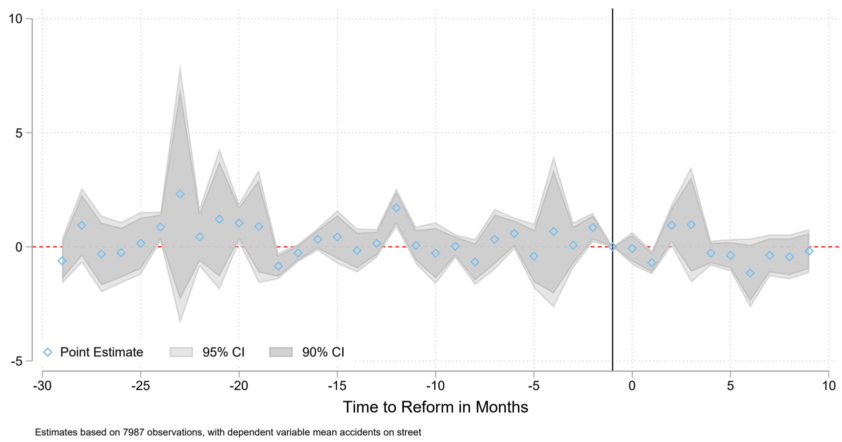
(c) Car accidents

Note: Own calculations. The graphs show the development of average street-level accidents separated by treatment status. They are presented by types of vehicles involved in the accidents, more precisely by overall accidents (5.A.4a), bike accidents (5.A.4b), and accidents with cars involved (5.A.4c). The vertical dashed line represents the timing of the installation of the first PUBL in the city on March, 25th in 2020.

Figure 5.A.5: Effects on Bike and Car Accidents



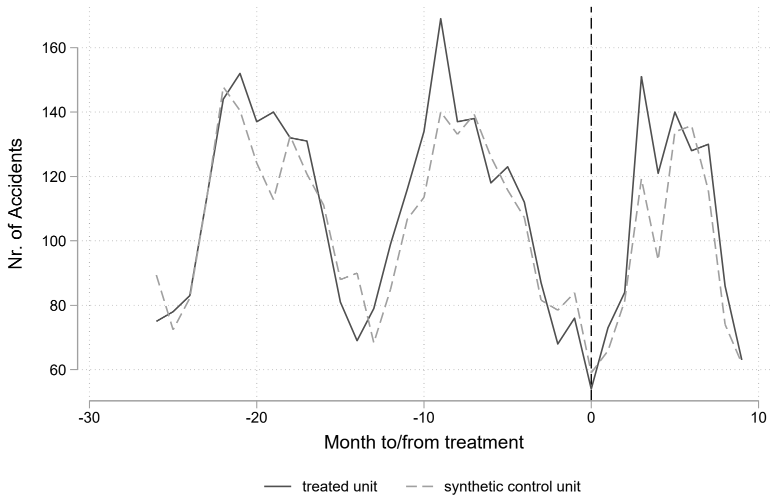
(a) Bike accidents



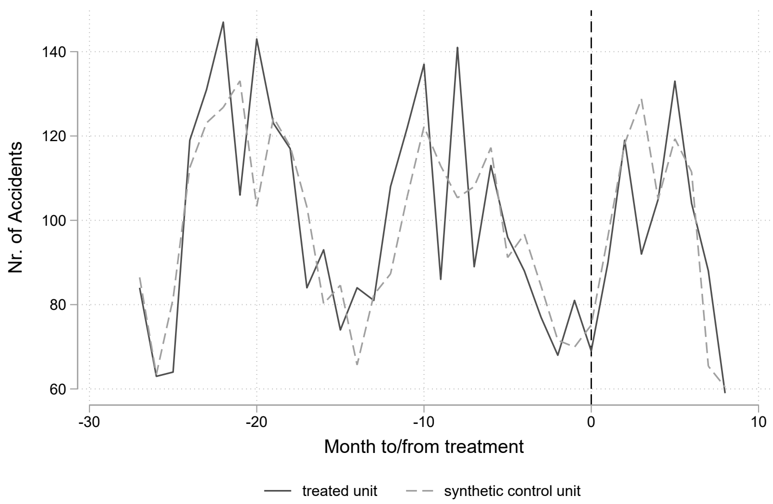
(b) Car accidents

Note: Own calculations. The graphs show the results of separate event study estimations as described in Equation 5.1. Outcomes are accidents with bicycles involved (Figure 5.A.5a), and accidents with cars involved (Figure 5.A.5b). Blue dots represent the main estimation coefficients of leads and lags. Confidence intervals are depicted as area shades in light (95% CI) and dark grey (90% CI). The vertical solid black line shows the time of treatment, which is anchored at 0. Leads and lags are the time before and after treatment in months. The estimations include street and month fixed effects as well as controls for road condition at the time of the accident and the type of street (primary, secondary etc.). Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running month variable.

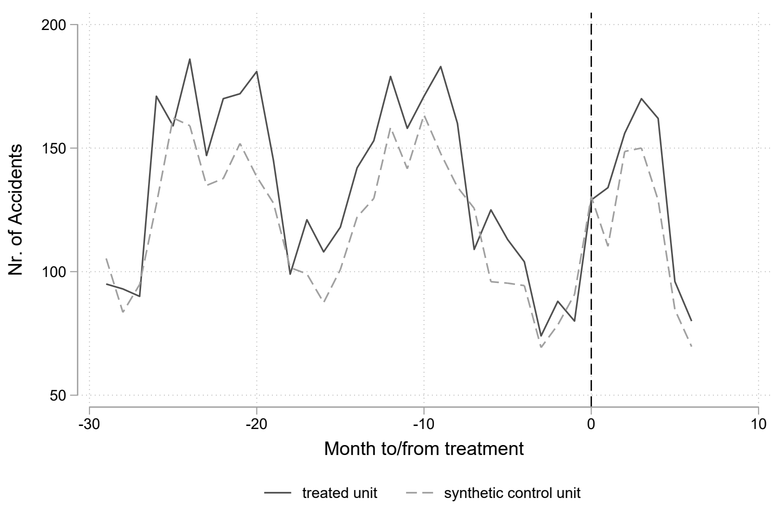
Figure 5.A.6: Effects on Accidents using synthetic control method



(a) Accidents Kreuzberg-Friedrichshain

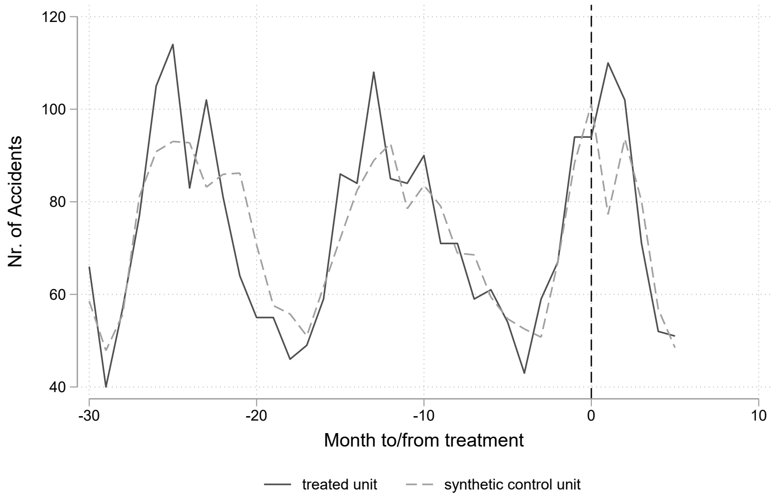


(b) Accidents Pankow

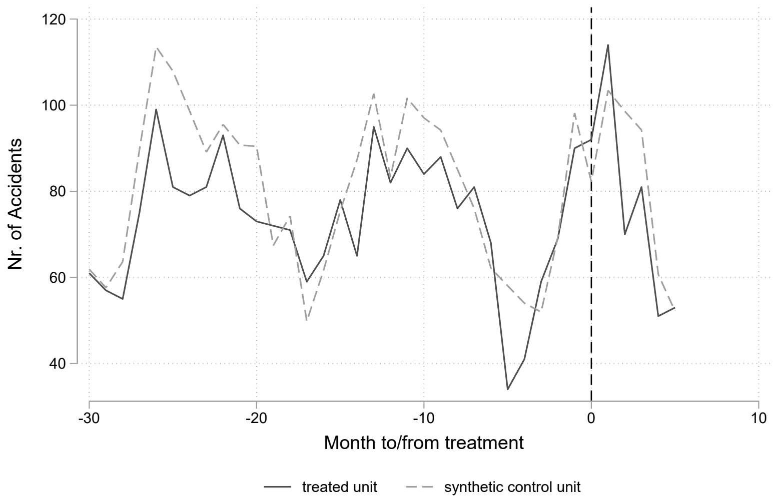


(c) Accidents Charlottenburg-Wilmersdorf

Figure 5.A.6: Effects on Accidents using synthetic control method



(d) Accidents Treptow-Köpenick

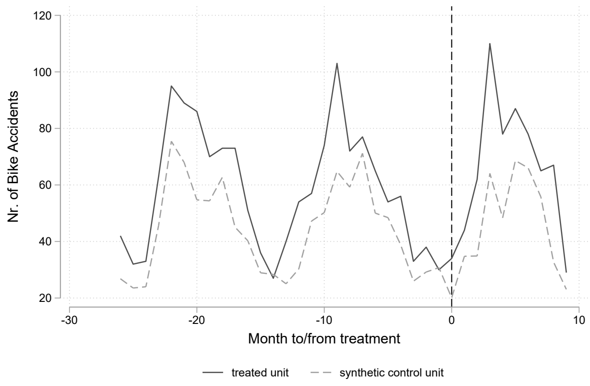


(e) Accidents Neukölln

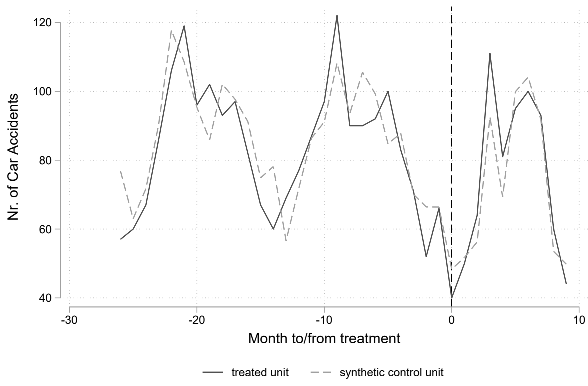
Note: Own calculations. The graphs show the results of separate synthetic control group estimations following Abadie et al. (2010). Outcome variable is the total number of accidents. Treated units are the respective Berlin districts as mentioned in each sub-caption in which a PUBL was installed. The time of treatment is the month of the first placement of a PUBL within the respective district. The synthetic control unit consists of potentially all municipalities in Germany outside of Berlin that did not receive a PUBL. The matching between treatment and control units is based on the respective monthly outcome variable prior to treatment, traffic space, space used for settlements, population, election participation, the share of green party voters, the unemployment rate, and the absolute number of accidents from 2018 until 2020.

Figure 5.A.7: Effects on Bike and Car Accidents using synthetic control method

Kreuzberg-Friedrichshain

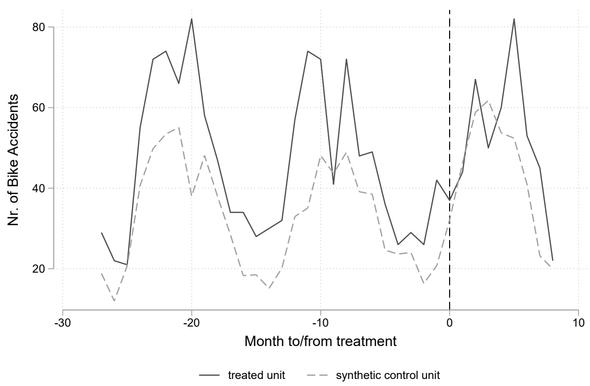


(a) Bike accidents

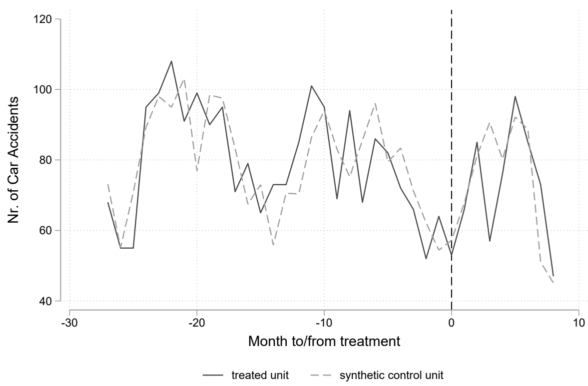


(b) Car accidents

Pankow

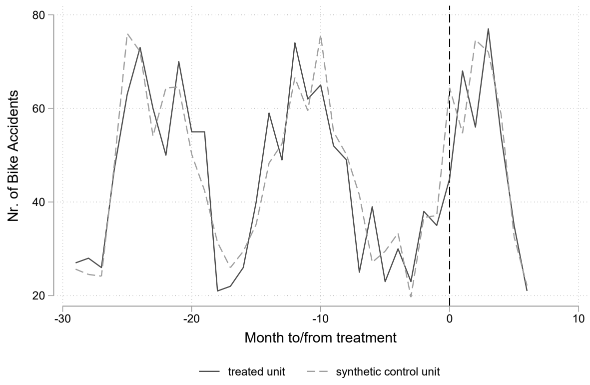


(c) Bike accidents

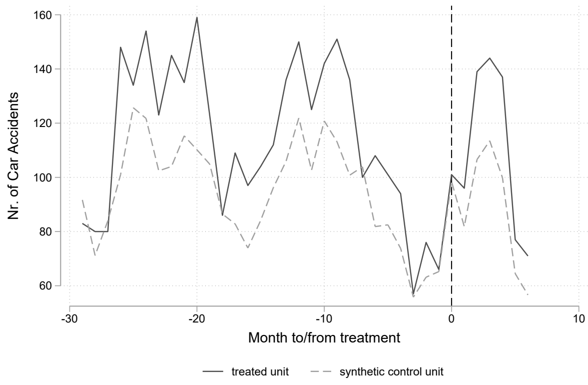


(d) Car accidents

Charlottenburg-Wilmersdorf



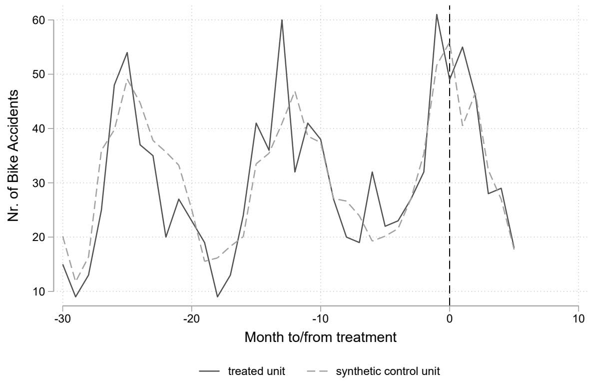
(e) Bike accidents



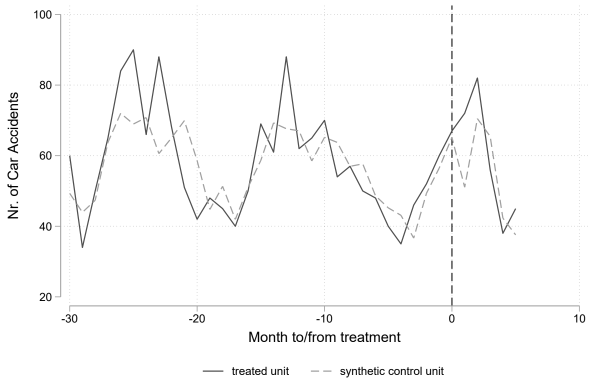
(f) Car accidents

Figure 5.A.7: Effects on Bike and Car Accidents using synthetic control method

Treptow-Köpenick

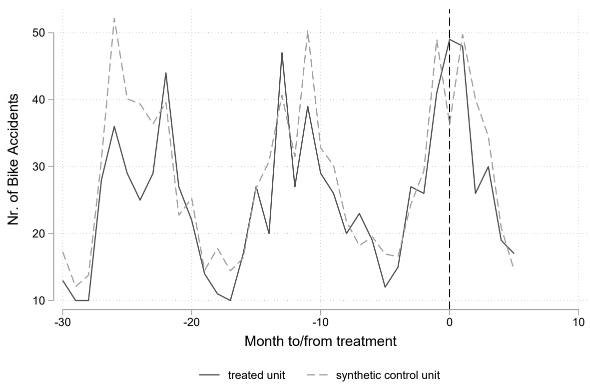


(g) Bike accidents

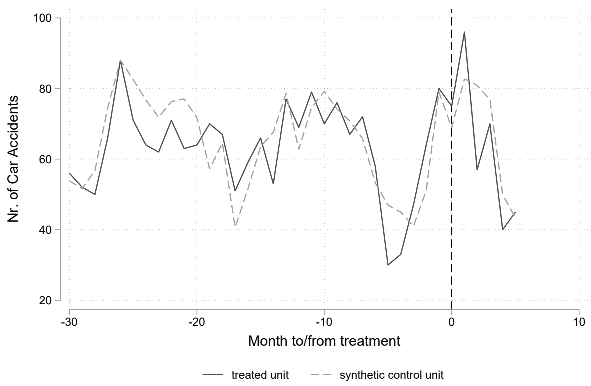


(h) Car accidents

Neukölln



(i) Bike accidents



(j) Car accidents

Note: Own calculations. The graphs show the results of separate synthetic control group estimations following Abadie et al. (2010). Outcome variables are accidents with bicycles involved in the left panel and those with cars involved in the right panel (those are not mutually exclusive and may partly contain the same accidents). Treated units are the respective Berlin districts as mentioned in each sub-caption in which a PUBL was installed. The time of treatment is the month of the first placement of a PUBL within the respective district. The synthetic control unit consists of potentially all municipalities in Germany outside of Berlin that did not receive a PUBL. The matching between treatment and control units is based on the respective monthly outcome variable prior to treatment, traffic space, space used for settlements, population, election participation, the share of green party voters, the unemployment rate, and the absolute number of accidents from 2018 until 2020.

5.B Tables

Table 5.B.1: Effect on weekly Volume and Speed (absolute and logged)

	Volume		Speed	
	(1)	(2)	(3)	(4)
Panel A: Outcome in absolute values				
1(PU lane)	-53.82***	0.550	-5.122***	-4.148***
	(5.354)	(7.547)	(0.311)	(0.538)
N	20196	20196	20194	20194
R^2	0.901	0.920	0.816	0.844
Stations	216	216	216	216
Interaction	No	Yes	No	Yes
Panel B: Log-transformed outcome				
1(PU lane)	-0.0292***	0.00788	-0.131***	-0.123***
	(0.00890)	(0.0135)	(0.00831)	(0.0139)
N	20194	20194	20194	20194
R^2	0.607	0.664	0.750	0.783
Stations	216	216	216	216
Interaction	No	Yes	No	Yes

Note: Own calculations. The table presents the coefficients of the treatment effects of separate two-way fixed effects estimations with vehicle volume and vehicle speed as dependent variables and data aggregated to weekly levels. Panel A shows the coefficients of interest with outcomes in absolute terms. Panel B shows the same for logged outcome variables. Even columns include an interaction term between a unique measuring station identifier and a post-lockdown dummy variable while uneven columns do not. All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, and an indicator for a change in speed limits. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. Standard errors are clustered at 1km × 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.2: Leave-one-out analyses by station

	(1) Volume	(2) Speed
Station 1	-0.00500	-0.125***
	(-0.39)	(-8.93)
Station 2	-0.00510	-0.129***
	(-0.39)	(-8.69)
Station 3	-0.00874	-0.124***
	(-0.67)	(-8.15)
Station 4	-0.000635	-0.121***
	(-0.05)	(-8.10)
Station 5	-0.0183	-0.147***
	(-1.60)	(-10.52)
Station 6	-0.00569	-0.120***
	(-0.45)	(-8.23)
Station 7	-0.0114	-0.120***
	(-0.90)	(-8.02)
Station 8	-0.00390	-0.115***
	(-0.30)	(-7.45)
Station 9	-0.00203	-0.123***
	(-0.15)	(-7.86)
Station 10	-0.00577	-0.117***
	(-0.45)	(-7.86)
Station 11	0.00779	-0.133***
	(0.61)	(-8.72)
Station 12	-0.00140	-0.128***
	(-0.11)	(-9.05)
Station 13	-0.00574	-0.124***
	(-0.45)	(-8.34)
Station 14	-0.00648	-0.120***
	(-0.49)	(-7.85)
Station 15	-0.00590	-0.0863***
	(-0.45)	(-6.48)
Station 16	-0.00579	-0.123***
	(-0.46)	(-8.43)
Station 17	-0.00564	-0.122***
	(-0.45)	(-8.36)
Station 18	-0.00556	-0.123***
	(-0.44)	(-8.47)
Station 19	-0.00619	-0.123***
	(-0.49)	(-8.40)
Station 20	-0.00567	-0.119***
	(-0.45)	(-8.40)
Station 21	-0.00506	-0.118***
	(-0.40)	(-8.34)
Station 22	-0.00567	-0.123***
	(-0.45)	(-8.30)
Station 23	-0.00520	-0.123***
	(-0.41)	(-8.33)

Note: Own calculations. The table presents outcomes of leaving out the respective station in the analyses. Control variables and limitations are the same as in Table 5.1. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.3: Peak hour results

	All Stations				No outlier stations			
	Volume		Speed		Volume		Speed	
	(1) ln	(2) abs	(3) ln	(4) abs	(5) ln	(6) abs	(7) ln	(8) abs
1(PU lane)	-0.00907	-7.060	-0.121***	-4.109***	-0.0117	-6.540	-0.0851***	-2.711***
	(0.0119)	(7.999)	(0.0145)	(0.540)	(0.0124)	(8.644)	(0.0134)	(0.463)
N	771418	772950	771404	771404	767695	769227	767681	767681
R^2	0.617	0.740	0.602	0.766	0.618	0.741	0.602	0.767
Stations treated	23	23	23	23	23	23	23	23
Stations Overall	215	215	215	215	214	214	214	214
Interaction	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Own calculations. The table presents the coefficients of the treatment effects of separate two-way fixed effects estimations with vehicle volume and vehicle speed as dependent variables. Only peak traffic hours between 6 a.m. and 9 a.m. as well as between 4 p.m. and 7 p.m. are considered. Even columns show outcomes in absolute terms, while uneven columns show outcomes with log-transformed dependent variables. All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, an indicator for a change in speed limits, and an interaction term between a station identifier and a post-lockdown dummy. Columns 1-4 include all stations of the main sample, columns 5-8 exclude outlier stations as identified by leave-on-out analyses. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.4: Results by hour of the day (Midnight - 11 a.m.)

	Log outcomes		Abs. outcomes	
	(1)	(2)	(3)	(4)
	Volume	Speed	Volume	Speed
hour 0	0.170***	-0.0786***	18.50***	-3.001***
	(8.00)	(-6.93)	(4.71)	(-6.44)
hour 1	0.158***	-0.0765***	11.61***	-2.915***
	(7.11)	(-6.94)	(4.35)	(-6.30)
hour 2	0.172***	-0.0677***	7.347***	-2.610***
	(6.88)	(-5.96)	(3.45)	(-5.55)
hour 3	0.106***	-0.0495***	3.510	-1.835***
	(5.39)	(-4.86)	(1.62)	(-4.38)
hour 4	0.0707***	-0.0612***	-1.335	-2.336***
	(4.68)	(-5.98)	(-0.60)	(-5.45)
hour 5	0.00323	-0.0612***	-22.52***	-2.324***
	(0.30)	(-5.80)	(-4.38)	(-5.11)
hour 6	-0.00452	-0.0992***	-21.15**	-3.507***
	(-0.38)	(-7.86)	(-2.94)	(-7.07)
hour 7	0.00145	-0.0970***	-17.59*	-3.315***
	(0.10)	(-6.51)	(-2.03)	(-5.75)
hour 8	-0.00224	-0.119***	-15.32	-3.979***
	(-0.13)	(-6.95)	(-1.54)	(-6.39)
hour 9	-0.00199	-0.121***	-6.457	-4.084***
	(-0.13)	(-8.04)	(-0.71)	(-7.33)
hour 10	-0.00640	-0.121***	-9.718	-4.031***
	(-0.44)	(-7.69)	(-1.32)	(-7.04)
hour 11	-0.0103	-0.131***	-11.20	-4.378***
	(-0.67)	(-8.05)	(-1.51)	(-7.43)

Note: Own calculations. The table presents outcomes of running the main analysis for each hour of the day separately from midnight (hour 0) until 11 a.m. (hour 11). All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, an indicator for a change in speed limits, and an interaction term between a station identifier and a post-lockdown dummy. Columns 1-4 include all stations of the main sample, columns 5-8 exclude outlier stations as identified by leave-on-out analyses. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.5: Results by hour of the day (12 a.m. - 11 p.m.)

	Log outcomes		Abs. outcomes	
	(1)	(2)	(3)	(4)
	Volume	Speed	Volume	Speed
hour 12	0.000502	-0.127***	-8.404	-4.232***
	(0.03)	(-8.31)	(-1.04)	(-7.58)
hour 13	0.00918	-0.139***	-13.91	-4.537***
	(0.32)	(-8.58)	(-1.65)	(-7.72)
hour 14	-0.0203	-0.145***	-9.566	-4.795***
	(-1.42)	(-8.08)	(-1.04)	(-7.60)
hour 15	-0.0201	-0.155***	-14.65	-5.054***
	(-1.14)	(-7.94)	(-1.35)	(-7.38)
hour 16	-0.0257	-0.158***	-12.44	-5.179***
	(-1.73)	(-8.88)	(-1.13)	(-8.39)
hour 17	-0.0278*	-0.137***	-7.773	-4.561***
	(-2.14)	(-7.97)	(-0.77)	(-7.54)
hour 18	-0.0151	-0.125***	4.003	-4.282***
	(-1.20)	(-7.32)	(0.43)	(-7.14)
hour 19	0.0150	-0.113***	25.17**	-4.008***
	(1.21)	(-7.82)	(3.05)	(-7.23)
hour 20	0.0415**	-0.104***	26.37***	-3.752***
	(3.15)	(-7.59)	(3.62)	(-7.07)
hour 21	0.0659***	-0.0921***	20.90**	-3.406***
	(3.36)	(-7.15)	(3.16)	(-6.57)
hour 22	0.111***	-0.107***	29.11***	-3.908***
	(4.89)	(-6.98)	(4.15)	(-7.09)
hour 23	0.127***	-0.0929***	29.04***	-3.555***
	(5.97)	(-7.45)	(4.35)	(-6.99)

Note: Own calculations. The table presents outcomes of running the main analysis for each hour of the day separately from 12 a.m. (hour 12) until 11 p.m. (hour 23). All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, an indicator for a change in speed limits, and an interaction term between a station identifier and a post-lockdown dummy. Columns 1-4 include all stations of the main sample, columns 5-8 exclude outlier stations as identified by leave-on-out analyses. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.6: Effects for cars and trucks

	Cars				Trucks			
	Volume		Speed		Volume		Speed	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Outcome in absolute values								
1(PU lane)	-52.26***	-10.70	-4.715***	-4.290***	-1.818	3.497	0.337	0.262
	(5.131)	(7.457)	(0.313)	(0.545)	(1.785)	(2.723)	(0.233)	(0.350)
N	1546526	1546526	1543307	1543307	1546526	1546526	1536225	1536225
R^2	0.735	0.748	0.752	0.767	0.546	0.559	0.558	0.570
Stations	215	215	215	215	215	215	215	215
Interaction	No	Yes	No	Yes	No	Yes	No	Yes
Panel B: Log-transformed outcome								
1(PU lane)	-0.0219***	-0.0112	-0.116***	-0.124***	-0.0499**	0.0249	0.0154*	0.00170
	(0.00827)	(0.0144)	(0.00880)	(0.0144)	(0.0214)	(0.0320)	(0.00858)	(0.0119)
N	1543329	1543329	1543307	1543307	1536335	1536335	1532624	1532624
R^2	0.622	0.643	0.579	0.601	0.601	0.618	0.489	0.500
Stations	215	215	215	215	215	215	215	215
Interaction	No	Yes	No	Yes	No	Yes	No	Yes

Note: Own calculations. The table presents the coefficients of the treatment effects of two-way fixed effects estimations for cars and trucks separately. Panel A shows the coefficients of interest with outcomes in absolute terms. Panel B shows the same for logged outcome variables. Even columns include an interaction term between a unique measuring station identifier and a post-lockdown dummy variable while uneven columns do not. All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, and an indicator for a change in speed limits. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. Standard errors are clustered at 1km × 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.7: Different lane samples

	Only 2 lanes		Only 3 lanes	
	(1)	(2)	(3)	(4)
	Volume	Speed	Volume	Speed
1(PU lane)	0.00195	-0.132***	-0.0253	-0.0951***
	(0.0156)	(0.0200)	(0.0176)	(0.0156)
N	1282105	1282065	261388	261388
R^2	0.623	0.614	0.699	0.537
Stations treated	16	16	7	7
Stations Overall	180	180	35	35
Interaction	Yes	Yes	Yes	Yes

Note: Own calculations. The table presents the coefficients of the treatment effects of two-way fixed effects estimations on vehicle volume and speed. Columns 1 and 2 only include two-lane streets in the sample. Columns 3 and 4 only include three-lane streets. All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, an indicator for a change in speed limits, and an interaction term between a station identifier and a post-lockdown dummy. Stations within a one kilometre radius of a treated street are excluded. Standard errors are clustered at 1km × 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.8: Spillover effects on surrounding stations

	750m		1km		1.5km	
	(1) Volume	(2) Speed	(3) Volume	(4) Speed	(5) Volume	(6) Speed
1(PU lane)	0.0809***	0.00526	0.0946***	-0.00740	0.0177	-0.00779
	(0.0249)	(0.0109)	(0.0202)	(0.00809)	(0.0161)	(0.00759)
N	2416577	2416351	2416577	2416351	2416577	2416351
R^2	0.798	0.621	0.798	0.621	0.798	0.621
Stations treated	54	54	84	84	129	129
Stations Overall	343	343	343	343	343	343
Interaction	Yes	Yes	Yes	Yes	Yes	Yes

Note: Own calculations.The table presents the coefficients of the treatment effects of two-way fixed effects estimations on vehicle volume and speed for streets surrounding the actually treated ones. Streets within a radius of 750m, 1km, and 1.5km respectively are now considered as treated units. Streets that actually received a PUBL are excluded from the estimations. All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, an indicator for a change in speed limits, and an interaction term between a station identifier and a post-lockdown dummy. Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.9: Long-term results

	All Stations		No outlier stations	
	(1) Volume	(2) Speed	(3) Volume	(4) Speed
1(PU lane)	-0.0268*	-0.130***	-0.0363**	-0.0932***
	(0.0140)	(0.0142)	(0.0158)	(0.0145)
N	1690820	1690780	1673979	1673939
R^2	0.647	0.578	0.647	0.577
Stations	215	215	213	213
Interaction	Yes	Yes	Yes	Yes

Note: Own calculations. The table presents the coefficients of the treatment effects of separate two-way fixed effects estimations on vehicle volume and speed for a sample including the months from March until May of 2021. Columns 1 and 2 include all stations of the main sample, columns 3 and 4 exclude outlier stations as identified by leave-on-out analyses. All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, an indicator for a change in speed limits, and an interaction term between a station identifier and a post-lockdown dummy. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.10: Different standard error clusters

	Station*week		Districts*week		Grids		Week	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Outcomes with all stations								
1(PU lane)	-0.00540	-0.122***	-0.00540	-0.122***	-0.00540	-0.122**	-0.00540	-0.122***
	(0.0115)	(0.0131)	(0.0138)	(0.0142)	(0.0301)	(0.0469)	(0.0103)	(0.00791)
N	1543493	1543453	1543493	1543453	1543493	1543453	1543493	1543453
R^2	0.650	0.607	0.650	0.607	0.650	0.607	0.650	0.607
Stations treated	23	23	23	23	23	23	23	23
Stations Overall	215	215	215	215	215	215	215	215
Interaction	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outcome	Volume	Speed	Volume	Speed	Volume	Speed	Volume	Speed
Panel B: Outcomes w/o outliers								
1(PU lane)	-0.00590	-0.0863***	-0.00590	-0.0863***	-0.00590	-0.0863**	-0.00590	-0.0863***
	(0.0122)	(0.0117)	(0.0141)	(0.0108)	(0.0314)	(0.0346)	(0.0107)	(0.00713)
N	1536042	1536002	1536042	1536002	1536042	1536002	1536042	1536002
R^2	0.650	0.607	0.650	0.607	0.650	0.607	0.650	0.607
Stations treated	23	23	23	23	23	23	23	23
Stations Overall	214	214	214	214	214	214	214	214
Interaction	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outcome	Volume	Speed	Volume	Speed	Volume	Speed	Volume	Speed

Note: Own calculations. The table presents the coefficients of the treatment effects of separate two-way fixed effects estimations with varying clusters of standard errors as described in the column headers. Panel A shows the coefficients of interest with all stations from the main sample. Panel B shows outcomes with outlier stations as identified by leave-on-out analyses being excluded from the sample. All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, an indicator for a change in speed limits, and an interaction term between a station identifier and a post-lockdown dummy. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.11: Results with different samples

	Full sample & hours		Full sample		Cycling Infr.		Highways	
	(1) Volume	(2) Speed	(3) Volume	(4) Speed	(5) Volume	(6) Speed	(7) Volume	(8) Speed
1(PU lane)	0.0206	-0.102***	-0.0303**	-0.116***	-0.0107	-0.145***	0.0251**	-0.0935***
	(0.0139)	(0.0126)	(0.0143)	(0.0144)	(0.0135)	(0.0151)	(0.0123)	(0.0141)
N	2980500	2980207	2002402	2002210	718615	718609	2962999	2962168
R^2	0.849	0.621	0.786	0.630	0.678	0.720	0.847	0.715
Stations treated	24	24	24	24	19	19	23	23
Stations Overall	283	283	283	283	102	102	450	450
Interaction	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Own calculations. The table presents the coefficients of the treatment effects of separate two-way fixed effects estimations with varying sample compositions. Columns 1 & 2 show results with all 24 hours of the day and the full sample including one and four-lane streets except for stations within a 1km radius of treated streets. Columns 3 & 4 make the same restrictions, but now only with times between 5 a.m. and 8 p.m. In columns 5 & 6 the sample is restricted to streets without bike lanes prior to treatment. Outcomes in columns 7 & 8 include observations from highways. All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, an indicator for a change in speed limits, and an interaction term between a station identifier and a post-lockdown dummy. Standard errors are clustered at 1km × 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.12: Placebo tests

	Jan2020		Aug2019		Apr2019	
	(1) Volume	(2) Speed	(3) Volume	(4) Speed	(5) Volume	(6) Speed
1(PU lane)	-0.00917	-0.0106	-0.00963	-0.00731	-0.0180**	0.00867
	(0.00817)	(0.00654)	(0.00708)	(0.00525)	(0.00724)	(0.00640)
N	948552	948543	948552	948543	948552	948543
R^2	0.650	0.606	0.650	0.606	0.650	0.606
Stations treated	23	23	23	23	23	23
Stations Overall	215	215	215	215	215	215

Note: Own calculations. The table presents the coefficients of the treatment effects of separate two-way fixed effects estimations with treatment being simulated at different points in time. All estimations include station fixed effects (FE), date FE, hour FE, a dummy whether construction work takes place, and an indicator for a change in speed limits. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. Standard errors are clustered at 1km × 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.13: Variation in construction work control

	Volume		Speed	
	(1)	(2)	(3)	(4)
1(PU lane)	-0.00540	-0.00592	-0.122***	-0.122***
	(0.0126)	(0.0128)	(0.0144)	(0.0144)
Construction work control:				
Dummy variable	-0.0445***		-0.0487***	
	(0.00652)		(0.00480)	
Categorical variable		-0.00722***		-0.00810***
		(0.00152)		(0.000942)
N	1543493	1543493	1543453	1543453
R^2	0.650	0.650	0.607	0.604
Stations	215	215	215	215
Interaction	Yes	Yes	Yes	Yes

Note: Own calculations. The table compares the coefficients of the treatment effects of separate two-way fixed effects estimations with variations of the construction work control. Uneven columns include a construction work dummy independent from type of construction work. Even columns show results with a construction work variable, which explicitly controls for type of construction work. All estimations include station fixed effects (FE), date FE, hour FE, an indicator for a change in speed limits, and an interaction term between a station identifier and a post-lockdown dummy. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.14: Heterogeneous timing

	Volume		Speed	
	(1)	(2)	(3)	(4)
Panel A: Outcomes with all stations				
1(PU lane)	-0.0378***	-0.00818	-0.127***	-0.117***
	(0.00808)	(0.0274)	(0.00991)	(0.0400)
N	1328753	1328753	1328739	1328739
R^2	0.626	0.652	0.588	0.613
Stations treated	23	23	23	23
Stations Overall	215	215	215	215
Interaction	No	Yes	No	Yes
Panel B: Outcomes w/o outliers				
1(PU lane)	-0.0353***	-0.0138	-0.0990***	-0.0907***
	(0.00831)	(0.0275)	(0.00876)	(0.0333)
N	1322325	1322325	1322311	1322311
R^2	0.626	0.652	0.589	0.612
Stations treated	22	22	22	22
Stations Overall	214	214	214	214
Interaction	No	Yes	No	Yes

Note: Own calculations. The table presents the coefficients of the treatment effects of separate two-way fixed effects estimations for two samples with homogeneous treatment timing. Therefore, all observations between the first and the last installation date of a PUBL are deleted from the sample. *Panel A* shows results for all treatment stations. *Panel B* excludes Station 15 from Table 5.B.2. Even columns include an interaction term between a unique measuring station identifier and a post-lockdown dummy variable while uneven columns do not. All estimations include station fixed effects (FE), date FE, hour FE, and an indicator for a change in speed limits. Stations within a one kilometre radius of a treated street are excluded and the sample is restricted to streets with two or three lanes. Standard errors are clustered at 1km $\times$ 1km grid cell level spanning the city times a running week variable. t statistics in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.B.15: Predictor balance (Friedrichshain-Kreuzberg)

	Treated	Synthetic
Traffic space	.2637255	.1923878
Space settlements	.6705883	.6443015
Population	290083	290159.5
Election participation	.773	.743195
Green Party Voters	21.28861	16.09941
Unemp. rate	9.7	9.403
Accidents 2018	1666	1693.033
Accidents 2019	1675	1709.326
Accidents 2020	1408	1405.261

Note: The table shows the predictor balance between treated unit and synthetic control group of the synthetic control group method with Friedrichshain-Kreuzberg as treated district.

Table 5.B.16: Predictor balance (Pankow)

	Treated	Synthetic
Traffic space	.1331137	.1409323
Space settlements	.5376865	.5178513
Population	409454	409117
Election participation	.793	.80261
Green Party Voters	14.6013	13.7961
Unemp. rate	9.7	8.8952
Accidents 2018	1578	1591.279
Accidents 2019	1552	1518.664
Accidents 2020	1350	1409.146

Note: The table shows the predictor balance between treated unit and synthetic control group of the synthetic control group method with Pankow as treated district.

Table 5.B.17: Predictor balance (Charlottenburg-Wilmersdorf)

	Treated	Synthetic
Traffic space	.1941568	.1835291
Space settlements	.5084248	.5557473
Population	342950	342976.5
Election participation	.786	.750991
Green Party Voters	15.35358	14.92943
Unemp. rate	9.7	9.3712
Accidents 2018	2013	2018.721
Accidents 2019	2048	2027.964
Accidents 2020	1644	1660.739

Note: The table shows the predictor balance between treated unit and synthetic control group of the synthetic control group method with Charlottenburg-Wilmersdorf as treated district.

Table 5.B.18: Predictor balance (Treptow-Köpenick)

	Treated	Synthetic
Traffic space	.0966434	.1098114
Space settlements	.3503845	.3659922
Population	273817	273322
Election participation	.766	.727015
Green Party Voters	7.744833	8.189585
Unemp. rate	9.7	9.2366
Accidents 2018	1115	1155.677
Accidents 2019	1119	1140.264
Accidents 2020	1128	1092.36

Note: The table shows the predictor balance between treated unit and synthetic control group of the synthetic control group method with Treptow-Köpenick as treated district.

Table 5.B.19: Predictor balance (Neukölln)

	Treated	Synthetic
Traffic space	.1684843	.1846853
Space settlements	.8032495	.7053794
Population	328666	327763.5
Election participation	.708	.757145
Green Party Voters	12.85852	11.76067
Unemp. rate	9.7	9.6853
Accidents 2018	1230	1266.639
Accidents 2019	1289	1287.227
Accidents 2020	1116	1121.362

Note: The table shows the predictor balance between treated unit and synthetic control group of the synthetic control group method with Neukölln as treated district.

Chapter 6

Conclusion and Outlook

This dissertation studies agglomeration diseconomies and policies aimed to solve them. The first part of the thesis consists of two chapters showing that poor air quality is a non-negligible cost in urban agglomerations, which is true on a worldwide scale. In the second part the focal point is then put on policy interventions that aim to transform urban life in order to reduce costs such as air pollution, traffic accidents, or congestion. With more than half of the global population living in cities, benefiting from their advantages but also suffering from their costs, the chapters of this dissertation contribute to important questions of overall welfare and its future development. In this concluding chapter I discuss limitations to the analyses and suggest avenues for future research. Subsequently, the findings are put into perspective across chapters.

The first two chapters of the thesis determine the contemporaneous effect of population density and agglomeration sizes on air pollution as specific cost factor. While the benefits of metropolitan areas, such as productivity advantages, better worker-firm matches or knowledge spillovers are well documented (e.g. Duranton and Puga, 2004), the economic costs are relatively unexplored.

In this regard, ***Chapter 2***, by means of IV estimates, causally identifies that air quality decreases with increasing population density with an elasticity of 0.12. Consequently, densely populated areas are more polluted, even though residents may produce lower per-capita emissions. This implies that urban residents are exposed to higher health risks than residents of sparsely populated regions. The study explains the possible causes for the results and elaborates that total commuting and household energy consumption are likely to have the greatest impact on local pollution, which is in accordance with a theoretical model presented. These findings suggest that city authorities may be well advised to consider sustainable public transport options and energy-efficient buildings in urban planning to minimize air pollution and associated health costs. The backdrop of the analysis is that it considers only one country, i.e. Germany. Since the historical developments and institutional characteristics are unequal across countries, and thus the mechanisms influencing the results may differ, the findings of *Chapter 2* are not readily generalizable to other countries and their specific circumstances.

Chapter 3 aims to address this issue by conducting estimates for the entire world and showing heterogeneities between countries. It relies on global data derived from satellite measurements, which provides a truly holistic perspective. The main results document considerable heterogeneity, e.g. across country income groups and continents, with results from *Chapter 2* being higher than the global average. The effects are largest in Asia and middle income countries. Another interesting finding is that the size of metropolitan areas, and in particular the number of commuters, is more strongly correlated with pollution exposure than population density. Sprawling cities thus appear to suffer more from local pollution and are more likely to contribute to it. Inferring from this, urban planners seem to have another lever to improve urban air quality, and that is dense rather than sprawling development. In a counterfactual simulation we moreover demonstrate that countries with large urban-rural differences in pollution may want to provide incentives for citizens to more equally spread across cities in order to increase overall social welfare. Despite efforts to incorporate causal estimates using an IV approach that exploits global historical population data as an instrument, these results are rather suggestive in nature given the small subsample we are able to instrument. Consequently, more research on causal effects for different countries in different states of development is required to complete the picture.[1] Such single-country studies would also allow to better elaborate on the channels that cause urban areas to be more polluted, given that more detailed city-level data is available. Importantly, the focus of the first part of the dissertation is on local pollution with its health consequences, but it does not deal

[1]By the time this dissertation was submitted, there was only one other study available as a working paper that examined the effect of population density on air quality in the US (Carozzi and Roth, 2020).

with global pollution such as CO_2, which is relevant for global warming. For the evaluation of the role of metropolitan areas for the environment, such considerations should be kept in mind.

The second part of the thesis is devoted to particular mechanisms that offer scope for improving urban life, for example in the form of better air quality. To this end, the impact of policies aimed at making predominantly metropolitan areas more sustainable and environmentally friendly is analysed. The policy measures examined target the transport sector, more specifically the use of public transport and cycling. By analysing these, the two chapters directly reference the mechanisms discussed in previous chapters that identified commuting as a major contributor to environmental problems in cities.

Chapter 4 finds that a substantial reduction in public transportation fares decreases air pollution, as measured by the AQI, by more than eight percent. The effects are particularly high in urban areas and those with a well-developed public transport network. Consequently, commuters appear to be diverted away from automobile travel as the price of alternative transportation modes decreases. This finding is in line with prior research about quantity related channels that make public transport more attractive and positively affect air quality (e.g. Bauernschuster et al., 2017; Gendron-Carrier et al., 2022). Accordingly, reducing fares seems to be a viable alternative for policymakers to expanding the route network in order to encourage people to use public transportation, especially in spatially limited metropolitan areas. However, the examined ticket fare reduction was only temporary and abolished after three months. Commuters may find switching to public transportation tolerable for a short period of time, but change their behaviour in the long run, for example, due to overcrowding at a given supply. Therefore, the question about mid- and long-term developments of a price reduction remains. Connected to this, an open question is to what extent a combination of ticket price reduction and increased public transport supply, for instance in the form of a higher frequency of trains, can deter commuters from using private vehicles and consequently improve air quality. Since some regions have reduced fares by as much as 90 percent, it would also be intriguing to know the impact of more moderate ticket price subsidies. As the expansion of public transportation services has accelerated in recent years[2] and the German

[2] see https://bmdv.bund.de/DE/Themen/Mobilitaet/OEPNV/Oeffentlicher-Personenverkehr-Kompakt/oeffentlicher-personenverkehr-kompakt.html.

federal government has announced plans to introduce a 49-euro ticket in 2023[3], there is continued leeway to explore these matters.

Finally, ***Chapter 5*** analyses the effects of bicycle lanes on congestion and traffic volumes. It thus examines a different type of environmentally friendly infrastructure on mechanisms that tend to cause air pollution. The results indicate a reduction in average vehicle speed and thus an increase in congestion on roads with new bicycle infrastructure by 8-12 percent and up to 16 percent during peak hours, compared to roads without changes. Since no effect on vehicle volumes was identified, the overall impact can be attributed to the reduction in space available for automobiles. Therefore, installing these new bike lanes at the expense of a car lane increased the costs of driving on these roads, but possibly lowered air quality given the documented correlation between congestion and local pollution in the literature (Beaudoin et al., 2015). Future studies should address the as yet unanswered question of the extent to which this result actually translates into changes in air pollution. Another subject analysed in the study that is important for road dynamics and costs are road accidents. While no impact on the absolute number of traffic collisions was found in this chapter, the question about the accident rate per cyclist could not be answered conclusively due to the lack of cyclist data. Detailed road-specific bicycle use data could address this shortcoming. One overall limitation of the analysis is that it examines short- to medium-term changes in road dynamics, while long-term results that can shed light on general equilibrium adjustments would be a fruitful addition to future research. Moreover, the chapter only considers the case study of Berlin. Whether and to what extent the results change in other circumstances, such as in urban areas with different public transportation options, remains an open question.

In summary, ***Chapters 2*** and ***3*** established that air pollution is a substantial cost factor for metropolitan areas, but subject to heterogeneities, while ***Chapters 4*** and ***5*** evaluate policy measures that have the potential to shape cities by addressing these same and other costs. The combination of the latter findings shows that it will probably take more than one measure to actually change cities in a sustainable way and make people in urban areas still "greener, healthier, and happier," as Edward Glaeser has phrased it (Glaeser, 2011). While more bike lanes are likely to alleviate the problem of accidents on city streets, motorists arguably need cheaper and better public transit alternatives to truly have incentives to

[3]see https://www.bundesregierung.de/breg-de/aktuelles/deutschlandticket-2134074.

switch their mode of transportation. Future studies will have to show whether these results will be assessed differently in the long term and to what extent further developments, such as the expansion and promotion of electro-mobility, can contribute to more sustainable cities.

List of Tables

List of Figures

List of Abbreviations

9ET 9-Euro-Ticket

A.D. Anno Domini

AOD Aerosol Optical Depth

API Application Programming Interface

AQI Air Quality Index

B.C. Before Christ

CBD Central Business District

CBSA Core-Based Statistical Area

CDU Christlich Demokratische Union

CI Confidence Interval

CO_2 Carbon dioxide

DiD Difference-in-Differences

DWD Deutscher Wetterdienst

ESDB European Soil Database

EU European Union

FEA Federal Environmental Agency

FDP Freie Demokratische Partei

FE Fixed Effects

FHKB Friedrichshain-Kreuzberg

FUA Functional Urban Area

GADM Global Administrative Areas

GDP Gross Domestic Product

GWR Geographically Weighted Regression

GIS Geographic Information System

GPW Gridded Population of the World (version 4)

IV Instrumental Variable

LMR Labour Market Region

LD Long Differences

LEZ Low Emission Zones

LCIA Local Conditional Independence Assumption

LPM Linear Probability Model

MSA Metropolitan Statistical Area

$\mu g/m^3$ Microgram per cubic meter

NO Nitrogen Monoxide

NO_2 Nitrogen Dioxide

NO_X Nitrogen Oxides

OECD Organisation for Economic Co-operation and Development

OLS Ordinary Least Squares

OSM Open Street Maps

O_3 Ozone

PM_{10} Particulate matters with aerodynamic diameter of 10 micrometer or less

$PM_{2.5}$ Particulate matters with aerodynamic diameter of 2.5 micrometer or less

PM_x Particulate matters of any size

PUBL Pop-Up Bike Lane

SE Standard Errors

SFD Spatial First Differences

SO_2 Sulphur Dioxide

SPD Sozialdemokratische Partei Deutschlands

SUTVA Stable Unit Treatment Value Assumption

TWFE Two-Way Fixed Effects

US United States

UN United Nations

VOC Volatile Organic Compounds

WHO World Health Organization

German Summary

Die vorliegende Dissertation stellt empirische Untersuchungen über den Zusammenhang von städtischem Leben und dessen ökonomische Kosten, insbesondere für die Umwelt, an. Dabei werden zum einen bestehende Forschungslücken des Einflusses von Bevölkerungsdichte auf die Luftqualität geschlossen und zum anderen innovative Politikmaßnahmen im Verkehrsbereich untersucht, die Ballungsräume nachhaltiger gestalten sollen. Im Zentrum der Betrachtungen stehen Luftverschmutzung, Staus und Verkehrsunfälle, die für Fragen der allgemeinen Wohlfahrt bedeutend sind und erhebliche Kostenfaktoren für urbanes Leben darstellen. Von ihnen ist ein beträchtlicher Anteil der Weltbevölkerung betroffen. Während im Jahr 2018 bereits 55% der Menschen weltweit in Städten lebten, soll dieser Anteil bis zum Jahr 2050 ungefähr 68% betragen (United Nations, 2019).

Die vier in sich geschlossenen Kapitel dieser Arbeit lassen sich in zwei Abschnitte aufteilen: In den Kapiteln 2 und 3 werden neue kausale Erkenntnisse über das komplexe Zusammenspiel von städtischen Strukturen und Luftverschmutzung erbracht. Kapitel 4 und 5 untersuchen anschließend politische Maßnahmen zur Förderung nicht-motorisierter Verkehrsmittel und deren Einfluss auf Luftqualität sowie Staugeschehen und Verkehrsunfälle.

Kapitel 2 analysiert den kausalen Zusammenhang zwischen dicht besiedelten Ballungsräumen und Luftqualität. Es handelt sich um die erste Publikation, die eine Antwort auf diese Frage gibt. Theoretisch gibt es zwei Möglichkeiten für die Wirkungsrichtung des Gesamteffekts: Einerseits verursachen mehr Menschen höhere Emissionen, beispielsweise durch das Pendeln zur Arbeit oder den Energieverbrauch beim Wohnen. Andererseits haben Menschen in Ballungsräumen, z.B. aufgrund kürzerer Pendelwege, vermehrt die Möglichkeit umweltfreundliche Verkehrsmittel wie Fahrräder oder den Öffentlichen Nahverkehr zu nutzen sowie in, im Vergleich zu Einfamilienhäusern, energieeffizienteren Hochhäusern zu leben.

Die Schätzungen der kausalen Effekte basieren auf aktuellen in Deutschland erhobenen Daten unter Nutzung einer Reihe ökonometrischer Methoden. So berücksichtigen *Long Difference* und *Fixed Effects* Analysen, im Vergleich zur einfachen Kleinste-Quadrate-Schätzung, nicht beobachtbare Variation, die zeitlich konstant ist. Die in der Studie präferierte Schätzgröße entstammt der Analyse anhand der *Instrumental Variable* Methode. Dabei müssen die gewählten Instrumente einen bedeutenden Einfluss auf die erklärende Variable, in unserem Falle Bevölkerungsdichte, haben, ohne direkt mit der zu erklärenden Variable, also aktuelle Luftverschmutzung, korreliert zu sein. In der Untersuchung wird argumentiert, dass diese Eigenschaften sowohl auf historische Bevölkerungsdichte als auch auf die geologische Beschaffenheit des Untergrunds einer Stadt zutreffen.

Im Ergebnis nimmt die Luftverschmutzung, gemessen an einem Luftqualitätsindex (AQI), mit der Bevölkerungsdichte mit einer Elastizität von rund 0.12 zu. Dicht besiedelte Ballungsräume sind dementsprechend, trotz eventueller geringerer Pro-Kopf-Emissionen, verschmutzter und die dort lebenden Menschen somit einem höheren Gesundheitsrisiko ausgesetzt. Die empirischen Ergebnisse stehen im Einklang mit einer Erweiterung des monozentrischen Stadtmodells, das in der Studie präsentiert wird.

Kapitel 3 ergänzt die vorherige Studie um eine Analyse von Bevölkerungsindikatoren und Luftqualität anhand weltweiter Daten. Im Vergleich zu Untersuchungen, die auf einzelnen Ländern basieren, bietet eine globale Analyse eine ganzheitliche Sicht: Probleme wie geografische oder institutionelle Eigenheiten einzelner Länder, wie beispielsweise der nationalen Umweltpolitik, können somit adressiert werden. Ein Hauptanliegen der Arbeit ist die Frage, inwiefern sich das Stadt-Land-Gefälle von Luftverschmutzung in verschiedenen Ländern unterscheidet. Ein weiterer Schwerpunkt liegt auf den Unterschieden zwischen dicht bebauten und zersiedelten Städten.

Deskriptive Analysen zeigen, dass etwa 75 Prozent der Weltbevölkerung Feinstaubkonzentrationen ausgesetzt sind, die von der WHO als gesundheitsschädlich eingestuften werden. Die Mehrheit der betroffenen Menschen lebt dabei in Städten. Anhand länderspezifischer Untersuchungen schätzen wir eine durchschnittliche Elastizität der Bevölkerungsgröße zu $PM_{2.5}$ von etwa 0,03 und zu NO_2 von 0,16. In Kombination mit den Ergebnissen aus Kapitel 2 bedeutet dies, dass das Stadt-Land-Gefälle von Luftverschmutzung in Deutschland über dem weltweiten Durchschnitt liegt. Wir stellen außerdem fest, dass die Effekte auf globaler Ebene eher durch die Größe des Ballungsraums als durch die Bevölkerungsdichte bestimmt werden. Außerdem wird das Gefälle eher durch größere Pendelgebiete als durch die Bevölkerung in den städtischen Zentren beeinflusst. Aus diesen Ergebnissen lässt sich ableiten, dass eine Abkehr vom Vorstadtmodell und eine Verdichtung von Städten einen Beitrag zur Luftverbesserung leisten kann und somit gesundheitsförderlich wäre.

Die Kapitel 4 und 5 der Dissertation widmen sich politischen Maßnahmen und deren Auswirkungen auf Luftqualität oder damit verbundenen Themen. So stellt sich beispielsweise die Frage, wie Menschen dazu bewegt werden können, vom Auto auf öffentliche Verkehrsmittel umzusteigen, um externe Effekte motorisierten Individualverkehrs in Form von Luftverschmutzung zu verringern. **Kapitel 4** leistet zu diesem Thema einen Beitrag, indem es eine groß angelegte Preissubvention für öffentliche Verkehrsmittel und deren Auswirkung auf Luftqualität untersucht. Die Einführung des 9-Euro-Tickets (9ET), das die monatlichen Preise für öffentliche Verkehrsmittel substantiell verringerte, dient hierbei als quasi-natürliches Experiment. Als Analysemethode dieser großen und landesweiten Preisanpassung dient der *Difference-in-Differences* (DiD)–Ansatz. Die *treatment group* ist der Monat Juni, der im Jahr 2022 im Gegensatz zu den vorangegangenen Jahren die Preisanpassung erfuhr. Der Monat Mai bildet die *control group*, da er ähnliche Merkmale wie der Juni aufweist. Die implizite Annahme ist, dass die Entwicklung der Luftqualität zwischen Mai und Juni 2022 im Vergleich zu den Vorjahren ohne Preisveränderung die gleiche gewesen wäre.

Im Ergebnis zeigt die Studie eine Verbesserung der Luftqualität, gemessen anhand des AQI, um etwa acht Prozent. Am größten ist der Effekt in städtischen Gebieten, an Werktagen und in Gegenden, in denen das öffentliche Verkehrsmittelnetz gut ausgebaut ist. Die Schätzungen ergeben zudem, dass der Effekt im Laufe der Zeit leicht abnimmt und dass die Luftverschmutzung nach Auslaufen der Maßnahme wieder zunimmt. Die Untersuchung schlussfolgert auf Grundlage der Schätzungen und unter Verwendung früherer Erkenntnisse über den Zusammenhang zwischen Luftqualität und Gesundheit, dass ein ermäßigter Fahrpreis für öffentliche Verkehrsmittel zu einer Verbesserung gesundheitsbezogener Ergebnisse führt. Diese haben das Potenzial, die tatsächlichen Kosten der Maßnahme zu amortisieren.

Obwohl Luftqualität im letzten Abschnitt der Arbeit, **Kapitel 5**, nicht gesondert betrachtet wird, ist dessen hauptsächlicher Analysegegenstand, nämlich Staus, eng damit verbunden. Sie führen neben ökonomisch bedeutenden Zeitverlusten für Pendler, zu verschwenderischem Kraftstoffverbrauch und in der Folge zu einem Anstieg von Emissionen. Zudem werden Verkehrsunfälle betrachtet, die darüber hinaus erhebliche externe Kosten verursachen. Untersuchungsgegenstand der Studie ist die Umwandlung von Autospuren in Fahrradwege und deren Folgen für das Geschehen auf den Straßen Berlins. Trotz der weltweit zunehmenden Bedeutung der Fahrradnutzung in großen Städten, befindet sich die Forschung zu den Auswirkungen infrastruktureller Maßnahmen für Fahrräder noch in den Anfängen. Aufgrund ihrer einfachen Implementierung und geringen Kosten ist eine solche Umwidmung von Straßen ein besonders wichtiges Forschungsobjekt, da sie in Städten mit wenig Platz eine praktikable Möglichkeit für den Bau neuer Fahrradinfrastruktur bietet.

Mein Forschungsansatz nutzt den Umstand, dass ab März 2020 in Berlin auf mehreren Straßen sogenannte Pop-Up Bike Lanes (PUBLs) installiert wurden, während der Großteil der Stadt keine neue Fahrradinfrastruktur erhielt. Aufgrund der sehr spontanen Umsetzung der Maßnahmen, die in Zeiten eines Lockdowns kurz nach Beginn der COVID-19 Pandemie stattfanden, war sowohl die zeitliche als auch räumliche Zuordnung der PUBLs so gut wie zufällig. Alle nicht zufälligen Faktoren, wie z.B. eine Mindestanzahl an vorhandenen Fahrspuren als Bedingung für die Installation, werden in der Studie adressiert. Zur Bewertung kausaler Effekte nutze ich als Methoden Variationen des klassischen *Difference-in-Differences*–Ansatzes, nämlich *Two-Way Fixed Effects*, ein *Event Study Design* und die *Synthetic Control Group* Methode.

Die Ergebnisse deuten auf eine Zunahme von Staus hin, die in den Hauptverkehrszeiten besonders hoch ist. Da sich das Fahrzeugaufkommen auf Straßen mit PUBL im Vergleich zu solchen ohne nicht signifikant unterscheidet, liegt der Schluss nahe, dass Autofahrer im untersuchten Zeitraum keinen Anlass sahen, auf andere Verkehrsmittel umzusteigen oder andere Wege zu nutzen. Die Staus werden dementsprechend durch den verringerten Platz für Autos verursacht. Gemäß Studienergebnissen kam es zu keinem Anstieg von Verkehrsunfällen durch die Maßnahme. Wird allerdings von einer Zunahme des Radverkehrs auf PUBL-Straßen ausgegangen, was von anderer Forschungsliteratur nahegelegt wird, so hieße das eine Verbesserung der Unfallraten pro Radfahrer. Insgesamt ist also die Umnutzung von

Autospuren in Fahrradwege mit Kosten verbunden, allerdings scheinen diese im Ergebnis moderat auszufallen.

Bibliography

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association 105*(490), 493–505.

Achakulwisut, P., M. Brauer, P. Hystad, and S. C. Anenberg (2019). Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO_2 pollution: estimates from global datasets. *The Lancet Planetary Health 3*(4), 166–178.

Ahlfeldt, G. M. and E. Pietrostefani (2019). The economic effects of density: A synthesis. *Journal of Urban Economics 111*, 93–107.

Aldeco, L., L. Barrage, and M. A. Turner (2019). Equilibrium particulate exposure. Working paper. https://matthewturner.org/papers/unpublished/Aldeco_Barrage_Turner_unp2021.pdf. Accessed December 2022.

Anderson, H. R. (2009). Air pollution and mortality: A history. *Atmospheric Environment 43*(1), 142–152.

Anderson, M. L. (2014). Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review 104*(9), 2763–2796.

Atmospheric Composition Analysis Group (2018). Surface NO_2 and surface $PM_{2.5}$ worldwide. https://sites.wustl.edu/acag/datasets/. Atmospheric Composition Analysis Group, Dalhousi University. Accessed December 2022.

Auffhammer, M., S. M. Hsiang, W. Schlenker, and A. Sobel (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy 7*(2), 181–198.

Auffhammer, M. and R. Kellogg (2011). Clearing the air? The effects of gasoline content regulation on air quality. *American Economic Review 101*(6), 2687–2722.

Basso, L. J. and H. E. Silva (2014). Efficiency and substitutability of transit subsidies and other urban transport policies. *American Economic Journal: Economic Policy 6*(4), 1–33.

Bauernschuster, S., T. Hener, and H. Rainer (2017). When labor disputes bring cities to a standstill: The impact of public transit strikes on traffic, accidents, air pollution, and health. *American Economic Journal: Economic Policy 9*(1), 1–37.

Beaudoin, J., Y. H. Farzin, and C.-Y. C. L. Lawell (2015). Public transit investment and sustainable transportation: A review of studies of transit's impact on traffic congestion and air quality. *Research in Transportation Economics 52*, 15–22.

Bell, M. L. and D. L. Davis (2001). Reassessment of the lethal london fog of 1952: novel indicators of acute and chronic consequences of acute exposure to air pollution. *Environmental health perspectives 109*(suppl 3), 389–394.

Bezirksamt Friedrichshain-Kreuzberg (2020a). Pressemitteilung vom 25.03.2020: Temporäre Einrichtung und Erweiterung von Radverkehrsanlagen während der Pandemie-Krise. Online resource: https://www.berlin.de/ba-friedrichshain-kreuzberg/aktuelles/pressemitteilungen/2020/pressemitteilung.911780.php. Accessed December 2022.

Bezirksamt Friedrichshain-Kreuzberg (2020b). Pressemitteilung vom 25.05.2020: Einrichtung von pandemieresilienter Infrastruktur in Form von temporären Radverkehrsanlagen. Online resource: https://www.berlin.de/ba-friedrichshain-kreuzberg/aktuelles/pressemitteilungen/2020/pressemitteilung.937004.php. Accessed December 2022.

Bhuyan, P., E. J. McCoy, H. Li, and D. J. Graham (2021). Analysing the causal effect of london cycle superhighways on traffic congestion. *The Annals of Applied Statistics 15*(4), 1999–2022.

Borck, R. (2019). Public transport and urban pollution. *Regional Science and Urban Economics 77*, 356–366.

Borck, R. and J. K. Brueckner (2018). Optimal energy taxation in cities. *Journal of the Association of Environmental and Resource Economists 5*(2), 481–516.

Borck, R. and P. Mulder (2022). Energy policies and pollution in developing country cities. Unpublished manuscript.

Borck, R. and M. Pflüger (2019). Green cities? urbanization, trade, and the environment. *Journal of Regional Science 59*(4), 743–766.

Borck, R. and P. Schrauth (2021). Population density and urban air quality. *Regional Science and Urban Economics 86*, 103596.

Borck, R. and T. Tabuchi (2019). Pollution and city size: can cities be too small? *Journal of Economic Geography 19*(5), 995–1020.

Breidenbach, P. and L. Eilers (2018). RWI-GEO-GRID: Socio-economic data on grid level. *Jahrbücher für Nationalökonomie und Statistik 238*(6), 609–616.

Brimblecombe, P. (2012). *The big smoke (Routledge Revivals): A history of air pollution in London since medieval times.* Routledge.

Brownstone, D. and T. F. Golob (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics 65*(1), 91–98.

Brueckner, J. K. (1987). Chapter 20 the structure of urban equilibria: A unified treatment of the muth-mills model. In *Urban Economics*, Volume 2 of *Handbook of Regional and Urban Economics*, pp. 821–845. Elsevier.

Buehler, R. and J. Pucher (2012). Cycling to work in 90 large american cities: new evidence on the role of bike paths and lanes. *Transportation 39*(2), 409–432.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Carozzi, F. and S. Roth (2020). Dirty density: air quality and the density of american cities. IZA Discussion Papers Nr. 13191, Institute of Labor Economics (IZA), Bonn.

Castells-Quintana, D., E. Dienesch, and M. Krause (2021). Air pollution in an urban world: A global view on density, cities and emissions. *Ecological Economics 189*, 107153.

Chen, S., P. Oliva, and P. Zhang (2022). The effect of air pollution on migration: evidence from China. *Journal of Development Economics 156*, 102833.

Chen, Y. and A. Whalley (2012). Green infrastructure: The effects of urban rail transit on air quality. *American Economic Journal: Economic Policy 4*(1), 58–97.

Cheshire, P. and S. Magrini (2008, 10). Urban growth drivers in a Europe of sticky people and implicit boundaries. *Journal of Economic Geography 9*(1), 85–115.

Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *American Economic Review 86*(1), 54–70.

Clarke, D. and K. Tapia-Schythe (2021). Implementing the panel event study. *The Stata Journal 21*(4), 853–884.

Combes, P.-P., G. Duranton, and L. Gobillon (2018). The costs of agglomeration: House and land prices in french cities. *Review of Economic Studies 86*(4), 1556–1589.

Combes, P.-P., G. Duranton, L. Gobillon, and S. Roux (2010). Estimating agglomeration economies with history, geology, and worker effects. In E. L. Glaeser (Ed.), *Agglomeration economics*, pp. 15–66. Chicago: University of Chicago Press.

Combes, P.-P. and L. Gobillon (2015). Chapter 5 - the empirics of agglomeration economies. In G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5 of *Handbook of Regional and Urban Economics*, pp. 247–348. Elsevier.

Costa, S., J. Ferreira, C. Silveira, C. Costa, D. Lopes, H. Relvas, C. Borrego, P. Roebeling, A. I. Miranda, and J. Paulo Teixeira (2014). Integrating health on air quality assessment — Review report on health risks of two major european outdoor air pollutants: PM and NO_2. *Journal of Toxicology and Environmental Health, Part B 17*(6), 307–340.

Currie, J., S. H. Ray, and M. Neidell (2011). Quasi-experimental studies suggest that lowering air pollution levels benefits infants' and children's health. *Health Affairs 30*(12), 2391–2399.

Currie, J. and R. Walker (2011). Traffic congestion and infant health: Evidence from E-ZPass. *American Economic Journal: Applied Economics 3*(1), 65–90.

Davis, D. R. and D. E. Weinstein (2002). Bones, bombs, and break points: the geography of economic activity. *American Economic Review 92*(5), 1269–1289.

Davis, N., J. Badger, A. N. Hahmann, B. Ohrbeck, H. Bjarke, T. Olsen, N. Gylling Mortensen, D. Heathfield, M. Onninen, G. Lizcano, and O. Lacave (2019). Global wind atlas v3. https://doi.org/10.11583/DTU.9420803.v1. Technical University of Denmark.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–96.

de Thé, C. B., B. Carantino, and M. Lafourcade (2021). The carbon 'carprint' of urbanization: New evidence from french cities. *Regional Science and Urban Economics 89*, 103693.

Die Bundesregierung (2022). 9-Euro-Ticket 52 Millionen Mal verkauft. Online resource: https://www.bundesregierung.de/breg-de/suche/faq-9-euro-ticket-2028756. Accessed December 2022.

Dijkstra, L., H. Poelman, and P. Veneri (2022). The EU-OECD definition of a functional urban area. OECD Regional Development Working Papers, No. 2019/11, OECD Publishing, Paris.

DLR (2022). Wie hat das 9-Euro-Ticket unsere Mobilität verändert? Online resource: https://www.dlr.de/content/de/artikel/news/2022/03/20220825_wie-hat-das-9-euro-ticket-unsere-mobilitaet-veraendert.html. Accessed December 2022.

Druckenmiller, H. and S. Hsiang (2018). Accounting for unobservable heterogeneity in cross section using spatial first differences. NBER Working Paper 25177. National Bureau of Economic Research.

Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, Volume 4, pp. 2063–2117. Elsevier.

Duranton, G. and D. Puga (2019). Urban growth and its aggregate implications. NBER Working Paper 26591. National Bureau of Economic Research.

Duranton, G. and M. A. Turner (2011). The fundamental law of road congestion: Evidence from us cities. *American Economic Review 101*(6), 2616–2652.

DWD Climate Data Center (CDC) (2022). Daily station data. Online resource: https://cdc.dwd.de/portal/. Accessed December 2022.

Edlin, A. S. and P. Karaca-Mandic (2006). The accident externality from driving. *Journal of Political Economy 114*(5), 931–955.

Ewing, R., R. Pendall, and D. Chen (2003). Measuring sprawl and its transportation impacts. *Transportation Research Record: Journal of the Transportation Research Board 1831*, 175–183.

Fageda, X., R. Flores-Fillol, and B. Theilen (2022). Price versus quantity measures to deal with pollution and congestion in urban areas: A political economy approach. *Journal of Environmental Economics and Management 115*, 102719.

FAO/IIASA (2012). Global agro-ecological zones (gaez v3. 0).

Feng, M., J. O. Sexton, S. Channan, and J. R. Townshend (2016). A global, high-resolution (30-m) inland water body dataset for 2000: First results of a topographic–spectral classification algorithm. *International Journal of Digital Earth 9*(2), 113–133.

Firket, J. (1936). Fog along the meuse valley. *Transactions of the Faraday Society 32*, 1192–1196.

Fowler, D., P. Brimblecombe, J. Burrows, M. R. Heal, P. Grennfelt, D. S. Stevenson, A. Jowett, E. Nemitz, M. Coyle, X. Liu, et al. (2020). A chronology of global air quality. *Philosophical Transactions of the Royal Society A 378*(2183), 20190314.

Freeman, R., W. Liang, R. Song, and C. Timmins (2019). Willingness to pay for clean air in China. *Journal of Environmental Economics and Management 94*, 188–216.

Galdo, V., Y. Li, and M. Rama (2019). Identifying urban areas by combining human judgment and machine learning: An application to India. *Journal of Urban Economics 125*, 103229.

Geddes, J. A., R. V. Martin, B. L. Boys, and A. van Donkelaar (2015). Long-term trends worldwide in ambient NO_2 concentrations inferred from satellite observations. *Environmental Health Perspectives 124*(3), 281–289.

Gehrsitz, M. (2017). The effect of low emission zones on air pollution and infant health. *Journal of Environmental Economics and Management 83*, 121–144.

Gendron-Carrier, N., M. Gonzalez-Navarro, S. Polloni, and M. A. Turner (2022). Subways and urban air pollution. *American Economic Journal: Applied Economics 14*(1), 164–96.

Geoportal Berlin (2021). Traffic network - street segments. Online resource: https://daten.odis-berlin.de/de/dataset/detailnetz_strassenabschnitte/. Accessed December 2022.

German Federal Office for Statistics (2022). 9-Euro-Ticket: Mobilität steigt deutlich auf kurzen Distanzen im Schienenverkehr. Online resource: https://www.destatis.de/DE/Presse/Pressemitteilungen/2022/07/PD22_284_12.html. Accessed December 2022.

Gifford, R. and L. Steg (2007). The impact of automobile traffic on quality of life. In *Threats from car traffic to the quality of urban life*. Emerald Group Publishing Limited.

Glaeser, E. (2011). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*. Penguin.

Glaeser, E. L. (1998). Are cities dying? *Journal of Economic Perspectives 12*(2), 139–160.

Glaeser, E. L. and M. E. Kahn (2010). The greenness of cities: Carbon dioxide emissions and urban development. *Journal of Urban Economics 67*(3), 404–418.

Godish, T. and J. S. Fu (2019). *Air quality*. CRC Press.

Goodman, A., J. Panter, S. J. Sharp, and D. Ogilvie (2013). Effectiveness and equity impacts of town-wide cycling initiatives in england: a longitudinal, controlled natural experimental study. *Social Science & Medicine 97*, 228–237.

Gordon, P. and W. Cox (2012). Cities in Western Europe and the United States: Do policy differences matter? *The Annals of Regional Science 48*(2), 565–594.

Green, C. P., J. S. Heywood, and M. Navarro (2016). Traffic accidents and the London congestion charge. *Journal of public economics 133*, 11–22.

Green, C. P., J. S. Heywood, and M. N. Paniagua (2020). Did the London congestion charge reduce pollution? *Regional Science and Urban Economics 84*, 103573.

Greenstone, M., R. Hornbeck, and E. Moretti (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy 118*(3), 536–598.

Gudipudi, R., T. Fluschnik, A. G. C. Ros, C. Walther, and J. P. Kropp (2016). City density and CO_2 efficiency. *Energy Policy 91*, 352–361.

Hamilton, T. L. and C. J. Wichman (2018). Bicycle infrastructure and traffic congestion: Evidence from dc's capital bikeshare. *Journal of Environmental Economics and Management 87*, 72–93.

Hammer, M. S., A. van Donkelaar, C. Li, A. Lyapustin, A. M. Sayer, N. C. Hsu, R. C. Levy, M. J. Garay, O. V. Kalashnikova, R. A. Kahn, et al. (2020). Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). *Environmental Science & Technology 54*(13), 7879–7890.

Handelsblatt (2022). "Es gibt einen Run auf das 9-Euro-Ticket" – Sieben Millionen Tickets verkauft. Online resource: https://www.handelsblatt.com/politik/deutschland/28387430.html. Accessed December 2022.

Heblich, S., A. Trew, and Y. Zylberberg (2021). East-side story: Historical pollution and persistent neighborhood sorting. *Journal of Political Economy 129*(5), 1508–1552.

Henderson, J. V., D. Nigmatulina, and S. Kriticos (2021). Measuring urban economic density. *Journal of Urban Economics 125*, 103188.

Henderson, J. V., T. Squires, A. Storeygard, and D. Weil (2018). The global distribution of economic activity: Nature, history, and the role of trade. *The Quarterly Journal of Economics 133*(1), 357–406.

Hilber, C. and C. Palmer (2014). Urban development and air pollution: Evidence from a global panel of cities. GRI Working Papers 175, Grantham Research Institute on Climate Change and the Environment.

Hymel, K. (2009). Does traffic congestion reduce employment growth? *Journal of Urban Economics 65*(2), 127–135.

INRIX (2021). 2021 inrix global traffic scorecard. Online resource: https://inrix.com/scorecard/. Accessed December 2022.

Jaensirisak, S., M. Wardman, and A. May (2005). Explaining variations in public acceptability of road pricing schemes. *Journal of Transport Economics and Policy (JTEP) 39*(2), 127–154.

Jin, J. and P. Rafferty (2017). Does congestion negatively affect income growth and employment growth? empirical evidence from US metropolitan regions. *Transport Policy 55*, 1–8.

Kahn, M. E. (2010). New evidence on trends in the cost of urban agglomeration. In E. L. Glaeser (Ed.), *Agglomeration economics*, pp. 339–354. Chicago: University of Chicago Press.

Karathodorou, N., D. J. Graham, and R. B. Noland (2010). Estimating the effect of urban density on fuel demand. *Energy Economics 32*(1), 86–92.

Knittel, C. R., D. L. Miller, and N. J. Sanders (2016). Caution, drivers! children present: Traffic, pollution, and infant health. *Review of Economics and Statistics 98*(2), 350–366.

Koh, H.-J., N. Riedel, and T. Böhm (2013). Do governments tax agglomeration rents? *Journal of Urban Economics 75*, 92–106.

Kosfeld, R. and A. Werner (2012). Deutsche Arbeitsmarktregionen – Neuabgrenzung nach den Kreisgebietsreformen 2007–2011. *Raumforschung und Raumordnung 70*(1), 49–64.

Kraus, S. and N. Koch (2021). Provisional covid-19 infrastructure induces large, rapid increases in cycling. *Proceedings of the National Academy of Sciences 118*(15), e2024399118.

Kummu, M., M. Taka, and J. H. Guillaume (2018). Gridded global datasets for gross domestic product and human development index over 1990–2015. *Scientific data 5*, 180004.

Lagravinese, R., F. Moscone, E. Tosetti, and H. Lee (2014). The impact of air pollution on hospital admissions: Evidence from Italy. *Regional Science and Urban Economics 49*, 278–285.

Lalive, R., S. Luechinger, and A. Schmutzler (2018). Does expanding regional train service reduce air pollution? *Journal of Environmental Economics and Management 92*, 744–764.

Lamsal, L., R. Martin, D. D. Parrish, and N. A. Krotkov (2013). Scaling relationship for NO_2 pollution and urban population size: A satellite perspective. *Environmental Science & Technology 47*(14), 7855–7861.

Lamsal, L., R. Martin, A. Van Donkelaar, M. Steinbacher, E. Celarier, E. Bucsela, E. Dunlea, and J. Pinto (2008). Ground-level nitrogen dioxide concentrations inferred from the satellite-borne ozone monitoring instrument. *Journal of Geophysical Research: Atmospheres 113*(D16).

LandScan (2018). Global population distribution data. OAK Ridge National Laboratory. https://landscan.ornl.gov/. Accessed December 2022.

Larson, W. and A. Yezer (2015). The energy implications of city size and density. *Journal of Urban Economics 90*, 35–49.

Li, H., D. J. Graham, and P. Liu (2017). Safety effects of the London cycle superhighways on cycle collisions. *Accident Analysis & Prevention 99*, 90–101.

Lloyd, C. T., A. Sorichetta, and A. J. Tatem (2017). High resolution global gridded data for use in population studies. *Scientific data 4*, 170001.

Margaryan, S. (2021). Low emission zones and population health. *Journal of Health Economics 76*, 102402.

Mattauch, L., M. Ridgway, and F. Creutzig (2016). Happy or liberal? making sense of behavior in transport policy design. *Transportation research part D: transport and environment 45*, 64–83.

Mayer, H. (1999). Air pollution in cities. *Atmospheric environment 33*(24), 4029–4037.

Mobycon (2020). Making safe space for cycling in 10 days - A guide to temporary bike lanes from Friedrichshain-Kreuzberg, Berlin. https://mobycon.com/wp-content/uploads/2020/05/FrKr-Berlin_Guide-EN.pdf. Accessed December 2022.

Moreno-Monroy, A., L. Maffenini, M. Schiavina, and P. Veneri (2019). GHS-FUA R2019A - GHS functional urban areas, derived from GHS-UCDB R2019A (2015). http://data.europa.eu/89h/347f0337-f2da-4592-87b3-e25975ec2c95. Last access: May 2022.

Moreno-Monroy, A. I., M. Schiavina, and P. Veneri (2020). Metropolitan areas in the world. Delineation and population trends. *Journal of Urban Economics 125*, 103242.

Mücke, H.-G. (2014). Gesundheitliche Auswirkungen von atmosphärisch beeinflussten Luftverunreinigungen. In J. Lozán, H. Grassl, L. Karbe, and G. Jendritzky (Eds.), *Warnsignal Klima: Gefahren für Pflanzen, Tiere und Menschen*, Volume 3.1.3, pp. 1–7. GEO Wissenschaftliche Auswertungen, Hamburg.

NASA Earth Observatory (2015). Seeing through the smokey pall: Observations from an grim Indonesian fire season. https://earthobservatory.nasa.gov/features/IndonesianFires. Accessed December 2022.

Natural Earth (2018). Natural earth version 4.1.0. https://www.naturalearthdata.com/. Accessed December 2022.

Newman, P. W. and J. R. Kenworthy (1989). Gasoline consumption and cities: A comparison of US cities with a global survey. *Journal of the American Planning Association 55*(1), 24–37.

Nivola, P. S. (1999, Fall). Are europe's cities better? *Public Interest* (137), 73–84.

Nunn, N. and D. Puga (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics 94*(1), 20–36.

OECD (2016). The economic consequences of outdoor air pollution. Organisation for Economic Co-operation and Development. https://doi.org/https://doi.org/10.1787/9789264257474-en. Accessed December 2022.

Oliveira, E. A., J. S. Andrade Jr, and H. A. Makse (2014). Large cities are less green. *Scientific Reports 4*, Article number: 4235.

Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. Powell, E. C. Underwood, J. A. D'amico, I. Itoua, H. E. Strand, J. C. Morrison, et al. (2001). Terrestrial ecoregions of the world: A new map of life on earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience 51*(11), 933–938.

Panagos, P., M. Van Liedekerke, A. Jones, and L. Montanarella (2012). European soil data centre: Response to european policy support and public data requirements. *Land Use Policy 29*(2), 329–338.

Parry, I. W. and K. A. Small (2009). Should urban transit subsidies be reduced? *American Economic Review 99*(3), 700–724.

Pope III, C. A. and D. W. Dockery (2006). Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association 56*(6), 709–742.

Ramankutty, N., J. A. Foley, J. Norman, and K. McSweeney (2002). The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Global Ecology and biogeography 11*(5), 377–392.

Reba, M., F. Reitsma, and K. C. Seto (2016). Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000. *Scientific data 3*, 160034.

Redding, S. J. and D. M. Sturm (2008). The costs of remoteness: Evidence from german division and reunification. *American Economic Review 98*(5), 1766–1797.

Santos, G., W. W. Li, W. T. Koh, et al. (2004). Transport policies in Singapore. *Research in Transportation Economics 9*(1), 209–235.

Sarzynski, A. (2012). Bigger is not always better: A comparative analysis of cities and their air pollution impact. *Urban Studies 49*(14), 3121–3138.

Schindler, M., G. Caruso, and P. Picard (2017). Equilibrium and first-best city with endogenous exposure to local air pollution from traffic. *Regional Science and Urban Economics 62*, 12–23.

Schrank, D., B. Eisele, T. Lomax, J. Bak, et al. (2015). 2015 urban mobility scorecard. *Texas A&M Transportation Institute 9*(2015), 10.

Schwela, D. and O. Zali (1998). *Urban traffic pollution.* CRC Press.

Süddeutsche Zeitung (2022). Weniger Stau durch Neun-Euro-Ticket. Online article: https://www.sueddeutsche.de/wirtschaft/9-euro-ticket-stau-tomtom-1.5612144. Accessed December 2022.

Shefer, D. and P. Rietveld (1997). Congestion and safety on highways: towards an analytical model. *Urban Studies 34*(4), 679–692.

Simeonova, E., J. Currie, P. Nilsson, and R. Walker (2021). Congestion pricing, air pollution, and children's health. *Journal of Human Resources 56*(4), 971–996.

Sommer, F., V. Dietze, A. Baum, J. Sauer, S. Gilge, C. Maschowski, R. Gieré, et al. (2018). Tire abrasion as a major source of microplastics in the environment. *Aerosol and air quality research 18*(8), 2014–2028.

Spiegel Online (2022). Grünenchefin Lang will Folgeregelung für 9-Euro-Ticket ausloten. Online article: https://www.spiegel.de/politik/deutschland/gruenen-chefin-ricarda-lang-will-folgeregelung-fuer-neun-euro-ticket-ausloten-a-bd1c5467-92a4-4108-beb6-5ccc2074e917. Accessed December 2022.

Summala, H., E. Pasanen, M. Räsänen, and J. Sievänen (1996). Bicycle accidents and drivers' visual search at left and right turns. *Accident Analysis & Prevention 28*(2), 147–153.

Tirachini, A. and O. Cats (2020). Covid-19 and public transportation: Current assessment, prospects, and research needs. *Journal of Public Transportation 22*(1), 1.

Treiber, M., A. Kesting, and C. Thiemann (2008). How much does traffic congestion increase fuel consumption and emissions? Applying a fuel consumption model to the NGSIM trajectory data. In *87th Annual Meeting of the Transportation Research Board, Washington, DC*, Volume 71, pp. 1–18.

Tu, R., J. Xu, A. Wang, M. Zhang, Z. Zhai, and M. Hatzopoulou (2022). Real-world emissions and fuel consumption of gasoline and hybrid light duty vehicles under local and regulatory drive cycles. *Science of The Total Environment 805*, 150407.

Umweltbundesamt (2022). Air quality data. Online resource: https://www.umweltbundesamt.de/en/data/air/air-data. Accessed December 2022.

United Nations (2019). World urbanization prospects: the 2018 revision.

Van den Elshout, S., H. Bartelds, H. Heich, and K. Léger (2012). CAQI air quality index. comparing urban air quality across borders.

Van den Elshout, S., K. Léger, and H. Heich (2014). CAQI common air quality index — update with $PM_{2.5}$ and sensitivity analysis. *Science of the total environment 488*, 461–468.

Van Donkelaar, A., R. V. Martin, M. Brauer, N. C. Hsu, R. A. Kahn, R. C. Levy, A. Lyapustin, A. M. Sayer, and D. M. Winker (2016). Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental Science & Technology 50*(7), 3762–3772.

Van Donkelaar, A., R. V. Martin, R. J. Spurr, and R. T. Burnett (2015). High-resolution satellite-derived $PM_{2.5}$ from optimal estimation and geographically weighted regression over north america. *Environmental science & technology 49*(17), 10482–10491.

VDV (2022). Bilanz zum 9-Euro-Ticket . Online resource: https://www.vdv.de/bilanz-9-euro-ticket.aspx. Accessed December 2022.

Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review 59*(2), 251–260.

Vitrano, C. (2021). Covid-19 and public transport - a review of the international academic literature. The Swedish Knowledge Centre for Public Transport. K2 Working Paper 2021:1.

Wolff, H. (2014). Keep your clunker in the suburb: Low-emission zones and adoption of green vehicles. *Economic Journal 124*(578), F481–F512.

World Bank (2019). World development indicators. The World Bank. https://datacatalog.worldbank.org/dataset/world-development-indicators. Accessed December 2022.

World Health Organization (2003). Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide: report on a who working group, bonn, germany 13-15 january 2003. Technical report, Copenhagen: WHO Regional Office for Europe.

World Health Organization (2006). Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide. World Health Organization. Regional Office for Europe. https://apps.who.int/iris/handle/10665/107823. Accessed December 2022.

World Health Organization (2015). Economic cost of the health impact of air pollution in europe: Clean air, health and wealth. Copenhagen: WHO Regional Office for Europe.

World Health Organization (2021). Who global air quality guidelines: particulate matter, ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization. https://apps.who.int/iris/handle/10665/345329. Accessed December 2022.

Yang, Y., X. Wu, P. Zhou, Z. Gou, and Y. Lu (2019). Towards a cycling-friendly city: An updated review of the associations between built environment and cycling behaviors (2007–2017). *Journal of Transport & Health 14*, 100613.

Yang, Z. and M. Tang (2018). Does the increase of public transit fares deteriorate air quality in Beijing? *Transportation Research Part D: Transport and Environment 63*, 49–57.

Zeit Online (2022). Bahn-Betriebsrat meldet 400 überfüllte Züge am Pfingstwochenende. Online Article: https://www.zeit.de/mobilitaet/2022-06/9-euro-ticket-deutsche-bahn-ueberlastung-pfingsten?. Accessed December 2022.

Zhou, Y., J. I. Levy, J. S. Evans, and J. K. Hammitt (2006). The influence of geographic location on population exposure to emissions from power plants throughout China. *Environment International 32*(3), 365–373.