ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Master Thesis [programme Policy Economics]

Brain Fog or Smog? Estimating the Effect of Air Pollution on Test Scores

Name student: Daniel Pritsch
Student ID number: 506483

Supervisor: Dr. K.F.J Spiritus
Second assessor: Prof.dr. H.D. Webbink

Date final version: 22-07-2022

Contents

1 Introduction

Since the popularization of human capital theory by Becker (1962), a wide range of research has centered around the subject. Such research is often interested in the returns to human capital investments in the form of education. For an overview of early quasi-experimental studies of this nature, see Card (1999). More recently, there has also been an interest in determining the effects of other factors that either positively or negatively influence human capital, particularly in early childhood. An overview of such studies is given by Currie and Almond (2011). Some of these studies focus on the effects of environmental factors, especially on children in utero. Almond et al. (2009), for example, study the effect of prenatal exposure to radioactive fallout in Sweden and find large negative effects on school outcomes, but not on health outcomes. Comparable results for scholastic performance have been found for early childhood lead exposure (see, e.g., Nilsson, 2009; Reyes, 2011).

A more recent strand of literature has studied how air pollution can influence cognitive performance. Such research contributes to our scientific understanding of the determinants of human capital, as decreased cognitive performance can harm human capital accumulation. Besides that, it also has strong societal relevance. Governments are increasingly becoming aware of the negative effects of air pollution, mainly with regards to its health consequences (UNEP, 2021). Possible negative effects on cognitive ability can provide an additional rationale for (stronger) policies to fight air pollution. Furthermore, it provides insights into the externalities associated with the emission of air pollutants. Such information is useful when considering what the optimal taxation of these pollutants would be. My thesis contributes to this literature by using a quasi-experimental method, the difference-in-differences methodology, to study the effect of air pollution on primary school test scores.

Previous studies in the field of economics have used student-fixed effects or instrumental variable models to investigate the effect of air pollution on test scores (see, e.g., Carneiro et al., 2021; Lavy et al., 2014; Zweig et al., 2009). Fixed effects models rely on the assumption that changes in air pollution levels are uncorrelated with other time-variant factors that affect test scores. This can be hard to justify when it is unclear what is causing the changes in air pollution levels. Carneiro et al. (2021) use wind as an instrumental variable for air pollution levels, with a necessary assumption being that wind direction only affects test scores through its effect on air pollution. This assumption is questionable, among others because wind direction can also affect the level of pollen in an area, which in turn can affect test scores (Bensnes, 2016). I take a novel approach in this area of research, by relying on a policy change in the form of environmental zones as an exogenous shift in air pollution levels. I hypothesize that such a policy lowers air pollution, and thereby improves test scores.

Environmental zones are areas in which certain (usually diesel) vehicles are prohibited from entering, with the aim of reducing the emission of harmful pollutants by these vehicles. I use a difference-in-differences model to assess the effects that environmental zones have on air pollution, and in turn what their effect is on test scores. For the effects on air pollution, I collect data from pollution measuring stations in the 40 largest cities in The Netherlands on monthly average values of PM_{10}, $PM_{2.5}$, NO_2, NO, O_3, and soot. I investigate the introduction of an environmental zone for cars in Rotterdam, Utrecht, and Arnhem, and use stations in cities with an environmental zone for trucks as control units. I argue that these serve as the best counterfactual for these cities, as Rotterdam, Utrecht, and Arnhem also had environmental zones for trucks in place before they introduced an environmental zone for cars. I consider all measuring stations in these three cities "treated" to capture potential spillover effects. I separately estimate the effects of each environmental zone on each pollutant using a "classical" difference-in-differences model, and informally test for parallel trends with event studies. I find evidence for a decrease of approximately 4% on PM_{10} concentrations in Rotterdam, and a decrease of approximately 4% on soot concentrations in Arnhem. I find no convincing evidence for the environmental zone in Utrecht.

Next, I examine whether the environmental zones have any effects on test scores. I use scores on the tests that all primary school students in The Netherlands take at the end of their final year. There are five tests to choose from, with the CITO test being by far the most popular. I have data on the specific test taken by a school and the average test score for each school year from 2010/2011 to 2018/2019. For these regressions, I employ the novel synthetic difference-in-differences methodology developed by Arkhangelsky et al. (2021). This method assigns weights to the control schools to make their average trend of test scores before the environmental zone parallel to the trend for the treated schools before the environmental zone. I again estimate the effects of the environmental zones for cars in Rotterdam, Utrecht, and Arnhem. Additionally, I can test the effect of an environmental zone for trucks in Arnhem. This environmental zone was introduced in 2014, which means that I can assess what effect this zone had on test scores. This was not possible for air pollution, as that dataset starts in 2014 and thus does not have a pre-intervention period. I first use standardized test scores for all tests as my dependent variable, and next only use schools that always take the most popular CITO test. I find non-significant results for all environmental zones, with the exception of the environmental zone for trucks in Arnhem. Here, I find a sizable and negative effect on standardized test scores of nearly 0.2 standard deviations. When using CITO test scores, this effect is no longer statistically significant. However, I now find that the environmental zone for cars in Rotterdam had a significant and positive effect on test scores, also of nearly 0.2 standard deviations. Test scores for schools in Rotterdam seem to be increasing more rapidly than test scores for the control schools in the years before the environmental zone was introduced. This might indicate that, if Rotterdam had not introduced an environmental zone, test scores in Rotterdam would have kept developing at a faster pace than the control schools did. This would lead to overestimation of the effect of the environmental zone on test scores. Additionally, in my robustness checks I find that there are composition effects present, where Rotterdam saw a decrease in the share of students from lower socioeconomic backgrounds after the introduction of the environmental zone. These results indicate that the increase in test scores for Rotterdam was partly caused by composition effects. All in all, although I find some evidence that environmental zones lower local air pollution, I do not find convincing proof that this leads to increases in test scores.

This thesis proceeds as follows. Section 2 contains the theoretical framework, which gives more background information on air pollution and past research on this topic. Section 3 describes the data that I have collected, and section 4 explains my methodology for both air pollution and test scores. Section 5 presents my results for the main regressions and multiple robustness checks. Finally, section 6 discusses my findings, and section 7 concludes.

2 Theoretical framework

This chapter serves as the theoretical background for my hypotheses. Section 2.1 explains what air pollution is, and section 2.2 gives an overview of the sources of different air pollutants. Section 2.3 then looks at how pollution can affect cognitive ability. Section 2.4 provides an overview of the different types of environmental zones and discusses the effects that environmental zones have on local air pollution. Finally, I formulate my hypotheses in section 2.5.

2.1 What is air pollution?

Before I start discussing the effects of air pollution, it is useful to explain what is meant by air pollution. The OECD defines air pollution as follows: “Air pollution is the presence of contaminant or pollutant substances in the air that do not disperse properly and that interfere with human health or welfare, or produce other harmful environmental effects” (OECD, 2001). Air pollution thus occurs when substances, or air pollutants, prevent some desired air quality from being reached. Five common air pollutants have standards set by the World Health Organization (WHO), with the goal of informing legislation and policy to reduce the health impacts of air pollution. These include particulate matter (PM), ozone (O_3), carbon monoxide (CO), sulfur dioxide (SO_2), and nitrogen dioxide (NO_2) (World Health Organization, 2021). Of these, PM can be classified into the size of the particles. Here, PM_{10} (particulate matter with a diameter smaller than 10 µm) and $PM_{2.5}$ (particulate matter with a diameter smaller than 2.5 µm) are the most common categories. PM differs from the other criteria air pollutants because it is a composite measure rather than a single chemical entity. It can include particles such as dust, soot, and smoke, but also particles formed by a reaction between other pollutants such as SO_2 and NO_2 (EPA, 2019; Vallero, 2014). Taken together, nitrogen dioxide and nitrogen monoxide (NO) are often referred to as nitrogen oxides (NO_x). Nitrogen dioxide is formed by a reaction of nitrogen monoxide with oxygen or ozone in the air (EEA, n.d.).

2.2 Sources of air pollution

The sources of pollution are numerous, and encompass both natural sources and man-made sources (Vallero, 2014). In addition to the air pollutants with WHO standards, I discuss soot separately here. For the composition of sources of each pollutant, I focus on The Netherlands, since this is the country of interest for my thesis.

Particulate matter (PM)

As mentioned, PM comprises many different substances. In 2016, the average concentrations of PM_{10} and $PM_{2.5}$ were 17.5 µg/m^3 and 10.2 µg/m^3, respectively. On average, more than three quarters of PM_{10} and nearly 90% of $PM_{2.5}$ in outdoor air came from human sources. This can be even higher near busy roads and industrial sites (RIVM, n.d.-a). As a comparison, the WHO guidelines on air pollution recommend an annual concentration of 15 µg/m^3 and 5 µg/m^3, for PM_{10} and $PM_{2.5}$ respectively (World Health Organization, 2021).

Of all human emission sources of PM_{10} in 2015, the highest share came from traffic and transport (36%), followed by agriculture (22%) and non-specified industry (18%). Within the traffic source, a majority came from buses and touring cars (37%) and shipping (23%). Trucks also made up approximately one fifth of the total, divided into light trucks (15%) and heavy trucks (4%). Cars emitted 12% of all traffic emissions. Compared to PM_{10}, a larger share of $PM_{2.5}$ comes from abroad and from road traffic (RIVM, n.d.-a). This might be caused by the fact that $PM_{2.5}$ is more likely to travel long distances than PM_{10} due to its size and weight (Vallero, 2014).

Nitrogen oxides (NO_x)

Of total NO_x emissions in 2014, traffic and transport were the main source by a substantial margin (64%). In 2015, the average concentration of NO_2 in outdoor air was 14.7 µg/m^3. Of this concentration, the largest part came from other countries (37%), with road traffic (27%), other traffic (14%) and shipping (9%) being the main national sources. Around large cities, the share coming from foreign sources decreases. In contrast, the share coming from road traffic is much higher in Utrecht (45%), Amsterdam (36%) and Rotterdam (30%). Near busy roads, average annual concentrations can be higher than 40 µg/m^3, with more than half of the total concentration coming from traffic (RIVM, n.d.-a). The WHO guidelines recommend an annual concentration of 10 µg/m^3 for NO_2 (World Health Organization, 2021).

Ozone (O_3)

O_3 is not directly emitted but is formed by a reaction between NO_x and volatile organic compounds. Sources of NO_x thus indirectly contribute to O_3 concentrations (RIVM, n.d.-a). O_3 concentrations are in the range of 40-55 μg/m^3, compared to the WHO guideline of a daily 8-hour mean of 100 μg/m^3 during peak season (RIVM, n.d.-b; World Health Organization, 2021). Although it seems like decreases in the concentration of NO_x would lead to decreases in the concentration of O_3, the relationship does not appear to be as clear. For example, Jhun et al. (2015) find evidence that decreases in NO_x led to decreases in peak O_3, but that these NO_x decreases were also associated with higher non-peak concentrations in O_3. Similarly, Shi and Brasseur (2020) find that although $PM_{2.5}$ and NO_2 decreased during the COVID-19 lockdown in China, O_3 concentrations increased.

Carbon monoxide (CO)

CO is mainly emitted through incomplete combustion, with traffic being the main source in 2012 (49%). Traffic emissions have been cut in halve between 1990 and 2013, and the 8-hour limit of 10,000 μg/m^3 has not been exceeded since 1994 (Compendium voor de Leefomgeving, 2014).

Sulfur dioxide (SO_2)

SO_2 is mainly emitted through the burning of coal and oil. Emissions from traffic have been reduced dramatically in recent decades due to decreases in the level of sulfur in fuels. Because of this, the share of SO_2 emissions originating from traffic is less than 2%. The main emission sources are shipping (30%), refineries (25%) and electricity generation (19%) (RIVM, n.d.-a).

Soot

Although soot is usually only considered as a part of PM, I discuss it separately here because the data for The Netherlands also report soot separately. Soot is mainly emitted through the burning of fossil fuels. Within transport, diesel vehicles are the main contributors to soot emissions. Average soot concentrations are in the range of 0.5-2.0 μg/m^3, and concentrations can be twice as high around busy streets compared to urban areas on average (RIVM, n.d.-c).

2.3 Air pollution and cognitive ability

When PM is inhaled, these particles can travel from the lungs to other organs, including the brain (Peters et al., 2006). Multiple studies have shown that there is an association between long-term air pollution exposure, in the form of $PM_{2.5}$, O_3 and NO_2, and risk of developing dementia and Alzheimer's disease (see, e.g., Carey et al., 2018; Jung et al., 2015; Peters et al., 2006; Wang et al., 2021; Younan et al., 2021). With regards to SO_2 and CO, research is somewhat limited, although epidemiological research does point to an increased risk of dementia associated with these pollutants (Fu and Yung, 2020). It should be noted that, although these studies try to control for many relevant covariates, they are ultimately measuring correlations. They thus cannot claim with certainty that air pollution is the cause of higher rates of dementia. However, these studies are still relevant because it is generally infeasible to study this relationship with a randomized controlled trial, at least for humans. Animals who were exposed to air pollution in experiments have shown developments in the brain that are similar to developments in the brains of humans which have been linked to Alzheimer's disease (Calderón-Garcidueñas et al., 2020). These findings indicate that air pollution can have lasting consequences on the brain, making it likely that it can affect cognitive performance more generally. Indeed, long-term exposure to PM, SO_2, CO, O_3 and NO_2 has been linked to a more general decline in cognitive performance (Chen and Schwartz, 2009; Park et al., 2022; Younan et al., 2020; Zhang et al., 2021).

For short-term exposure to air pollution, experiments are a more feasible research method. Regardless, experimental research on this topic is relatively scarce, with the best evidence on the effect on cognitive performance coming from Shehab and Pope (2019). These authors run two experiments where the subjects take three types of cognitive tests under low concentrations of PM, and the same tests under a higher level of PM. In the first experiment, subjects were exposed to candle burning for an hour, which is a source of different types of PM. In the second experiment, subjects commuted next to a busy road for thirty minutes. This meant that they were exposed to different pollutants, such as PM, NO_x and CO. The results indicate that both sources of air pollution led to a decline in cognitive performance on the Mini-Mental State Examination, which is a test to measure overall cognitive ability. Because both experiments find very similar effect sizes, the authors argue that PM is the cause of this decline. They also find that the outdoor commuting experiment led to declines in automatic detection speed on a different test. For controlled search speed and a final test on colors and words, no effect is found for either source of pollution.

There are also several papers from the economic literature that use observational data to link air pollution and cognitive performance. Lavy et al. (2014) have data on high school test scores and pollution in Israel from 2000-2002. Each student takes multiple tests at different locations at the end of each year, which means the authors can include city, school, and student fixed effects. They combine data on test scores with data on the average concentration of $PM_{2.5}$ and CO on the day of the test in the city that each school is located in. These pollution concentrations are measured on the Air Quality Index (AQI), which ranges from 0 to 500, with values above 150 being considered unhealthy. This makes their results harder to compare to other findings, but they find that a 10 unit increase in the AQI for $PM_{2.5}$ lowers test scores by 1.9% of a standard deviation. Similarly, they find that a 10 unit increase in the AQI for CO lowers test scores by 3.5% of a standard deviation. Carneiro et al. (2021) have data on test scores in Brazilian university entrance tests, which are spread out over two days. These tests are taken at exam venues in more than 250 municipalities, which allows the authors to combine test score data with data on the average concentration of PM_{10} and O_3 in a municipality on the day of the test. Because the tests are spread out over two days, they can include student and exam fixed effects. They also include weather controls, in the form of wind speed, humidity, and temperature, as these factors can be correlated with pollution levels and also affect test scores. Using this model, they find that a 10 μg/m^3 increase in PM_{10} on examination days lowers test scores by 8% of a standard deviation. Next, they also use an instrumental variables model with wind direction as an instrumental variable for air pollution. Wind direction can influence the concentration of air pollution in a city by carrying air from a highly polluting area. Here, the assumption is that wind direction in a municipality on the day of the test only affects test scores in that municipality through its effect on air pollution in that municipality. With this model, they again find that a 10 μg/m^3 increase in PM_{10} on examination days decreases test scores by 8% of a standard deviation.

Although these studies form an important step in assessing the effects of air pollution on cognitive performance, there are reasons to question the reliability of their results. Fixed effects models rely on the assumption that changes in air pollution are not correlated with changes in any unobserved factors that influence test scores, which can be questionable when it is unclear where the variation in air pollution comes from. Using wind direction as an instrumental variable relies, amongst others, on the assumption that wind direction only affects test scores through its effect on air pollution levels. This assumption is questionable, for instance because wind direction can also influence the amount of pollen in an area, which has been shown to negatively influence test scores (Bensnes, 2016). Depending on the relative presence of pollen-producing plants near sources of air pollution, this could lead to either overestimation or underestimation of the true effect.

2.4 Environmental zones and air pollution

Environmental zones in The Netherlands can be broadly divided into three categories of vehicles they target: diesel cars and delivery vans, diesel trucks, and diesel buses. As of May 2022, there are four cities with an environmental zone for diesel cars and delivery vans, 15 cities with an environmental zone for diesel trucks, and four cities with an environmental zone for diesel buses (Milieuzones in Nederland, n.d.-a, n.d.-b, n.d.-c). On January 1, 2020, new standards were put in place to make environmental zones more comparable across cities. Before then, cities were largely able to set their own standards, which resulted in a variety of different rules among environmental zones. Regardless, these zones had the same overarching principle: vehicles below a certain emission class were not allowed to enter. These emission classes usually use different pollutants, such as PM, CO and NO_x, to assign a value of how clean a vehicle is (ANWB, n.d.). Besides these main categories, there is also an environmental zone for mopeds in Amsterdam and The Hague (Gemeente Amsterdam, n.d.; Gemeente Den Haag, 2022). As I will explain in section 4, I use the introduction of environmental zones for *cars* in Utrecht, Arnhem and Rotterdam, and an environmental zone for *trucks* in Arnhem in my analysis. Table A1 provides a full overview of environmental zones for trucks and cars per city.

Multiple reports have focused on predicting the effects of environmental zones on air pollution, as well as on assessing the actual effects after an environmental zone has been introduced.

I discuss a few of these reports here, to get an idea of the general findings of these studies. Additional focus will be given to the environmental zones in Utrecht, Arnhem and Rotterdam. Nieman et al. (2010) study the introduction of environmental zones for trucks in 11 cities in July of 2007. Based on license plate scans and model calculations, they find that in 2010, PM_{10} concentrations decreased by 0.15 to 0.25 $\mu g/m^3$ near roads with concentrations above EU standards. They predict that this effect size will stay similar in 2013 and 2015. On average, they find that PM_{10} concentrations are 0.02 to 0.08 $\mu g/m^3$ lower for streets inside the environmental zones. For NO_2, they find no effect in 2010, but predict that there could be decreases up to 0.3 $\mu g/m^3$ near roads with concentrations above EU standards. On average, they expect NO_2 decreases of 0.02 to 0.09 $\mu g/m^3$ inside environmental zones in 2013 and 2015. The effects for PM_{10} and NO_2 are both smaller than the predicted effects in 2007. This was caused by three factors: new trucks being less clean than was initially expected, diesel particulate filters that are encouraged by environmental zones increasing NO_2 concentrations, and imperfect compliance.

Eijk and Voogt (2016) assess the effect that the introduction of an environmental zone for cars in Utrecht in 2015 had on soot concentrations. Their analysis is twofold: they perform a license plate scan to see how the composition of cars has changed, and also use actual measures of soot concentrations. They compare data in 2014 with data in 2015 to assess the impact of the environmental zone, controlling for a prognosis of what the composition of cars in 2015 would have been without the environmental zone. The license plate scan revealed large differences in the composition of cars in Utrecht, which they estimate to have led to a decrease in PM (including soot) but not NO_x. The results based on measured soot concentrations differ, however. Although they find that the contribution of traffic to soot concentrations decreased by 16%, they cannot conclude that this was caused by the environmental zone rather than by other factors. It is not clear why they do not find an effect of the environmental zone on air pollution. One organization appealed the decision to introduce the environmental zone in Utrecht by arguing that it led to traffic bottlenecks because of vehicles that had to take a detour (Trouw, 2017). Although this appeal was unsuccessful at preventing the environmental zone from being introduced, it might give a possible explanation for the limited effect on air pollution that is found.

With regards to the environmental zone for cars that was introduced in Arnhem in 2019, no extensive analysis has been performed so far. However, there was a sharp decrease in the number of old diesels driving in the city center, from 1,800 in 2017 to 150 in 2019 (Van der Vegt, 2019). Van de Poll et al. (2017) calculated that an environmental zone for cars in Arnhem could result in decreases of 17% for soot, nearly 10% for $PM_{2.5}$, 3% for PM_{10}, and 4% for NO_x in 2020.

In January 2016, Rotterdam extended the area of its environmental zone for trucks and simultaneously introduced an environmental zone for cars in this same area (Rubio, 2015). Based on scans of the composition of cars in Rotterdam, the expected emissions were calculated in 2017 and compared to 2015. The policies were estimated to have led to a 13% decrease in soot emissions by traffic and a 4% decrease in NO_x traffic emissions (Gemeente Rotterdam, 2018). In the year following the introduction of this environmental zone, the amount of old diesel cars in Rotterdam nearly decreased by half (CBS, 2017). In contrast to other cities, Rotterdam abolished its environmental zone for cars after some time. This was done in two stages, allowing old gasoline cars back from July 2018 onwards and allowing old diesel cars back from January 2020 onwards (ANWB, 2019).

2.5 Hypotheses

Based on my theoretical framework, I formulate the following hypotheses:

Hypothesis I: Environmental zones lower concentrations of air pollution.

Hypothesis II: Environmental zones improve test scores.

The second hypothesis follows naturally from the first hypothesis. Many of the studies that I have discussed, suggest that air pollution lowers cognitive performance. If environmental zones succeed in lowering air pollution, then they should also cause an increase in cognitive performance. I hypothesize that such an increase will be visible in improved test scores. The next section discusses the data that I have available to test these hypotheses.

3 Data

3.1 Air pollution data

For air pollution, I use data from Luchtmeetnet. This website reports monthly values for a wide range of air pollutants, such as PM, O_3, NO_x, SO_2 and CO, but only has data for measuring stations rather than averages for all municipalities. I collect data from all cities with an environmental zone, as well as all cities from the G40-network that did not implement an environmental zone (“40 steden”, n.d.). I do not collect data for air pollutants with very few measuring stations. This means I only have data on the monthly concentration of PM_{10}, $PM_{2.5}$, NO_2, NO, O_3, and soot, given in µg/m^3. These data go back to 2014 at the earliest, depending on the measuring station. However, most measuring stations have some missing data, either because they started measuring later than 2014 or because they do not measure a certain pollutant. As with test scores, I drop observations for 2020 and 2021 due to the COVID-19 crisis. I drop data from Amsterdam, due to multiple types of environmental zones being introduced in Amsterdam. In total, this leaves me with data for 37 measuring stations from 16 municipalities. Of these stations, 11 are located in cities that introduce an environmental zone for cars, and 22 are located in cities that have an environmental zone for trucks at some point between 2014 and 2019. Table A2 presents summary statistics for different groups of measuring stations.

3.2 Test score data

My main dependent variable of interest is the result on the final exams for primary school students in The Netherlands. At the end of primary school, all students in The Netherlands (who are 11-12 years old at the time) are required to participate in a centralized test. Each school can choose which test to take each year, and they currently have five options to choose from (Ministerie van Algemene Zaken, n.d.). The test scores lead to a recommendation for the high school level the student should go to. If this recommendation is higher than the recommendation that was given by their teacher beforehand, the definitive recommendation can be adjusted upwards (Inspectie van het Onderwijs, n.d.). Annual data on which test a school chose and the average test score per school are publicized by Dienst Uitvoering Onderwijs (DUO). These data are available for the school years 2010/2011 to 2020/2021, with the exception of 2019/2020 when all tests were cancelled due to COVID-19. This means that I have access to data on average test scores at the school level taken in 10 years. This dataset also contains information on the number of students who took the test, for example. A more extensive dataset is available for the school years 2016/2017 to 2020/2021, which has additional information on the postal code area and municipality of a school, among others. Combining these two datasets allows me to use the more extensive data period of the former, but to also add information on the location of the school based on the latter. Here, I assume that in school years 2010/2011 to 2015/2016, the location of each school was the same as the location of the school in the school year 2016/2017. Although this will likely not be factually correct for all schools, I believe errors will be small in size and essentially random, not biasing my estimates. Additional data on school characteristics are also available annually for the entire data period. These include the number of students by gender and non-Dutch background, and a school weight that serves as an indication for the socioeconomic status of the student population for each school. Finally, I drop observations for the final school year 2020/2021, to prevent my findings from being driven by outliers caused by COVID-19. This means that the final school year I take into consideration is 2018/2019.

Following Carneiro et al. (2021), standardized test scores will be calculated and used as the dependent variable. The approach taken here will differ because there are different tests available. For each school, the standardized test score will be calculated by taking the difference between the school's average test score and the average test score across all schools for that test in that year, and dividing it by the standard deviation. Table A3 provides summary statistics on the data for test scores.

4 Methodology

This chapter is structured as follows. I first explain why using a standard OLS-regression is unlikely to retrieve the true causal effect of air pollution on test scores. I then introduce the basic idea of the strategy I use instead, namely the difference-in-differences methodology. Section 4.1 describes how I use the difference-in-differences methodology to test my first hypothesis, that environmental zones lower air pollution, in more detail. Section 4.2 then discusses the synthetic difference-in-differences methodology I use to investigate my second hypothesis, that environmental zones improve test scores. This section starts with an explanation of the synthetic control method, as it is closely related to the synthetic difference-in-differences and can help to understand the method better. Furthermore, it is useful to discuss what the advantages are of the synthetic difference-in-differences over the synthetic control method that led me to use the former instead of the latter. I then proceed with a more in-depth explanation of the synthetic difference-in-differences method as I use it.

First, it is worthwhile to consider why I do not simply run an OLS-regression of the average test score on the average pollution level near a school. The main issue is that local pollution levels are correlated with other factors that affect test scores, such as socioeconomic status (Savelkoul et al., 2010). Since socioeconomic status of parents affects school performance of children, this creates geographical variation in test scores that is *correlated* with air pollution but not *caused* by it.[1] A simple OLS-regression would thus result in a biased estimate of the effect of air pollution on test scores, as this estimate would also capture these other factors that are correlated with air pollution and affect test scores.

In order to get a more reliable causal estimate of the relationship, I use a difference-in-differences approach to investigate the effect of environmental zones in cities in The Netherlands. To test my first hypothesis, that environmental zones lower air pollution, I use a standard difference-in-differences approach. To test my second hypothesis, that environmental zones improve test scores, I use the newer synthetic difference-in-differences method developed by Arkhangelsky et al. (2021). An explanation of the synthetic difference-in-differences method follows in section 4.2.

The difference-in-differences methodology is used to estimate the effect of a certain policy, also referred to as intervention or treatment, on a certain outcome. In its most basic form, it is used when there are data on two groups and two time periods. The two groups can be single aggregate units such as states or countries, but also consist of multiple units such as firms or schools. One of the groups introduces the policy of interest, and is thus named the treatment group. The other group does not introduce the policy of interest, and is referred to as the control group. The two time periods will generally be two consecutive years or months, but could also be further apart. The important thing is that the policy of interest is introduced between the two time periods. This ensures that there is one time period when neither group is treated, and one time period after the treatment group has introduced the policy and has thus become treated. In order to retrieve the effect of the policy, or the

[1] See Zumbuehl & Dillingh (2020) for evidence on educational disparities between different socioeconomic groups.

treatment effect, you need to know what the counterfactual scenario is for the treatment group. In other words, what would the value of the outcome variable have been for the treatment group if it had not been treated? The difference-in-differences methodology estimates this counterfactual by assuming that, in the absence of treatment, the outcome variable would have evolved in a similar fashion in both groups. This is also known as the parallel trends assumption. This is a fundamentally untestable assumption, as the counterfactual scenario is never observed. After all, the treatment group cannot simultaneously be treated and untreated in the last time period (Cunningham, 2021).

The treatment that I am interested in is the introduction of environmental zones. My analysis essentially consists of two parts. I first estimate the effect of environmental zones on air pollution to test my first hypothesis, and the second part investigates the effect of environmental zones on test scores to test my second hypothesis. The units of analysis are measuring stations for the first hypothesis, and schools for the second hypothesis. I consider a school or measuring station treated if it is located in a city with an active environmental zone. My reasoning to consider all schools and stations in one of these cities treated, as well as potential issues with this definition, are discussed further in section 4.1. Because the difference-in-differences methodology uses changes in treatment, there are four environmental zones of interest in my dataset for test scores, which covers 2011 until 2019: an environmental zone for cars in Utrecht in 2015, an environmental zone for trucks in Arnhem in July 2014, an environmental zone for cars in Arnhem in 2019, and an environmental zone for cars in Rotterdam in 2016. Although there are also changes in the environmental zone in Amsterdam during this period, I would be unable to isolate the effect of one particular environmental zone in this city. I thus exclude Amsterdam from my sample.

4.1 Air pollution methodology

This section precedes as follows. I first provide the technical details of my difference-in-differences analysis for my first hypothesis. I then discuss the identifying assumption, the parallel trends assumption, more closely. Finally, I discuss potential issues with this analysis.

The first regression I run investigates my first hypothesis, that environmental zones lower air pollution. The regression equation for this is as follows:

$$Pollution_{mt} = \alpha_0 + \sum_{t=1}^{T} \alpha_t \, monthofyear_t + \gamma TreatedEnvironmentalZone_{mt} + \mu_m + \varepsilon_{mt}, \quad (1)$$

where $Pollution_{mt}$ is the concentration of a certain type of air pollutant at monitoring station m in month t. These pollutants are PM_{10}, $PM_{2.5}$, NO_2, NO, O_3, and soot, in $\mu g/m^3$, and there are T months in total in my dataset. $TreatedEnvironmentalZone_{mt}$ is equal to one when a station is both in the city with the environmental zone of interest and that environmental zone is active in month t. $monthofyear_t$ is a separate dummy variable for each month in my data period, since the air pollutant concentrations are monthly averages. These are the time fixed effects. Finally, μ_m are monitoring station fixed effects, and the error terms ε_{mt} are clustered at the municipality level. The parameter of interest is γ, which gives the effect that the environmental zone of interest has on pollution levels. Based on my first hypothesis, I expect γ to be negative, as I hypothesized that the introduction of an environmental zone lowers air pollution concentrations.

Due to missing data in 2014 for the monitoring stations in Arnhem, I cannot estimate the effect of the truck zone in Arnhem on air pollution. This means I run this regression three times in total, to get a separate effect for each of the car zones in Utrecht, Arnhem and Rotterdam. When estimating the effect of the car zone in Utrecht, I drop observations from Arnhem and Rotterdam, and vice versa. Before these cities implemented an environmental zone for cars, they all had an environmental zone for trucks in place. I therefore use monitoring stations in cities with an environmental zone for trucks as control units in all three regressions. I deem this to be the most appropriate counterfactual for each of the cities, as this corresponds to the situation in these cities before they introduced a car zone. The goal of the control stations is to serve as the counterfactual trend for how air pollution would have developed in the absence of the car zone. Cities that keep their truck zone when one of these cities adds an additional car zone seem more appropriate for this than cities that never had an environmental zone in the first place. It is important to note that all these control units have a truck zone in place in the entire data period, meaning that there are no changes in treatment status for them. I also drop data before July 2014 for the regression of Arnhem for this reason, as that is when Arnhem implemented its truck zone.

There are multiple reasons why I choose to consider every station in a city with an active environmental zone, rather than only the stations actually within the zone, as treated. It is partly due to data limitations, as none of the measuring stations in Utrecht and Arnhem are located inside the environmental zone and only one school is in Arnhem, for example. Using treatment at the city-level thus allows for more data on treated units to estimate the treatment effects. It is also motivated by the fact that spillover effects are highly likely within cities, due to pollution being carried across the environmental zone borders and changes in vehicle composition also affecting other parts of these cities. In addition, this allows me to capture potential increases in air pollution in areas outside the environmental zones caused by vehicles having to take detours. Some opponents of environmental zones have claimed that environmental zones are ineffective for this reason, but based on previous studies I expect such effects to be small. Similarly, it limits potential composition changes, which could arise if wealthier families move to areas within a new environmental zone. I assume that these spillover and composition effects are only present within cities, and not across cities. Although this likely "waters down" potential effects of environmental zones, I consider this the most valid approach to use.

I choose to run a separate regression for each environmental zone instead of running the regression once on my full sample, where I would have variation in treatment timing due to Utrecht, Arnhem and Rotterdam implementing environmental zones in different years. Such a model would produce a single estimate of the variance-weighted average effect of environmental zones on the treated (Goodman-Bacon, 2021). My reasoning to instead run separate regressions is both practical and theoretical. It allows me to obtain distinct effects per city, which is particularly useful if effects are heterogenous across cities, as section 2.4 suggested they might be. Furthermore, Goodman-Bacon (2021) showed that variation in treatment timing can result in biased estimates when treatment effects are not constant over time, which is plausible in my set-up. By running separate regressions, treatment only occurs at one point in time, meaning that this bias cannot arise.

In my context, the parallel trends assumption implies that the development of air pollution concentrations in Rotterdam, Utrecht, and Arnhem if they had not introduced an environmental zone for cars would have been the same as the actual development of air pollution in cities that only have an environmental zone for trucks. I perform two standard informal tests to check the validity of the assumption. These assumptions can give some degree of evidence for the parallel trends assumption, but are never able to prove it. After all, the counterfactual scenario of air pollution in Rotterdam, Utrecht, and Arnhem without the car zone is not observed in the periods when they have a car zone in place. The first test is that I run each regression again with a municipality-specific linear time trend added to the regression. This essentially relaxes the parallel trends assumption, as it allows treated and control municipalities to follow different linear air pollution trends (Angrist & Pischke, 2009). More specifically, this extrapolates any linear pre-treatment difference in trends between the treatment and control stations to the post-treatment periods (Rambachan and Roth, 2022). If the coefficients do not change much compared to the base model, this is generally considered to make the results more credible (Wing et al., 2018). However, this is still not unequivocal proof of the parallel trends assumption holding. It simply relaxes the assumption somewhat, to the assumption that any linear differences before an environmental zone is introduced would have remained in the absence of the environmental zone. Moreover, when the difference is not linear, this solution will not be valid (Rambachan and Roth, 2022). Secondly, I run an event study to examine the extent to which stations in Rotterdam, Utrecht, and Arnhem were already on a different air pollution trend compared to the control stations. This event study is essentially the same as the base difference-in-differences model, but it includes pre-treatment leads and post-treatment lags of the $TreatedEnvironmentalZone_{mt}$ coefficient. The regression, based on Cunningham (2021), looks as follows:

$$Pollution_{mz} = \eta_m + \lambda_z + \sum_{z=-q}^{-2} \theta_z \, TreatedEnvironmentalZone_{mz} + \sum_{z=0}^{n} \delta_z \, TreatedEnvironmentalZone_{mz} + \epsilon_{mz}, \quad (2)$$

where treatment occurs at month z=0. $TreatedEnvironmentalZone_{mz}$ is equal to one when a monitoring station is in a city with an active environmental zone of interest. η_m are monitoring station fixed effects, and λ_z are month fixed effects. q is the number of treatment leads, and n is the number of treatment lags. These are simply the number of months before and after the environmental zone is introduced, respectively. The error terms ϵ_{mz} are clustered at the municipality level. If monitoring stations in Utrecht, Arnhem, and Rotterdam are on parallel trends to the control stations, θ_z would be expected to be equal to zero. δ_z give a treatment coefficient per month, and can thus show whether the effect of environmental zones on air pollution is dynamic and changes over time. Note that z=-1 is the reference month. The θ_z and δ_z coefficients thus show how much bigger or smaller the difference in air pollution is between the treatment and control stations for each month compared to this reference month, after accounting for month and station fixed effects. Standard errors are clustered at the municipality level.

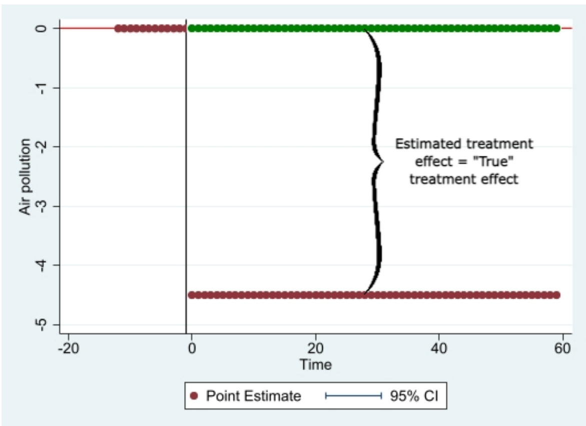
Event studies produce intuitive graphs that plot all of the θ_z and δ_z coefficients, which allows me to visually assess the parallel trends assumption. If the θ_z coefficients are equal to zero, this indicates that the treated monitoring stations were on a parallel trend for an air pollutant in the pre-treatment period. This does not mean that all of these monitoring stations did not experience an increase or decrease in air pollution concentrations over time. It merely means that the difference between these groups stayed constant over time.

As an example of what non-parallel pre-treatment trends mean, consider the example in Figure 1. In Panel A of Figure 1, the pre-treatment lag coefficients θ_z are equal to zero. This means that the concentration of the air pollutant was changing by the same amounts in the treatment stations and control stations each period. As soon as the environmental zone is introduced in period 0, the treatment stations see a drop in the air pollutant concentration. Because there was no difference in trends between treatment and control stations in the pre-treatment period, it seems that the control stations provide a credible counterfactual for the treatment stations. The treatment effect that a difference-in-differences regression estimates then seems likely to be equal to the ”true” treatment effect. Note, again, that this is not definitive proof for the parallel trends assumption. Just because the two trends moved parallel before the environmental zone, does not necessarily mean that they would have continued to do so in the absence of the environmental zone. However, I can be more confident in this assumption than if the two trends already did not move parallel before the environmental zone. For an example of this, consider Panel B of Figure 1. Here, the θ_z pre-treatment lag coefficients are positive in the first time period and then decrease over time to the period when the environmental zone is introduced. I can then say that the treated stations are on a “downward trend” compared to the control stations. This means that pollution concentrations seemed to be decreasing more rapidly (increasing less rapidly) for the treated stations compared to the control stations. It then seems likely that these treated stations would have continued to have more rapidly decreasing (less rapidly increasing) pollution concentrations after the treatment period, even if the environmental zone had not been introduced. This is indicated by the green dots, which show the actual counterfactual trend of the treated stations. The fact that the δ_z post-treatment lead coefficients are negative, is then partly due to stations in Utrecht being on a downward trend compared to the control stations rather than representing the actual effect of the environmental zone. This results in the difference-in-differences regression overestimating the treatment effect, as can be seen in Panel B. Consider, also, what the effect would be of including a municipality-specific linear time trend in the difference-in-differences regression, which is the first informal test I discussed. This would be successful in retrieving the ”true” treatment effect, since the difference in trends between the treatment and control stations is linear.

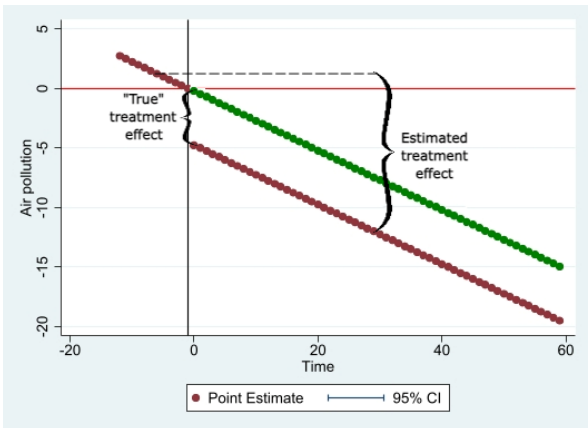
To summarize, I investigate the effect of environmental zones on air pollution through a difference-in-differences model. I run a separate regression for each of the car zones in Utrecht, Arnhem, and Rotterdam. All monitoring stations in these cities are considered treated once the environmental zone for cars is introduced, and monitoring stations in cities with an environmental zone for trucks serve as control stations. I expect to find that environmental zones lower concentrations of air pollutants. I informally test the parallel trends assumption by adding municipality-specific linear time trends and by running event studies.

Figure 1: Illustration of parallel trends assumption with event study

Panel A: Example of event study with parallel trends



Panel B: Example of event study with non-parallel trends



Note: Event study regressions in the format of equation 2, constructed by the author based on simulated data. X-axis shows the number of months away from the introduction of the environmental zone, which happens in month 0. Vertical line is the last pre-treatment period, which is the reference period. Red dots indicate the estimated treatment lead and lag coefficients, green dots indicate the true counterfactual outcomes for the treatment stations.

4.2 Test score methodology

I now turn to my second hypothesis, that environmental zones improve test scores. For this part, I use my dataset with annual school-level data of scores on the final tests in primary schools. The research method is the synthetic difference-in-differences, which is based on the classic difference-in-differences methodology and the more recent synthetic control method. I first give a brief explanation of the synthetic control method, followed by an intuitive explanation of the synthetic difference-in-differences methodology. This is followed by a more technical explanation of how this synthetic difference-in-differences model looks in my case.

The synthetic control method (SCM) has been developed and popularized in the past two decades. The goal of this method is to investigate the effect of policies for a small number of large units, such as cities or countries, on an aggregate outcome (Abadie, 2021). For example, Abadie et al. (2010) study the effect of a form of anti-tobacco legislation in California on cigarette sales in California. This legislation was Proposition 99, which imposed a tax on packs of cigarettes, amongst others. The SCM can be seen as a type of comparative case study, where the effect of Proposition 99 is estimated by comparing the evolution of cigarette sales per capita in California to states that are similar but did not introduce anti-tobacco legislation. Before the SCM was introduced, a comparative case study would have likely picked one state to compare California to. However, it can be difficult to hand-pick a state that is a credible counterfactual for California. Instead, the SCM is built on the premise that a combination of unaffected states can serve as a better comparison than a single unaffected state. It does this by using a data-driven approach to construct a synthetic control state for California. The authors first select a set of states that serve as the candidate states, from which certain states will be selected as part of the synthetic control. These candidate states are states that did not introduce any large anti-tobacco legislation in the data period. They then select characteristics that predict cigarette sales per capita, such as GDP per capita, the price of cigarettes, the share of youths, and cigarette sales per capita itself in multiple years before Proposition 99 was introduced. The synthetic control is constructed by assigning different weights to each candidate state, with the weights summing to one. This linear combination of candidate states forms a synthetic control for California, with the weights being chosen to minimize the difference in the pre-intervention characteristics between the synthetic control and California. The goal is to make the cigarette sales per capita trend of the synthetic control equal to cigarette sales per capita in California, in the period before Proposition 99 was introduced. Including pre-intervention characteristics other than cigarette sales per capita itself help in ensuring that the synthetic California does not follow California's cigarette sales per capita because of chance. Cigarette sales can be volatile, but these other characteristics make it more convincing that cigarette sales in California and the synthetic control are driven by similar underlying factors. This makes it more likely that cigarette sales per capita would have evolved in a similar fashion if Proposition 99 had not been introduced. The treatment effect of Proposition 99 is then the observed difference in cigarette sales per capita, between California and its synthetic control, after Proposition 99 is introduced (Abadie et al., 2010; Abadie et al., 2015; Abadie, 2021).

The synthetic difference-in-differences (SDID) methodology combines characteristics of the classical difference-in-differences methodology with characteristics of the synthetic control method. The SDID can be used on panel data to compare multiple treated units, who become treated after some time period in this dataset, to multiple untreated units, who are never treated. The SDID then chooses unit weights, using a form of penalized least squares, so that the pre-treatment outcome trends of the untreated units run parallel to the pre-treatment outcome trends of the treated units. In contrast to the synthetic control method, only the outcome variable is used for this matching, and not any covariates.[2] Additionally, the outcome trends only have to run parallel rather than be equal in levels. It also selects time weights, by solving for a form of least squares, to remove the importance of pre-treatment periods that are very different from the post-treatment periods. After those weights are chosen, it runs a standard difference-in-differences regression on the reweighted panel. The treatment effect, τ, is given by solving the following two-way fixed effects regression:

$$\left(\hat{\tau}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \underset{\tau,\mu,\alpha,\beta}{\arg\min}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau)^2 \hat{\omega}_i \hat{\lambda}_t\right\}. \tag{3}$$

Here, Y_{it} is the average standardized test score for school i at time t, μ is an intercept, α_i is a school-fixed effect, and β_t is a time-fixed effect. W_{it} is a treatment indicator with a value of 1 if a school is in the city with the environmental zone of interest, and a value of 0 otherwise. $\hat{\omega}$ are the unit weights and $\hat{\lambda}$ are the time weights (Arkhangelsky et al., 2021). I expect τ to be positive, as my second hypothesis is that environmental zones improve test scores. Before this regression can be run, $\hat{\omega}$ and $\hat{\lambda}$ have to be chosen. Arkhangelsky et al. (2021) provide a full overview of the methods that are used to choose each of those groups of weights, and I discuss them here in a more limited form.

To choose the unit weights, the following equation is solved:

$$(\hat{\omega}_0, \hat{\omega}^{sdid}) = \underset{\omega_0 \in \mathbb{R}, \omega \in \Omega}{\operatorname{argmin}} \sum_{t=1}^{T_{pre}}\left(\sum_{i=1}^{N_{co}}\omega_i Y_{it} - \frac{1}{N_{tr}}\sum_{i=N_{co}+1}^{N} Y_{it} + \omega_0\right)^2 + \zeta^2 T_{pre}\|\omega\|_2^2. \tag{4}$$

I discuss the main intuition of this equation here, with a full algebraic overview of these terms in Appendix B. I first focus on the first part of the equation. N_{co} is the number of schools in cities without an environmental zone, N_{tr} is the number of schools in the city with the environmental zone of interest, and N is the total number of schools. This equation is solved by assigning a unit weight ω_i to each control school i. As the first two terms between brackets show, this is done by minimizing the squared difference between the average test score of these reweighted control schools and the

[2] Covariates can also be used with the SDID, but they serve a different function compared to the synthetic control method. Covariates are not used in constructing the unit weights, as they are for the synthetic control. Instead, covariates can be included by running the SDID on the residual of the regression of the outcome variable on the covariates (Arkhangelsky et al., 2021). I also ran all my regressions with the share of migrants, share of females, and measures of the school weight included as covariates. These covariates did not change my results either quantitatively or qualitatively, but generally only increased my standard errors due to a lower number of observations because of missing values for these variables. I thus do not report the results of these regressions.

average test score in treated schools. The sum is taken over each of the T_{pre} years before the environmental zone of interest is introduced. However, there is also the additional ω_0 term. This is not a weight that is specific to each unit, but instead a simple constant that allows for a constant difference in average test scores between the two groups of schools. This weight relaxes the requirement that the trends in average test scores for the two groups have to be equal to each other, and instead only have to run parallel to each other. This is possible because the main regression equation includes school-fixed effects, which capture differences between schools that stay constant over time. The final term, which is outside brackets, is meant to put some restrictions on the unit weights that are chosen. This can be referred to as a penalization term, hence the name penalized least squares. This term is minimized by keeping the unit weights closer to zero. The goal of this term is to ensure that the set of weights that solve the equation is unique, as well as to get a more dispersed set of weights (Arkhangelsky et al., 2021). This is also known as regularization, and is particularly useful in situations with more control units than pre-treatment periods. Not including regularization can then result in an imprecise estimator (Doudchenko and Imbens, 2016).

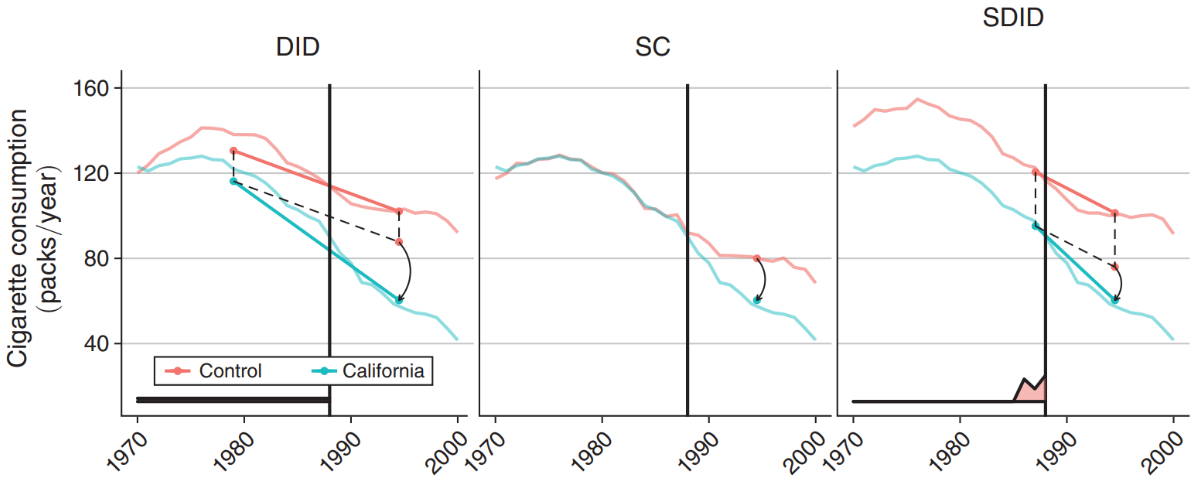
The time weights, on the other hand, are chosen by solving the following equation:

$$(\hat{\lambda}_0, \hat{\lambda}^{sdid}) = \underset{\lambda_0 \in \mathbb{R}, \lambda \in \Lambda}{\operatorname{argmin}} \sum_{i=1}^{N_{co}} \left(\sum_{t=1}^{T_{pre}} \lambda_t Y_{it} - \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^{T} Y_{it} + \lambda_0 \right)^2. \tag{5}$$

These time weights assign a weight to each year before the environmental zone is introduced. They do this by minimizing the squared sum of differences between the test score of each control school in the reweighted years and the average test score of this control school in the years after the environmental zone is introduced. However, there is again a constant term λ_0 included as well, which allows for a constant difference between these averages (Arkhangelsky et al., 2021). Importantly, there is no regularization in the form of a penalization term here. One of the reasons for this is that the authors assume that units can be interchangeable in this setting, but time periods might not be (Athey, 2021). They thus allow the time weights to limit the importance of years before the environmental zone was introduced that differ substantially from the years after the environmental zone was introduced, which can improve precision (Arkhangelsky et al., 2021).

Figure 2, taken from Arkhangelsky et al. (2021), shows the difference between the classical difference-in-differences, the synthetic control method, and the synthetic difference-in-differences, using the example of Proposition 99. The classical difference-in-differences simply compares California to the average of all other states, and calculates the treatment effect by assuming parallel trends. The synthetic control method first tries to construct a synthetic California that is very close in the level of cigarette sales per capita to California, and then calculates the treatment effect by taking the difference between their cigarette sales after Proposition 99 is introduced. Finally, the synthetic difference-in-differences merely tries to match the average trend of cigarette sales per capita of the control states to the trend in California. The treatment effect is calculated by assuming that the reweighted control states are on a parallel trend to California.

Figure 2: Comparison of classical difference-in-differences, synthetic control, and synthetic difference-in-differences



Note: Adapted from "Synthetic Difference-in-Differences", by D. Arkhangelsky, S. Athey, D.A. Hirshberg, G.W. Imbens, and S. Wager, 2021, *American Economic Review, 111*(12), p. 4095 (https://doi.org/10.1257/aer.20190159). Copyright American Economic Association; reproduced with permission of the American Economic Review. Each graph shows the trend of cigarette sales per capita in California and control states per year. The vertical lines indicate the year Proposition 99 was introduced. Arrows indicate the estimated treatment effect for the method listed at the top of each graph.

The main advantage of the SDID over a synthetic control method for my analysis is that the SDID can be used on panels with a large number of treated and control units, whereas the synthetic control method is usually applied to a small number of aggregate units. It would be possible to first calculate the mean test score per year for each city, and then apply the synthetic control method. However, this would substantially decrease my number of observations. Since I have data at the school-level, the SDID allows me to use my more extensive school panel data set to get more precise estimates. The main disadvantage of the difference-in-differences methodology over the SDID is that the treatment and control units might not have parallel trends using raw data, even before treatment occurs. In contrast, the SDID assigns weights to the control units to make the pre-intervention trends parallel to the treated units, and then applies the DID method on this reweighted panel (Arkhangelsky et al., 2021). I am only able to use the SDID for test scores, because the air pollution data have missing values for some periods. The SDID method will not work when the outcome variable has missing values.

As in section 4.1, I consider every school in the city with the environmental zone of interest as treated to capture potential spillover effects. Here, I do perform a robustness check where only those schools which are actually within the borders of the environmental zone are considered treated. This means I drop the schools within the treated cities that are not inside the borders of the environmental zones. Not doing this would likely result in spillover effects within cities biasing my estimates (Cunningham, 2021).

I run a separate regression for each environmental zone of interest. My test score data set covers more years than my air pollution data set, so that I also have test score data before an environmental zone for trucks was introduced in Arnhem in July 2014. This means that in addition to the three car zones in Utrecht, Arnhem, and Rotterdam, I can also investigate the effect of the truck zone in Arnhem for this section. Control units for the regressions of the car zones are schools in cities with environmental zones for trucks. This is based on the same logic as in section 4.1, as these cities had a truck zone in place before they introduced a car zone. For the regression of the truck zone in Arnhem, I use units without any environmental zone as controls because Arnhem did not have any environmental zone in place before it introduced its truck zone. Arnhem introduced its truck zone in July 2014, and its car zone in January 2019. I thus drop the period before July 2014 for my regression of the car zone, and I drop the period after December 2018 for my regression of the truck zone. Not doing this would result in multiple "treatments" occurring in Arnhem in each analysis.

After this method produces an estimate of the treatment effect, I want to get some idea of how unlikely it is that this effect is actually due to chance. Just like the synthetic control method, the synthetic difference-in-differences method cannot use conventional standard errors to evaluate whether the result is statistically significant. Instead, Arkhangelsky et al. (2021) suggest three ways to calculate standard errors with this approach. I opt to use placebo standard errors, which, in contrast to the other two options, are also reliable if the number of treated schools is small. Calculating the standard errors by using placebo evaluations is also similar to the approach usually taken when using the synthetic control method. In essence, this method takes the control schools, assigns a fake treatment to some of them, and then repeats the main regression. This produces placebo treatment effects for these control schools. The reported p-value for the treatment effect of interest is then the proportion of placebo treatment effects that is larger in absolute size than the actual estimated treatment effect of interest (Abadie, 2021; Arkhangelsky et al., 2021). For example, assume I get an estimate of the treatment effect that is equal to 0.1. If all the placebo treatment effects are larger than 0.1, the estimated treatment effect of interest was most likely a chance finding rather than the "true" treatment effect. This will be reflected by the p-value, which would then be equal to 1. Following conventional p-value thresholds of 0.01, 0.05 and/or 0.1, I would then conclude that I do not find evidence for an effect of an environmental zone on test scores. I run 200 placebo replications for each regression to calculate the p-value.

To summarize, I investigate the effect of environmental zones on test scores by using the synthetic difference-in-differences method. I run a separate regression for each of the car zones in Utrecht, Arnhem, and Rotterdam, and for the truck zone in Arnhem. All monitoring stations in these cities are considered treated once the environmental zone of interest is introduced. The control stations are stations in cities with a truck zone for the regression of car zones, and stations in cities without any environmental zone for the regression of the truck zone. The synthetic difference-in-differences assigns weights to these control schools to match the average trend of test scores for treated schools before the environmental zone is introduced. In line with my second hypothesis, I expect to find that environmental zones lower concentrations of air pollutants.

5. Results

This section reports my results for both hypotheses. Section 5.1 shows my results of the effect of environmental zones on different air pollutants, and section 5.2 present my findings of the effect of environmental zones on test scores. Finally, section 5.3 includes robustness checks to see whether my results stand up to scrutiny.

5.1 Air pollution results

I start by reporting the results of the environmental zone in Utrecht. Table 1 reports the results of my standard regression, which is Equation 1. These findings indicate that the introduction of an environmental zone for cars in Utrecht had no significant effect on concentrations of NO_2, NO and PM_{10}. Surprisingly, I find that concentrations of $PM_{2.5}$ and O_3 actually increased by more than 2 µg/m^3. Table 2 includes municipality-specific linear time trends and finds similar results. The sign of each coefficient is the same, but the effect is now significant for NO_2 and PM_{10} and no longer significant for O_3. My preferred specification is the one in table 2, as this is better able to account for possible diverging air pollution trends between Utrecht and other cities. As explained, this extrapolates any linear differences in air pollution trends that already existed before the environmental zone was introduced. Overall, these findings point towards a negative effect of the environmental zone on PM_{10} concentrations, but a positive effect on NO_2 and $PM_{2.5}$ concentrations.

Next, I investigate to what extent monitoring stations in Utrecht were on a parallel trend compared to stations from other cities by estimating an event study. These results are reported in Figure 3. The coefficient in the first month in Panel A of Figure 3 is approximately -4. The interpretation of this coefficient is that the difference in NO_2 concentrations between the first month and the last month before the car zone is introduced, is 4 µg/m^3 lower for monitoring stations in Utrecht than for the control stations. Overall, Figure 3 shows that the station- and time-fixed effects do not fully capture differences in pre-treatment trends for some pollutants. In particular, stations in Utrecht seem to be on an upward trend compared to the control stations for $PM_{2.5}$, and for NO_2 to a lesser extent. This can be seen by looking at the general pattern of the coefficients before the environmental zone is introduced. Pre-treatment trends do not show clear systematic divergences for NO, PM_{10} and O_3. Visual inspection of the post-treatment coefficients seems to point towards a negative effect on PM_{10}, a positive effect for O_3, and no effect for NO. However, it is hard to get a clear picture of the effects based on these event study graphs due to the time-fixed effects not fully picking up monthly patterns in pollutant concentrations. The upward pre-treatment trends for NO_2 and $PM_{2.5}$ might, however, explain why I found positive coefficients in Table 1, to the extent that this reveals non-parallel trends between these groups in the absence of treatment. Apparently, the municipality-specific trends in Table 2 are not able to fully capture these pre-treatment trend differences. In general, my conclusion from the findings for Utrecht is that there is some evidence for decreases in PM_{10}, and less convincing evidence for increases in NO_2 and $PM_{2.5}$. I interpret these findings as the environmental zone in Utrecht having very moderate to no effects on air pollution, with possible decreases in PM_{10} concentrations.

Table 1

Effects of the environmental zone for cars in Utrecht

	(1) NO_2	(2) NO	(3) PM_{10}	(4) $PM_{2.5}$	(5) O_3
Car zone	0.996	0.334	-0.598	2.398***	2.868**
	(0.812)	(0.671)	(0.287)	(0.092)	(0.816)
Observations	722	730	614	410	507
Station fixed effects	Y	Y	Y	Y	Y
Month of year fixed effects	Y	Y	Y	Y	Y
Municipality-specific trend	N	N	N	N	N

Note: Dependent variable is concentration of pollutant listed at the top of each column in µg/m^3. Standard errors are reported between brackets. Stars indicate p-values, with the following values: *p<0.1 **p<0.05 ***p<0.01.

Table 2

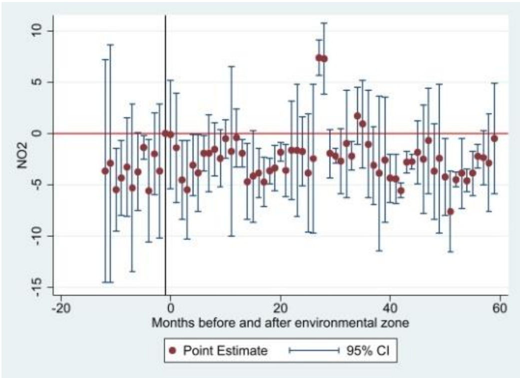
Effects of the environmental zone for cars in Utrecht with municipality-specific trends

	(1) NO_2	(2) NO	(3) PM_{10}	(4) $PM_{2.5}$	(5) O_3
Car zone	1.832**	0.193	-0.781**	2.241***	0.413
	(0.418)	(0.262)	(0.251)	(0.116)	(0.325)
Observations	722	730	614	410	507
Station fixed effects	Y	Y	Y	Y	Y
Month of year fixed effects	Y	Y	Y	Y	Y
Municipality-specific trend	Y	Y	Y	Y	Y

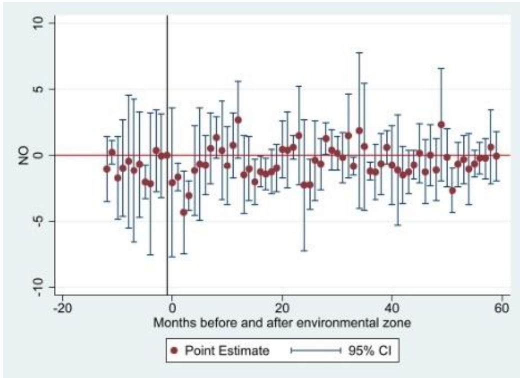
Note: Dependent variable is concentration of pollutant listed at the top of each column in µg/m^3. Standard errors are reported between brackets. Stars indicate p-values, with the following values: *p<0.1 **p<0.05 ***p<0.01.

Figure 3: Event studies of environmental zone for cars in Utrecht

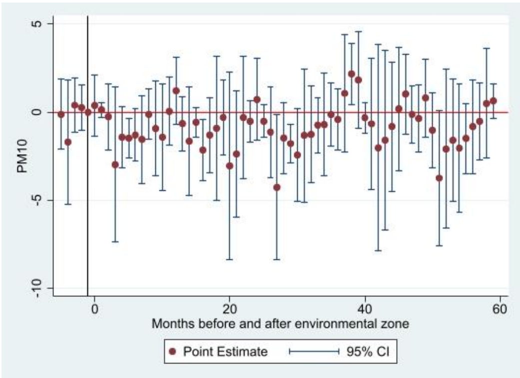
Panel A: Effect on NO_2 in μg/m^3



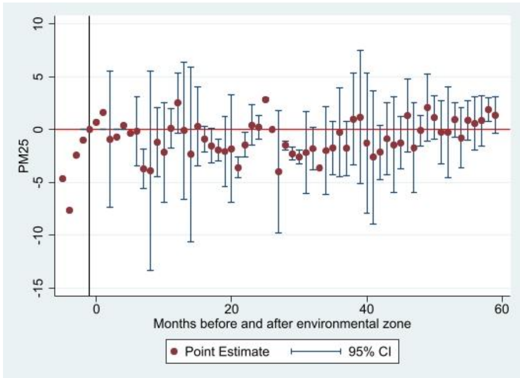
Panel B: Effect on NO in μg/m^3



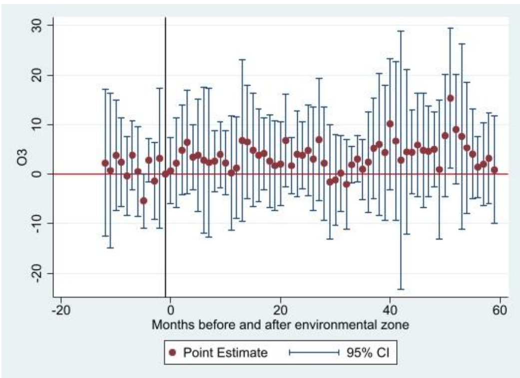
Panel C: Effect on PM_{10} in μg/m^3



Panel D: Effect on $PM_{2.5}$ in μg/m^3



Panel E: Effect on O_3 in μg/m^3



Note: Each panel represents an event study regression of the effect of the environmental zone for cars in Utrecht on the pollutant listed on the y-axis of each graph, using equation 2. The vertical line represents the last month before the environmental zone is introduced, which is the reference period. Dots before this line represent the treatment leads, and dots after this line represent the treatment lags. The lines around each dot represent the 95% confidence interval of the point estimate.

Next, I examine the results of the environmental zone for cars in Arnhem. Table 3 reports the results of Equation 1, where I find statistically significant and negative results for all pollutants. In particular, this corresponds to decreases of approximately 9% for NO_2, 2.5% for NO, 3.5% for PM_{10}, and 13% for soot, compared to their pre-treatment means in Arnhem. When including municipality-specific linear time trends, the effect only remains negative and significant for PM_{10}, although the estimated effect is now also larger in size. On the contrary, it becomes positive and significant for NO. For soot, the effect is similar in size, leading me to believe that its insignificance is driven by the large standard error rather than there actually not being an effect. Turning to Figure 4, there is a seemingly downward pre-treatment trend for NO_2 and no clear visual effect of the environmental zone after it is introduced. For NO, the trends seem to be comparable in the 2-3 years before the car zone is introduced. I believe that the positive effect found in Table 4 is mainly driven by the outlier 4 months after the environmental zone is introduced, whereas there seems to be a visible decline in the last months of my data period. For PM_{10}, trends move in a parallel fashion in the two years leading up to the environmental zone, but do not show any clear change afterwards. Finally, there is no evidence of a -trend difference for soot before the environmental zone is introduced, but there does seem to be a clear decrease in soot concentrations in the year after the environmental zone is introduced. All in all, I find evidence for a decrease in soot concentrations at Arnhem's monitoring stations, but limited evidence for an effect on the other pollutants. The effect size for soot is somewhat smaller than the effect predicted by Van de Poll et al. (2017), which was 17%.

Table 3

Effects of the environmental zone for cars in Arnhem

	(1) NO_2	(2) NO	(3) PM_{10}	(4) Soot
Car zone	-3.032**	-0.620*	-0.748*	-0.200***
	(0.440)	(0.289)	(0.313)	(0.004)
Observations	452	454	498	249
Station fixed effects	Y	Y	Y	Y
Month of year fixed effects	Y	Y	Y	Y
Municipality-specific trend	N	N	N	N

Note: Dependent variable is concentration of pollutant listed at the top of each column in $\mu g/m^3$. Standard errors are reported between brackets. Stars indicate p-values, with the following values: *p<0.1 **p<0.05 ***p<0.01.

Table 4

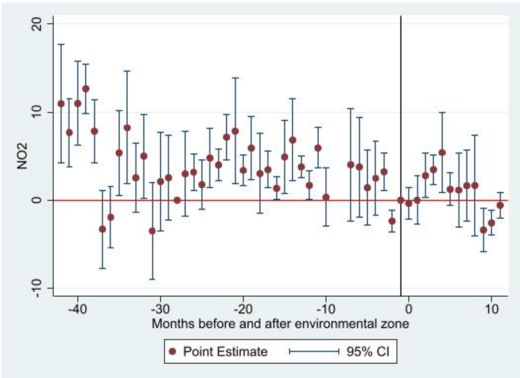
Effects of the environmental zone for cars in Arnhem with municipality-specific trends

	(1) NO_2	(2) NO	(3) PM_{10}	(4) Soot
Car zone	0.190 (0.254)	4.537*** (0.569)	-2.801*** (0.543)	-0.147 (0.073)
Observations	452	454	498	249
Station fixed effects	Y	Y	Y	Y
Month of year fixed effects	Y	Y	Y	Y
Municipality-specific trend	Y	Y	Y	Y

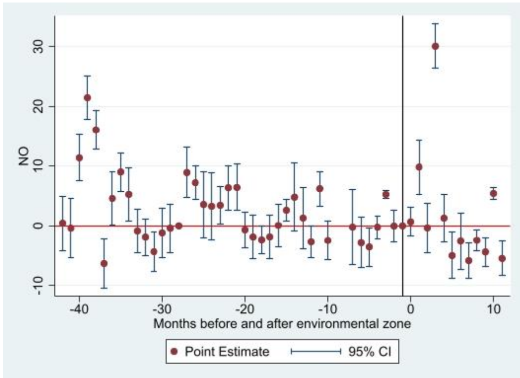
Note: Dependent variable is concentration of pollutant listed at the top of each column in µg/m^3. Standard errors are reported between brackets. Stars indicate p-values, with the following values: *p<0.1 **p<0.05 ***p<0.01.

Figure 4: Event studies of environmental zone for cars in Arnhem

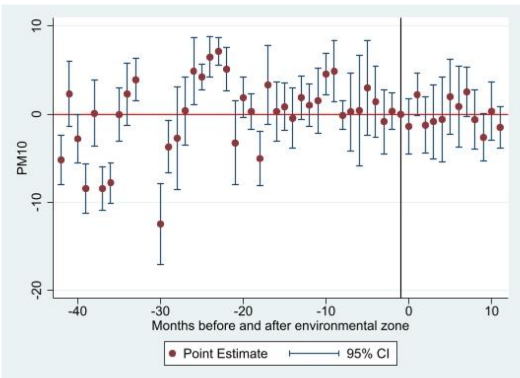
Panel A: Effect on NO_2 in µg/m^3



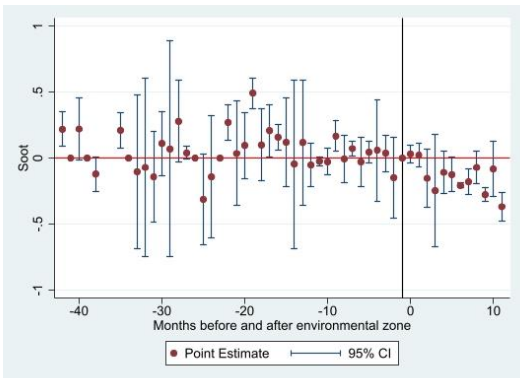
Panel B: Effect on NO in µg/m^3



Panel C: Effect on PM_{10} in µg/m^3



Panel D: Effect on soot in µg/m^3



Note: Each panel represents an event study regression of the effect of the environmental zone for cars in Arnhem on the pollutant listed on the y-axis of each graph, using equation 2. The vertical line represents the last month before the environmental zone is introduced, which is the reference period. Dots before this line represent the treatment leads, and dots after this line represent the treatment lags. The lines around each dot represent the 95% confidence interval of the point estimate.

Finally, I examine the effects of the environmental zone in Rotterdam, where Table 5 presents the results of my main regression. These findings indicate that the environmental zone in Rotterdam led to decreases in PM_{10} and soot concentrations, but to an increase in the concentration of NO_2.

When adding a municipality-specific time trend in Table 6, the sign of each effect remains the same. However, the effect for soot is no longer statistically significant. Since the coefficients are so similar, however, this might point towards the larger standard error being the culprit rather than there actually being no effect. The similarity of coefficients between these tables seems to indicate that the initial model was able to ensure parallel trends in the pre-treatment period fairly well already. I check this by looking at the event studies in Figure 5. For NO_2, there seems to be somewhat of an upward trend visible in the pre-treatment period for Rotterdam compared to other stations. Again, this could explain why I find a positive coefficient here. For the other pollutants, the pre-treatment trends seem to run fairly parallel between the two groups. In terms of the post-treatment effects, both NO and soot show an initial decrease in concentrations in the first few months after treatment. However, this effect seems to fade away rather quickly. For PM_{10}, on the other hand, concentrations seem to decrease in a slower but more persistent manner. Based on these event studies and my difference-in-difference regressions, I conclude that the environmental zone in Rotterdam seems to have decreased PM_{10} concentrations, but not affected concentrations of other pollutants. Compared to the pre-treatment mean PM_{10} concentration in Rotterdam, the effect is equal to a decrease of approximately 4%.

To summarize this section, I find very modest evidence for my first hypothesis, that environmental zones lower air pollution. For Utrecht, I find practically no effect, except for possible decreases in in PM_{10} concentrations. This effect size is in the range of -0.5 to -1 $\mu g/m^3$. In Arnhem, there seems to be a small effect on soot concentrations, but no clear effects on other pollutants. The effect on soot concentrations is approximately -0.2 $\mu g/m^3$. Finally, the environmental zone in Rotterdam only seems to have lowered PM_{10} concentrations by nearly 1 $\mu g/m^3$.

Table 5

Effects of the environmental zone for cars in Rotterdam

	(1) NO_2	(2) NO	(3) PM_{10}	(4) $PM_{2.5}$	(5) O_3	(6) Soot
Car zone	0.765**	-0.382	-0.841***	-0.527	-0.134	-0.026*
	(0.191)	(0.251)	(0.124)	(0.296)	(1.174)	(0.007)
Observations	936	941	862	575	575	448
Station fixed effects	Y	Y	Y	Y	Y	Y
Month of year fixed effects	Y	Y	Y	Y	Y	Y
Municipality-specific trend	N	N	N	N	N	N

Note: Dependent variable is concentration of pollutant listed at the top of each column in μg/m^3. Standard errors are reported between brackets. Stars indicate p-values, with the following values: *p<0.1 **p<0.05 ***p<0.01.

Table 6

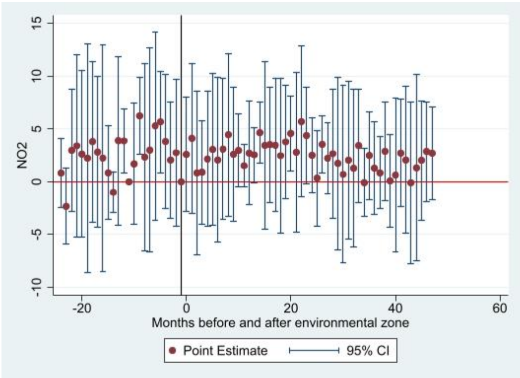
Effects of the environmental zone for cars in Rotterdam with municipality-specific trends

	(1) NO_2	(2) NO	(3) PM_{10}	(4) $PM_{2.5}$	(5) O_3	(6) Soot
Car zone	0.980***	-0.171	-0.952***	-0.008	-0.620	-0.023
	(0.110)	(0.277)	(0.132)	(0.484)	(0.925)	(0.028)
Observations	936	941	862	575	575	448
Station fixed effects	Y	Y	Y	Y	Y	Y
Month of year fixed effects	Y	Y	Y	Y	Y	Y
Municipality-specific trend	Y	Y	Y	Y	Y	Y

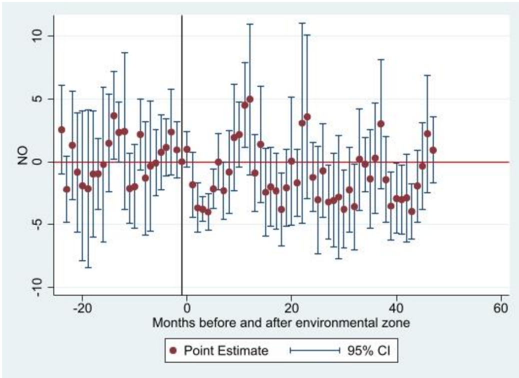
Note: Dependent variable is concentration of pollutant listed at the top of each column in μg/m^3. Standard errors are reported between brackets. Stars indicate p-values, with the following values: *p<0.1 **p<0.05 ***p<0.01.

Figure 5: Event studies of environmental zone for cars in Rotterdam

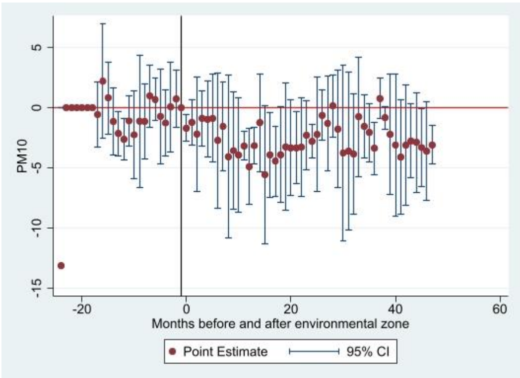
Panel A: Effect on NO_2 in $\mu g/m^3$



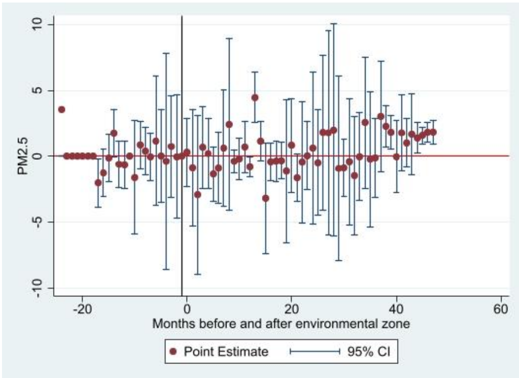
Panel B: Effect on NO in $\mu g/m^3$



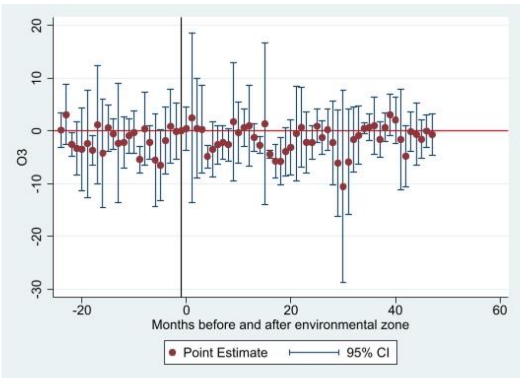
Panel C: Effect on PM_{10} in $\mu g/m^3$



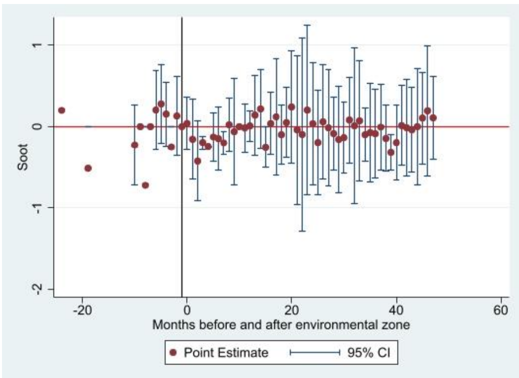
Panel D: Effect on $PM_{2.5}$ in $\mu g/m^3$



Panel E: Effect on O_3 in $\mu g/m^3$



Panel F: Effect on soot in $\mu g/m^3$



Note: Each panel represents an event study regression of the effect of the environmental zone for cars in Utrecht on the pollutant listed on the y-axis of each graph, using equation 2. The vertical line represents the last month before the environmental zone is introduced, which is the reference period. Dots before this line represent the treatment leads, and dots after this line represent the treatment lags. The lines around each dot represent the 95% confidence interval of the point estimate.

5.2 Test score results

After establishing the effects that the environmental zones had on air pollution in Utrecht, Arnhem and Rotterdam, I can now estimate the effect that these have had on test scores. Table 7 reports the results of my regression following Equation 3, with the synthetic difference-in-differences approach.

The results of these regressions in Table 7 largely point to there being no effect of environmental zones on test scores, with the exception of the environmental zone for trucks in Arnhem. Unfortunately, I was not able to estimate the effect of this environmental zone on air pollution in section 5.1. This effect on test scores would point to the truck zone having increased local air pollution. Figure 6 allows me to assess the pre-treatment fit of the reweighted control schools for all regressions. These reweighted control schools are meant to have a parallel trend in test scores on average to the treated schools in the pre-treatment period. Visual inspection can indicate whether this synthetic diff-in-diff is able to match the treated schools well. The fits seem to be particularly good for the environmental zones in Arnhem and Rotterdam, where trends move in a parallel fashion in the pre-treatment period. The fit for Utrecht in Panel A is not as good, and the post-treatment effects do not show a clear pattern. Panel B of Figure 6 seems to indicate that standardized test scores diverge between Arnhem and the synthetic counterfactual immediately after the environmental zone is introduced. Panel C and D do not show any such effects.

One other possibility is that these effects are not actually driven by a decrease in performance, but rather by changes in the test that is chosen. Schools in Arnhem might have switched to tests where they would have ranked lower in the distribution of test scores regardless of whether or not the environmental zone had been introduced. To investigate whether this explains these findings, I restrict my regressions to schools which always took the CITO test, which is the most popular test. This ensures a better comparability of test scores across years, both within schools and across schools, but also leads to less observations for my regressions. Doing this leaves me with 33 schools from Utrecht, 35 schools from Arnhem, and 68 schools from Rotterdam. For these regressions, I calculate standardized test scores over the sample of schools that always use the CITO test.

Table 7

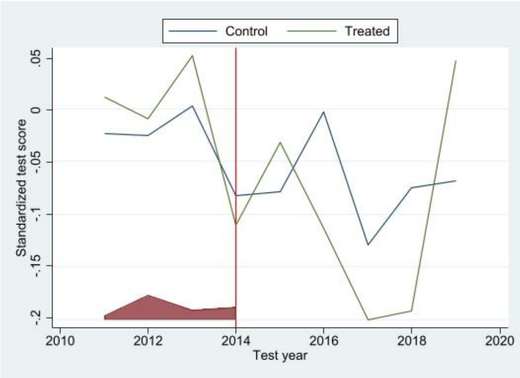
Effects of environmental zones on standardized test scores

	(1) Car zone in Utrecht	(2) Truck zone in Arnhem	(3) Car zone in Arnhem	(4) Car zone in Rotterdam
Environmental zone	-0.040 (0.067)	-0.189* (0.097)	0.012 (0.137)	0.055 (0.063)

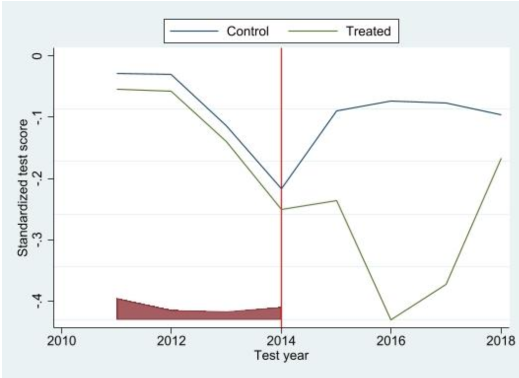
Note: Dependent variable is average standardized test score on final exam. Each column represents a single regression to estimate the effect of one environmental zone in a city. Standard errors are reported between brackets and calculated using placebo replications. Stars indicate p-values, with the following values: *p<0.1 **p<0.05 ***p<0.01.

Figure 6: Synthetic difference-in-difference estimates of environmental zones

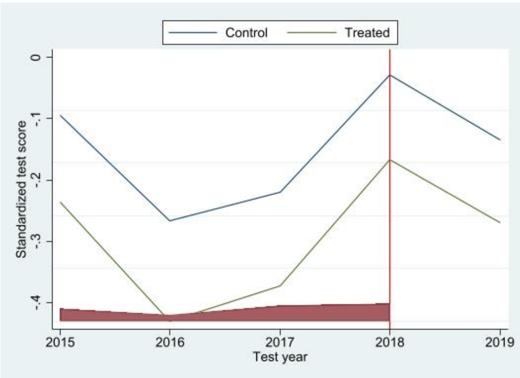
Panel A: Effect on environmental zone for cars in Utrecht on standardized test scores



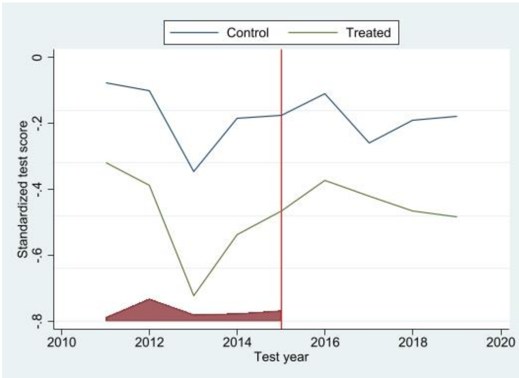
Panel B: Effect on environmental zone for truck in Arnhem on standardized test scores



Panel C: Effect on environmental zone for cars in Arnhem on standardized test scores



Panel D: Effect on environmental zone for cars in Rotterdam on standardized test scores



Note: Each panel represents a synthetic difference-in-differences regression to estimate the effect of the environmental zone listed at the top of each panel on standardized test scores. The red line represents the last year before the environmental zone of interest becomes active. The green line shows the trend in average standardized test score for schools in the city where the environmental zone of interest is introduced. The blue line shows the trend in average standardized test score for the reweighted control schools. The red area at the bottom of each graph indicate the time weights and show which pre-treatment years receive more importance in the regression.

I report the results of the regression with only CITO-schools in Table 8. The sign of each coefficient has stayed the same, but sizes have changed. In particular, I no longer find that the truck zone in Arnhem has had a statistically significant negative effect on test scores. On the other hand, I now find that the car zone in Rotterdam increased test scores by 0.2 standard deviations. Figure 7 allows me to assess to what extent the synthetic difference-in-difference control is able to match the pre-intervention trend of the treated schools. For Utrecht, test scores seem to be on a somewhat downward trend compared to the reweighted control schools. The difference in 2011 is approximately 0.1 standard deviation, which has nearly disappeared in 2014, the last pre-treatment year. If this trend is an indication of how Utrecht's counterfactual of not having an environmental zone would have evolved, the coefficient I find is a lower bound and should be expected to be higher in reality. For both environmental zones in Arnhem, in Panel B and C, the two trends move in a parallel fashion. This gives me more confidence that the effects I find are close to the true effects, although the negative coefficient for the truck zone remains puzzling. Finally, Panel D shows that test scores in Rotterdam are on an upward trend compared to the control schools. The difference is nearly 0.5 standard deviation in 2011, but less than 0.4 standard deviation in 2015. If this trend would have continued in the absence of the environmental zone, the coefficient in Table 8 is an upper bound for the effect of the car zone on test scores in Rotterdam.

It should be noted that an effect size of 0.2 standard deviation is very sizable in the education literature. For example, Bloom et al. (2008) find that American students increase their test scores by approximately 0.4 SD from grade 5 to grade 6. These results are not entirely comparable, as they also mention that student-level deviations are generally larger than school-level standard deviations. However, they provide some indication of the effect size.

Table 8

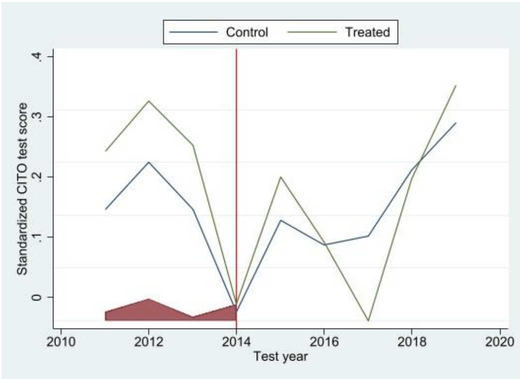
Effects of environmental zones on standardized CITO test scores

	(1) Car zone in Utrecht	(2) Truck zone in Arnhem	(3) Car zone in Arnhem	(4) Car zone in Rotterdam
Environmental zone	-0.076 (0.085)	-0.136 (0.097)	0.005 (0.135)	0.209*** (0.076)

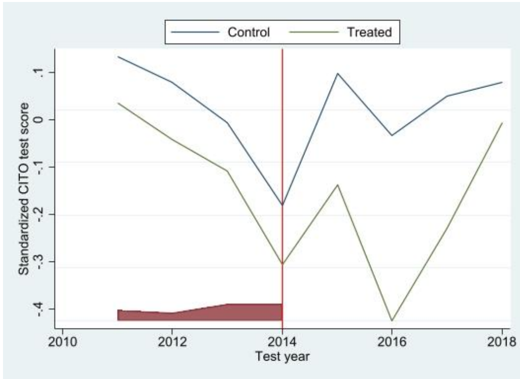
Note: Dependent variable is average standardized CITO test score on final exam. Each column represents a single regression to estimate the effect of one environmental zone in a city. Standard errors are reported between brackets and calculated using placebo replications. Stars indicate p-values, with the following values: *p<0.1 **p<0.05 ***p<0.01.

Figure 7: Synthetic difference-in-difference estimates of environmental zones

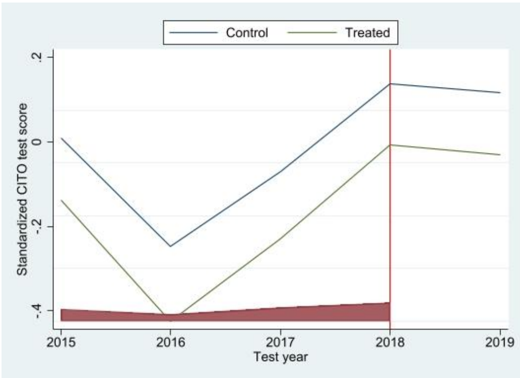
Panel A: Effect on environmental zone for cars in Utrecht on standardized CITO test scores



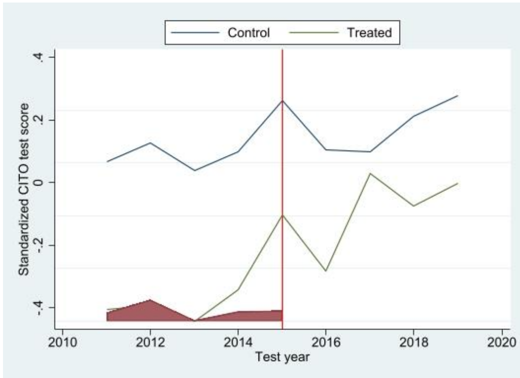
Panel B: Effect on environmental zone for truck in Arnhem on standardized CITO test scores



Panel C: Effect on environmental zone for cars in Arnhem on standardized CITO test scores



Panel D: Effect on environmental zone for cars in Rotterdam on standardized CITO test scores



Note: Each panel represents a synthetic difference-in-differences regression to estimate the effect of the environmental zone listed at the top of each panel on standardized CITO test scores. The red line represents the last year before the environmental zone of interest becomes active. The green line shows the trend in average standardized test score for schools in the city where the environmental zone of interest is introduced. The blue line shows the trend in average standardized test score for the reweighted control schools. The red area at the bottom of each graph indicate the time weights and show which pre-treatment years receive more importance in the regression.

5.3 Robustness checks

This section serves to corroborate my findings of section 5.2 in particular, since the scope for testing the results of section 5.1 is very limited due to scarce data. The results of my robustness checks can be found in Appendix C.

I first investigate the effect on schools that are actually located inside the environmental zones. As control units for the effects of environmental zones for cars, I use schools located in environmental zones for trucks in other cities. As control units for the effects of the environmental zone for trucks, I use schools that are never in an environmental zone. Table C1 shows the results of these regressions. Here, I find no statistically significant effect for any of the environmental zones. All effects are positive and small, with the exception of the environmental zone for trucks in Arnhem. That estimate is large and negative, but still not statistically significant. Although Panel B of Figure C1 shows that the pre-intervention fit is very good, the standard error is very large, most likely due to the fact that this regression only uses 1 school located inside the environmental zone in Arnhem. Overall, these results do not show any effects of the environmental zones on test scores.

Next, I test whether the environmental zones caused compositional changes in the student populations of schools. For this, I use the variables relating to the school weight. Students are assigned a weight based on the education level of their parents, where the weight can take on the values 0, 0.3, and 1.2. A student receives a weight of 0.3 if both of their parents completed no more than one of the lower levels of secondary education. A weight of 1.2 is received when one parent has only completed primary education and the other parent has at a maximum completed one of the lower levels of secondary education. If a student does not fall into either category, they receive a weight of 0 (DUO, 2014-2018). For each school, I calculate the number of students with a weight of 1.2 as a share of the total number of students in that school per year. If the type of sorting occurs where families of higher socioeconomic statuses move to areas with environmental zones, I would expect the share of students with a weight of 1.2 to decrease following the introduction of an environmental zone.

The results of the synthetic difference-in-differences, where the share of students with the highest weight is the dependent variable, are shown in Table C2. The coefficients are negative but statistically insignificant, with the exception of Rotterdam in column 4. Figure C2 shows that pre-treatment trends run parallel for all regressions, making it more likely that the parallel trends assumption holds. For Rotterdam, I find that there is a statistically significant decrease in the share of students with a weight of 1.2 following the introduction of the environmental zone for cars, in the order of 0.7 percentage points. This might explain why, in section 5.2, I found that the environmental zone led to an increase in CITO test scores in Rotterdam. It was not (only) through the mechanism of less air pollution improving cognitive performance, but rather by inducing a change in student composition towards more students of high socio-economic status, that the environmental zone increased test scores.

6 Discussion

Although I find some indications in section 5.1 that environmental zones lower air pollution, I do not find convincing proof that environmental zones improve test scores. This is contradictory to the findings of most studies on this topic, which show that air pollution lowers cognitive performance. This section discusses possible reasons for this discrepancy.

One issue relates to the possible presence of anticipation effects. Anticipation effects can bias results, and one assumption of the synthetic control method is in fact that they are not present (Adabie, 2021). Although the assumptions of the synthetic difference-in-differences have not yet been formalized as such, this seems like a necessary assumption for this method as well. In my setting, anticipation effects could arise if individuals already switch to cleaner vehicles before the environmental zone is introduced, or if families of higher socioeconomic status already move to areas inside environmental zones before the zone becomes active. My results in section 5.1 did not seem to show evidence of the former, as effects (where visible) only started after the intervention was put in place. The latter might be more concerning, if the synthetic difference-in-differences matches control units to the treated who already saw a change in test scores before the environmental zone was actually active. It could then be matching the treated schools to control schools that saw a similar shock in test scores by chance, rather than actually having similar processes affecting the development of test scores. I thus quickly discuss when the environmental zones of interest were announced, which provides an indication of whether anticipation effects were possible. The environmental zone in Utrecht was announced in November of 2013, which is six months before the final pre-treatment test in Utrecht (Het Parool, 2013). Arnhem's environmental zone for trucks was only officially decided a week before it was enacted, and its car zone was announced just a month before the final pre-treatment test (*Besluit Milieuzone Arnhem,* 2014; Van der Vegt, 2019a). Finally, the environmental zone in Rotterdam was announced one month after the final pre-intervention test ("Gemeenteraad stemt in", 2015). This thus indicates that the scope for anticipation effects is likely limited in my analyses.

Another possible explanation is that the average effect on air pollution in a city is different from the effect I find. This could be the case because the measuring stations for which I have data are not randomly chosen. Instead, they are usually put either near a street, or away from busy streets to serve as a background station. This choice of locations for measuring stations could thus either lead to underestimation or overestimation of the average effect of environmental zones on local air pollution. Furthermore, even if I were to assume that the locations of these stations are randomly chosen, my findings do not provide a full picture. This is caused by the fact that I have missing data for practically all measuring stations, which means I was unable to estimate the effect on all air pollutants of interest. Finally, because these stations are generally not actually located inside the environmental zone, I am unable to see whether there is a larger effect in the area covered by an environmental zone than in surrounding areas.

A final explanation, which is a more fundamental problem, is that environmental zones have too limited of an effect on air pollution for them to affect test scores. I found that Rotterdam's environmental zone caused concentrations of PM_{10} to decrease by approximately 0.9 $\mu g/m^3$, and Arnhem's environmental zone for cars caused concentrations of soot to decrease by 0.2 $\mu g/m^3$. As a comparison, Carneiro et al. (2021) find that a 10 $\mu g/m^3$ increase in PM_{10} lowers test scores by 8% of a standard deviation. If the effects are linear, this would imply that Rotterdam would see an increase in test scores of 0.008 standard deviations due to its environmental zone. It might be the case that air pollution does not have a negative effect on cognitive performance below a certain threshold value, similar to the way that there are threshold values set for adverse impacts of pollutants on health. Contrarily, it might be the case that the effect is there, but that my research method is not the best way to find it. Average test scores for schools can fluctuate from year to year, and the difference-in-difference model only finds the true effect if the parallel trends assumption holds. If the treated schools are not on a perfectly parallel trend to the control schools, it is questionable whether this method would be able to isolate an effect of just 0.008 SD. This is also reflected in the size of my standard errors, which in some regressions showed to be more than ten-fold this expected effect.

7 Conclusion

Does air pollution lower test scores? Experimental evidence on the effect of short-term exposure to air pollution by Shehab and Pope (2019) indicates that it would, but the effect of long-term exposure can only be tested using observational data. My thesis is not able to corroborate the findings of previous economic studies, which found that higher air pollution on test days lowers test scores. I find no clear evidence that the environmental zone in Utrecht had an effect on air pollution. I do find that the environmental zone in Rotterdam lowered concentrations of PM_{10}, and that the environmental zone for cars in Arnhem lowered concentrations of soot. However, this does not seem to translate to increases in test scores. I do find positive results on test scores for Rotterdam, but these seem to be contaminated by non-parallel trends and composition effects. These findings thus indicate that environmental zones do not have the additional bonus of improving cognitive performance, besides their intended effects on human health.

For future research, my thesis indicates that policy introductions are not the best approach to estimate the causal effect of air pollution on cognitive performance going forward. Environmental zones are one of the primary policies to combat local air pollution, but they seem to have too limited of an effect to credibly estimate their impact on test scores. This thus indicates that other exogenous sources of variation in air pollution will have to be used in future research to retrieve the causal effect of air pollution on test scores, or cognitive performance more generally. In this regard, I believe the approach taken by Carneiro et al. (2021) to be promising. Wind direction can lead to plausibly exogenous variation in air pollution, if they can isolate this relationship from other mechanisms through which wind direction influences test scores, such as its effect on pollen levels. I look forward to seeing what this exciting literature brings in the coming years.

References

40 steden. (n.d.). G40 Stedennetwerk. https://www.g40stedennetwerk.nl/40-steden.

Abadie, A., Diamond, A., & Hainmueller, J. (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.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*(2), 495-510.

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, *59*(2), 391-425.

Almond, D., Edlund, L., & Palme, M. (2009). Chernobyl's subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in Sweden. *The Quarterly Journal of Economics*, *124*(4), 1729-1772.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University press.

ANWB. (n.d.). *Wat is de emissieklasse van mijn auto?* https://www.anwb.nl/auto/nieuws-en-tips/wat-is-de-emissieklasse-van-mijn-auto.

ANWB. (2019, December 6). *Milieuzone Rotterdam wordt opgeheven*. https://www.anwb.nl/verkeer/nieuws/nederland/2018/juni/milieuzone-rotterdam-wordt-opgeheven

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, *111*(12), 4088-4118.

Athey, S. (2021, January 13). *Susan Athey: Synthetic Difference in Differences* [Video]. YouTube. https://www.youtube.com/watch?v=r2DzGAigTl4&t=2219s

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, *70*(5, Part 2), 9-49.

Bensnes, S. S. (2016). You sneeze, you lose: The impact of pollen exposure on cognitive performance during high-stakes high school exams. *Journal of Health Economics*, *49*, 1-13.

Besluit Milieuzone Arnhem. (2014, June 25). Staatscourant van het Koninkrijk der Nederlanden. https://zoek.officielebekendmakingen.nl/stcrt-2014-18031.html#extrainformatie

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*(4), 289–328.

Calderón-Garcidueñas, L., Herrera-Soto, A., Jury, N., Maher, B. A., González-Maciel, A., Reynoso-Robles, R., ... & Varela-Nallar, L. (2020). Reduced repressive epigenetic marks, increased DNA damage and Alzheimer's disease hallmarks in the brain of humans and mice exposed to particulate urban air pollution. *Environmental Research*, *183*, 109226.

Card, D. (1999). The Causal Effect of Education on Earnings. In O.C. Fashenfelter & D. Card (Eds.), *Handbook of Labor Economics: Volume 3A* (pp. 1801-1863). Elsevier. https://doi.org/10.1016/S1573-4463(99)03011-4.

Carey, I. M., Anderson, H. R., Atkinson, R. W., Beevers, S. D., Cook, D. G., Strachan, D. P., ... & Kelly, F. J. (2018). Are noise and air pollution related to the incidence of dementia? A cohort study in London, England. *BMJ open*, *8*(9), e022404.

Carneiro, J., Cole, M. A., & Strobl, E. (2021). The Effects of Air Pollution on Students' Cognitive Performance: Evidence from Brazilian University Entrance Tests. *Journal of the Association of Environmental and Resource Economists*, *8*(6), 1051-1077.

CBS. (2017, July 4). *Helft oude diesels in Rotterdam in 2016 van de straat*. https://www.cbs.nl/nl-nl/nieuws/2017/27/helft-oude-diesels-in-rotterdam-in-2016-van-de-straat

Chen, J. C., & Schwartz, J. (2009). Neurobehavioral effects of ambient air pollution on cognitive performance in US adults. *Neurotoxicology*, *30*(2), 231-239.

Compendium voor de Leefomgeving. (2014, October 9). *Koolmonoxide in lucht, 1990–2013*. https://www.clo.nl/indicatoren/nl0465-koolmonoxide#:%7E:text=Bronnen,de%20verkeersemissies%20in%20Nederland%20gehalveerd.

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

Currie, J., & Almond, D. (2011). Human Capital Development before Age Five. In O.C. Fashenfelter & D. Card (Eds.), *Handbook of Labor Economics: Volume 4B* (pp. 1315-1486). Elsevier. https://doi.org/10.1016/S0169-7218(11)02413-0.

Doudchenko, N., & Imbens, G. W. (2016). *Balancing, regression, difference-in-differences and synthetic control methods: A synthesis* (No. w22791). National Bureau of Economic Research.

DUO. (2014–2018). *Toelichting leerlingen bo per gewicht en leeftijd* [Dataset]. DUO. https://duo.nl/open_onderwijsdata/images/toelichting-leerlingen-bo-gewicht-leeftijdpdf.pdf

EEA. (n.d.). Nitrogen oxides. In *EPER Chemicals glossary.* https://www.eea.europa.eu/help/glossary/eper-chemicals-glossary/nitrogen-oxides-nox#:~:text=The%20term%20'nitrogen%20oxides'%20(,air%20to%20form%20nitrogen%20dioxide.

Eijk, A., & Voogt, M. (2016, March). *Effectmeting milieuzone personen- en bestelverkeer in Utrecht*. TNO.

EPA. (2019, December). *Integrated Science Assessment for Particulate Matter*. https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=347534#tab-3

Fu, P., & Yung, K. K. L. (2020). Air pollution and Alzheimer's disease: a systematic review and meta-analysis. *Journal of Alzheimer's Disease*, *77*(2), 701-714.

Gemeente Amsterdam. (n.d.). *Milieuzone brom- en snorfietsen*. https://www.amsterdam.nl/veelgevraagd/?productid=%7BB9B8F66D-0366-4082-A871-D3FBB789BC25%7D#:%7E:text=Sinds%201%20januari%202018%20geldt,in%20de%20bebouwde%20kom%20rijden.

Gemeente Den Haag. (2022, January 10). *Milieuzone oude brom- en snorfietsen*. https://www.denhaag.nl/nl/in-de-stad/verkeer-en-vervoer/milieuzone-oude-brom-en-snorfietsen.htm

Gemeente Rotterdam. (2018, February). *Vervolgevaluatie Koersnota Luchtkwaliteit*. https://www.rotterdam.nl/nieuws/minder-uitstoot-verkeer/Rapport-Gezondere-Lucht-_SPREAD.pdf

Gemeenteraad stemt in met milieuzone Rotterdam. (2015, May 29). *Rijnmond*. https://www.rijnmond.nl/nieuws/129742/gemeenteraad-stemt-in-met-milieuzone-rotterdam

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254-277.

Het Parool. (2013, November 1). Oude dieselauto's niet meer welkom in Utrecht. *Het Parool*. https://www.parool.nl/nieuws/oude-dieselauto-s-niet-meer-welkom-in-utrecht~b959066a/

Inspectie van het Onderwijs. (n.d.). *Wanneer is bijstelling van het schooladvies mogelijk?* Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/onderwerpen/overgang/vraag-en-antwoord/wanneer-is-bijstelling-van-het-schooladvies-mogelijk

Jhun, I., Coull, B. A., Zanobetti, A., & Koutrakis, P. (2015). The impact of nitrogen oxides concentration decreases on ozone trends in the USA. *Air Quality, Atmosphere & Health*, *8*(3), 283-292.

Jung, C. R., Lin, Y. T., & Hwang, B. F. (2015). Ozone, particulate matter, and newly diagnosed Alzheimer's disease: a population-based cohort study in Taiwan. *Journal of Alzheimer's Disease*, *44*(2), 573-584.

Lavy, V., Ebenstein, A., & Roth, S. (2014). *The impact of short term exposure to ambient air pollution on cognitive performance and human capital formation* (No. w20648). National Bureau of Economic Research.

Milieuzones in Nederland. (n.d.-a). *Autobussen*. https://milieuzones.nl/autobussen

Milieuzones in Nederland. (n.d.-b). *Personen- en bestelauto's*. https://milieuzones.nl/personen-en-bestelautos

Milieuzones in Nederland. (n.d.-c). *Vrachtauto's*. https://milieuzones.nl/vrachtautos

Ministerie van Algemene Zaken. (n.d.). *Toegestane eindtoetsen basisschool*. Rijksoverheid.nl. https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-eindtoets-basisschool/toegestane-eindtoetsen-basisschool

Nieman, F., Motshagen, R., Griffioen-Akerboom, A., van den Brink, R. M. M., van Valsteren, R., & Colon, P. (2010, October). *Landelijke effectstudie milieuzones vrachtverkeer, effecten op de luchtkwaliteit*. Goudappel Coffeng en Buck Consultants International. https://www.milieuzones.nl/sites/default/files/Effectstudie_2010.pdf

Nilsson, J. P. (2009). The long-term effects of early childhood lead exposure: Evidence from the phase-out of leaded gasoline. *Institute for Labour Market Policy Evaluation (IFAU) Work. Pap*.

OECD. (2001). Air pollution. In *Glossary of Statistical Terms*. https://stats.oecd.org/glossary/detail.asp?ID=86#:~:text=Air%20pollution%20is%20the%20presence,produce%20other%20harmful%20environmental%20effects.

Park, S. Y., Han, J., Kim, S. H., Suk, H. W., Park, J. E., & Lee, D. Y. (2022). Impact of long-term exposure to air pollution on cognitive decline in older adults without dementia. *Journal of Alzheimer's Disease*, (Preprint), 1-11.

Peters, A., Veronesi, B., Calderón-Garcidueñas, L., Gehr, P., Chen, L. C., Geiser, M., ... & Schulz, H. (2006). Translocation and potential neurological effects of fine and ultrafine particles a critical update. *Particle and fibre toxicology*, *3*(1), 1-13.

Rambachan, A., & Roth, J. (2022). *A More Credible Approach to Parallel Trends*. Working Paper.

Reyes, J. W. (2011). Childhood lead and academic performance in Massachusetts. *Childhood*, *11*(3).

RIVM. (n.d.-a). *Bronnen per component van luchtverontreiniging*. https://www.rivm.nl/ggd-richtlijn-medische-milieukunde-luchtkwaliteit-en-gezondheid/toelichting-en-tools-luchtkwaliteit/toelichting-en-tools-luchtkwaliteit/Bronnen-per-component

RIVM. (n.d.-b). *Ozon (O3)*. Samen meten aan luchtkwaliteit. https://www.samenmetenaanluchtkwaliteit.nl/ozon-o3#:%7E:text=De%20concentratie%20voor%20ozon%20varieert,op%20de%20website%20luchtmeetnet.nl.

RIVM. (n.d.-c). *Roet (EC)*. Samen meten aan luchtkwaliteit. https://www.samenmetenaanluchtkwaliteit.nl/roet-ec#:%7E:text=Gemiddeld%20is%20de%20roetconcentratie%20in,0.5%2D1.0%20%C2%B5g%2Fm3.&text=Er%20zijn%20voor%20roet%20geen%20wettelijke%20grenswaarden%20vastgesteld.&text=In%20het%20Landelijk%20Meetnet%20Luchtkwaliteit%20wordt%20roet%20gemeten%20met%20automatische%20monitoren.

Rubio, I. (2015, April 30). Rotterdam krijgt grootste milieuzone van Nederland. *AD*. https://www.ad.nl/rotterdam/rotterdam-krijgt-grootste-milieuzone-van-nederland~a685d240/

Savelkoul, M., Schuit, A. J., & Storm, I. (2010). *Terugdringen van gezondheidsachterstanden door gemeentelijk beleid: een literatuurverkenning naar effectiviteit van fysieke en sociale omgevingsmaatregelen* (No. 270161003). RIVM. https://www.rivm.nl/bibliotheek/rapporten/270161003.pdf

Shehab, M. A., & Pope, F. D. (2019). Effects of short-term exposure to particulate matter air pollution on cognitive performance. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-44561-0.

Shi, X., & Brasseur, G. P. (2020). The response in air quality to the reduction of Chinese economic activities during the COVID-19 outbreak. *Geophysical Research Letters*, *47*(11), e2020GL088070.

Trouw. (2017, February 8). Milieuzone Utrecht mag blijven. *Trouw*. https://www.trouw.nl/duurzaamheid-natuur/milieuzone-utrecht-mag-blijven~be9b2458/?referrer=https%3A%2F%2Fwww.google.com%2F

UNEP. (2021). *Actions on Air Quality: A Global Summary of Policies and Programmes to Reduce Air Pollution*. United Nations Environment Programme (UNEP). https://www.unep.org/resources/report/actions-air-quality-global-summary-policies-and-programmes-reduce-air-pollution

Vallero, D. (2014). *Fundamentals of Air Pollution* (4th ed.). Elsevier Gezondheidszorg.

Van de Poll, T., Teeuwisse, S., Haxe, L., & Regterschot, E. (2017, June). *Onderzoek effecten en kosten uitbreiding milieuzone Arnhem*. Royal HaskoningDHV. https://repository.officiele-overheidspublicaties.nl/externebijlagen/exb-2018-66581/1/bijlage/exb-2018-66581.pdf

Van der Vegt, E. (2019a, January 1). Milieuzone voor oude dieseltjes in Arnhem van kracht. *De Gelderlander*. https://www.gelderlander.nl/arnhem/milieuzone-voor-oude-dieseltjes-in-arnhem-van-kracht~a5e26ac2/#:%7E:text=Het%20gaat%20om%20een%20uitbreiding,centrumring%20en%20in%20de%20binnenstad.

Van der Vegt, E. (2019b, March 31). Effect milieuzone Arnhem nu al zichtbaar: honderden oude diesels minder door de stad. *De Gelderlander*. https://www.gelderlander.nl/arnhem-e-o/effect-milieuzone-arnhem-nu-al-zichtbaar-honderden-oude-diesels-minder-door-de-stad~ae71b805/?referrer=https%3A%2F%2Fwww.google.com%2F

Wang, X., Younan, D., Petkus, A. J., Beavers, D. P., Espeland, M. A., Millstein, J., ... & Chen, J. C. (2021). Heterogeneous associations of air quality improvement with domain-specific cognitive function in older women. *Alzheimer's & Dementia*, *17*, e056585.

World Health Organization. (2021). *WHO global air quality guidelines*. World Health Organization.

Younan, D., Petkus, A. J., Widaman, K. F., Wang, X., Casanova, R., Espeland, M. A., ... & Chen, J. C. (2020). Particulate matter and episodic memory decline mediated by early neuroanatomic biomarkers of Alzheimer's disease. *Brain*, *143*(1), 289-302.

Younan, D., Wang, X., Millstein, J., Petkus, A. J., Beavers, D. P., Espeland, M. A., ... & Chen, J. C. (2021). Association of air quality improvement with slower decline of cognitive function in older women. *Alzheimer's & Dementia*, *17*, e056162.

Zhang, X., Chen, X., & Zhang, X. (2018). The impact of exposure to air pollution on cognitive performance. *Proceedings of the National Academy of Sciences*, *115*(37), 9193-9197.

Zumbuehl, M., & Dillingh, R. (2020, december). *Ongelijkheid van het jonge kind*. CPB. https://www.cpb.nl/sites/default/files/omnidownload/CPB-Notitie-Ongelijkheid-van-het-jonge-kind.pdf

Zweig, J. S., Ham, J. C., & Avol, E. L. (2009). Air pollution and academic performance: Evidence from California schools. *National Institute of Environmental Health Sciences*, 1-35.

Appendix A: Data summary

Table A1

Overview of environmental zones

Municipality	Environmental zone for trucks	Environmental zone for cars
Alkmaar	-	-
Almelo	-	-
Almere	-	-
Alphen aan den Rijn	-	-
Amersfoort	-	-
Apeldoorn	-	-
Arnhem	July 1, 2014	January 1, 2019
Assen	-	-
Breda	July 1, 2007	-
Delft	January 1, 2010	-
Den Bosch	September 1, 2007	-
Den Haag	April 16, 2008	-
Deventer	-	-
Dordrecht	-	-
Ede	-	-
Eindhoven	July 1, 2007	-
Emmen	-	-
Enschede	-	-
Gouda	-	-
Groningen	-	-
Haarlem	January 1, 2022	-
Haarlemmermeer	-	-
Heerlen	-	-
Helmond	-	-
Hengelo	-	-
Hilversum	-	-
Hoorn	-	-
Leeuwarden	-	-
Leiden	January 1, 2010	-

Lelystad	-	-
Maastricht	March 1, 2010	-
Nijmegen	-	-
Oss	-	-
Rijswijk	September 10, 2010	-
Roosendaal	-	-
Rotterdam	September 16, 2007	January 1, 2016
Schiedam	-	-
Sittard-Geleen	-	-
Tilburg	September 1, 2007	-
Utrecht	July 1, 2007	January 1, 2015
Venlo	-	-
Zaanstad	-	-
Zoetermeer	-	-
Zwolle	-	-

Note: Overview of G40 cities and other cities that introduced environmental zones. Amsterdam and Maasvlakte Rotterdam are not included. Table contains date of introduction for each type of environmental zone per city.

Table A2

Data summary for air pollution

Panel A: Stations in Utrecht					
Variable	Observations	Mean	Standard deviation	Minimum	Maximum
PM_{10}	127	19.550	4.510	12.64	32.19
$PM_{2.5}$	125	11.171	4.692	1.96	22.8
NO_2	205	26.614	8.035	10.33	61.14
NO	209	10.945	7.912	1.04	43.32
O_3	137	41.111	16.332	9.19	78.15
Soot	-	-	-	-	-
Panel B: Stations in Arnhem					
Variable	Observations	Mean	Standard deviation	Minimum	Maximum
PM_{10}	92	22.052	6.364	5.19	39.29
$PM_{2.5}$	-	-	-	-	-
NO_2	57	33.318	8.014	20.72	52
NO	57	24.936	10.684	9.97	53.23
O_3	-	-	-	-	-

Soot	50	1.452	0.385	0.8	2.47
		Panel C: Stations in Rotterdam			
Variable	Observations	Mean	Standard deviation	Minimum	Maximum
PM_{10}	375	21.801	5.097	9.76	40.84
$PM_{2.5}$	290	12.872	4.650	4.68	27.04
NO_2	419	33.665	8.928	15.27	58.07
NO	420	15.884	12.360	0.88	67.43
O_3	205	39.065	14.653	7.29	74.56
Soot	246	1.534	0.623	0.5	3.58
		Panel D: Stations in cities with only a truck zone			
Variable	Observations	Mean	Standard deviation	Minimum	Maximum
PM_{10}	870	20.763	4.588	8.76	36.08
$PM_{2.5}$	609	12.776	4.682	2.62	28.24
NO_2	1,203	30.068	10.420	10.81	63.05
NO	1,208	14.886	14.416	0.1	88.1
O_3	575	43.736	15.418	9.86	88.78
Soot	602	1.328	0.591	0.4	3.92

Note: Summary statistics for air pollution. One observation corresponds to one station in one year. Values are given in μg/m^3. Each panel corresponds to a separate group of observations.

Table A3

Data summary for test scores

		Panel A: Schools in Utrecht			
Variable	Observations	Mean	Standard deviation	Minimum	Maximum
CITO score	685	534.955	4.737	518.33	547.87
Route 8 score	22	205.942	12.878	179	225.43
AMN score	1	393.42	-	393.42	393.42
Dia score	5	360.376	7.474	350.38	370.21
IEP score	70	80.613	5.031	65.8	89.92
Standardized score	780	-0.031	1.080	-4.097	3.197
Share migrants	791	0.291	0.282	0	0.997

Share with weight 1.2	791	0.127	0.169	0	0.734
School weight	791	30.963	57.392	0	361
Panel B: Schools in Arnhem					
Variable	Observations	Mean	Standard deviation	Minimum	Maximum
CITO score	387	534.178	4.157	518.93	543.61
Route 8 score	3	211.79	13.499	197.62	224.5
AMN score	1	412.91	-	412.91	412.91
Dia score	-	-	-	-	-
IEP score	4	82.263	5.482	75.21	88.19
Standardized score	395	-0.216	0.967	-3.843	2.194
Share migrants	398	0.223	0.239	0	0.937
Share with weight 1.2	398	0.072	0.105	0	0.515
School weight	398	15.073	35.105	0	228
Panel C: Schools in Rotterdam					
Variable	Observations	Mean	Standard deviation	Minimum	Maximum
CITO score	1,200	532.718	4.990	513.2	546.15
Route 8 score	5	208.808	14.961	186.64	227.16
AMN score	2	391.67	14.326	381.54	401.8
Dia score	6	356.038	3.555	349.31	358.69
IEP score	183	77.400	5.200	63.61	89.75
Standardized score	1,391	-0.521	1.114	-5.504	2.790
Share migrants	1,425	0.394	0.266	0	1
Share with weight 1.2	1,425	0.167	0.162	0	0.772
School weight	1,425	44.701	57.021	0	396
Panel D: Schools in cities with only a truck zone					
Variable	Observations	Mean	Standard deviation	Minimum	Maximum
CITO score	4,387	534.568	4.774	515.1	546.92

Route 8 score	241	202.664	10.397	171.92	226.93
AMN score	14	410.315	26.211	358	449.7
Dia score	29	359.037	4.555	348.25	366.63
IEP score	434	79.975	4.995	61.83	93.47
Standardized score	5,090	-0.128	1.084	-5.130	2.963
Share migrants	5,300	0.309	0.279	0	1
Share with weight 1.2	5,300	0.116	0.143	0	0.772
School weight	5,300	29.613	47.952	0	497
Panel E: Schools in cities with no environmental zone					
Variable	Observations	Mean	Standard deviation	Minimum	Maximum
CITO score	7,145	534.589	3.994	515.4	546.78
Route 8 score	573	203.851	10.384	170.5	235.88
AMN score	26	421.771	39.695	355.25	572.29
Dia score	58	359.5878	3.509	352.15	366.42
IEP score	1,315	80.122	5.031	57.52	92.55
Standardized score	9,059	-0.114	0.936	-4.865	3.061
Share migrants	9,780	0.153	0.198	0	1
Share with weight 1.2	9,780	0.059	0.095	0	0.924
School weight	9,780	10.666	24.557	0	279

Note: Summary statistics for test scores. One observation corresponds to one school in one year. Each panel corresponds to a separate group of observations.

Appendix B: SDID methodology

The equations in this Appendix are all provided by Arkhangelsky et al. (2021). The unit weights are chosen as follows:

$$(\widehat{\omega}_0, \widehat{\omega}^{sdid}) = \underset{\omega_0 \in \mathbb{R}, \omega \in \Omega}{\operatorname{argmin}} \sum_{t=1}^{T_{pre}} \left(\sum_{i=1}^{N_{co}} \omega_i Y_{it} - \frac{1}{N_{tr}} \sum_{i=N_{co}+1}^{N} Y_{it} + \omega_0 \right)^2 + \zeta^2 T_{pre} \|\omega\|_2^2, \text{ with}$$

$\Omega = \left\{ \omega \in \mathbb{R}_+^N : \sum_{i=1}^{N_{co}} \omega_i = 1, \omega_i = N_{tr}^{-1} \; for \; all \; i = N_{co}+1, \dots, N \right\}$, where $\mathbb{R}_+$ is the positive real line.

The regularization parameter ζ is set as follows: $\zeta = (N_{tr} T_{post})^{1/4} \hat{\sigma}$, with

$\hat{\sigma}^2 = \frac{1}{N_{co}(T_{pre}-1)} \sum_{i=1}^{N_{co}} \sum_{t=1}^{T_{pre}-1} (\Delta_{it} - \overline{\Delta})^2$, $\Delta_{it} = Y_{i(t+1)} - Y_{it}$, and $\overline{\Delta} = \frac{1}{N_{co}(T_{pre}-1)} \sum_{i=1}^{N_{co}} \sum_{t=1}^{T_{pre}-1} \Delta_{it}$.

The time weights are chosen as follows:

$$(\hat{\lambda}_0, \hat{\lambda}^{sdid}) = \underset{\lambda_0 \in \mathbb{R}, \lambda \in \Lambda}{\operatorname{argmin}} \sum_{i=1}^{N_{co}} \left(\sum_{t=1}^{T_{pre}} \lambda_t Y_{it} - \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^{T} Y_{it} + \lambda_0 \right)^2, \text{ with}$$

$$\Lambda = \left\{ \lambda \in \mathbb{R}_+^T : \sum_{t=1}^{T_{pre}} \lambda_t = 1, \lambda_t = T_{post}^{-1} \; for \; all \; t = T_{pre}+1, \dots, T \right\}.$$

Appendix C: Robustness checks

Table C1

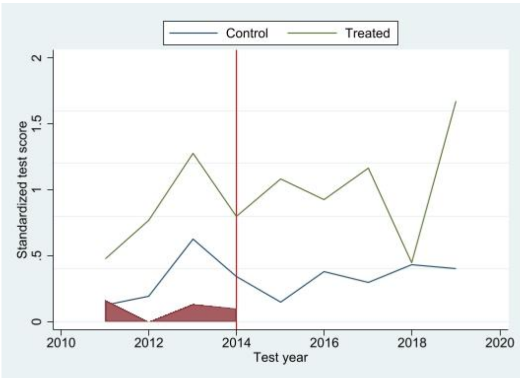
Effects of environmental zones on standardized test scores for schools located inside environmental zone

	(1) Car zone in Utrecht	(2) Truck zone in Arnhem	(3) Car zone in Arnhem	(4) Car zone in Rotterdam
Environmental zone	0.244 (0.268)	-0.780 (0.615)	0.082 (0.780)	0.064 (0.175)

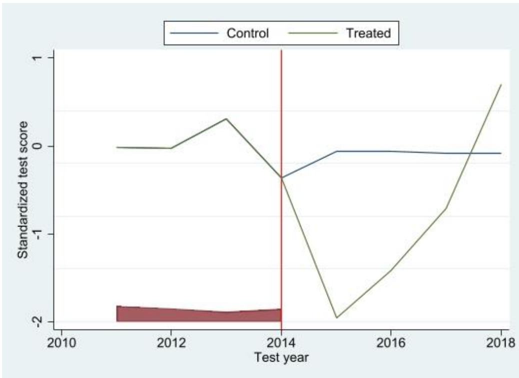
Note: Dependent variable is average standardized test score on final exam. Each column represents a single regression to estimate the effect of one environmental zone in a city. Standard errors are reported between brackets and calculated using 200 placebo replications, with the exception of column 4 which uses bootstrapping. Stars indicate p-values, with the following values: *p<0.1 **p<0.05 ***p<0.01.

Figure C1: Synthetic difference-in-difference estimates of environmental zones

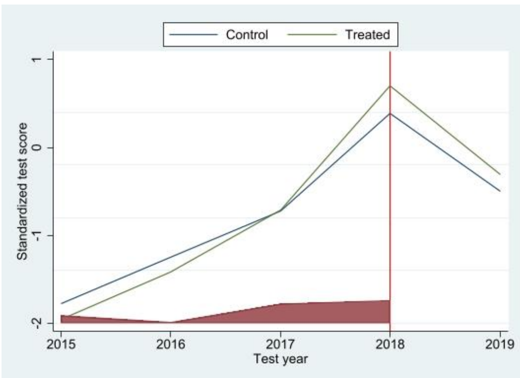
Panel A: Effect on environmental zone for cars in Utrecht on standardized test scores



Panel B: Effect on environmental zone for truck in Arnhem on standardized test scores



Panel C: Effect on environmental zone for cars in Arnhem on standardized test scores



Panel D: Effect on environmental zone for cars in Rotterdam on standardized test scores

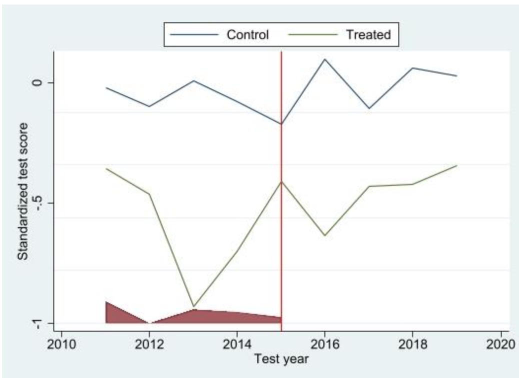


Table C2

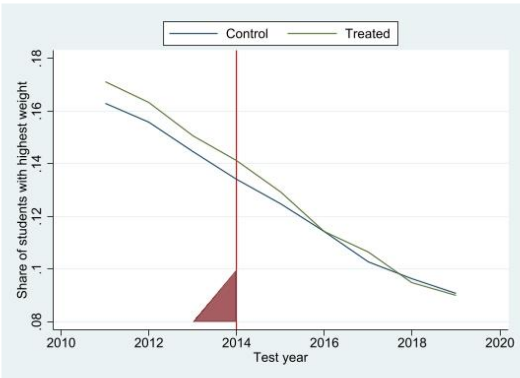
Effects of environmental zones on the share of students with the highest weight

	(1) Car zone in Utrecht	(2) Truck zone in Arnhem	(3) Car zone in Arnhem	(4) Car zone in Rotterdam
Environmental zone	-0.006 (0.005)	-0.002 (0.006)	-0.002 (0.002)	-0.007* (0.004)

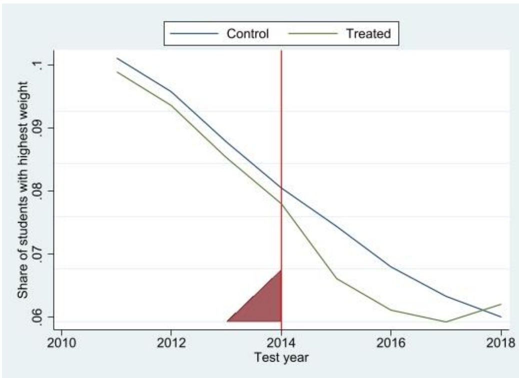
Note: Dependent variable is share of students with a weight of 1.2. Each column represents a single regression to estimate the effect of one environmental zone in a city. Standard errors are reported between brackets and calculated using 200 placebo replications. Stars indicate p-values, with the following values: *p<0.1 **p<0.05 ***p<0.01.

Figure C2: Synthetic difference-in-difference estimates of environmental zones

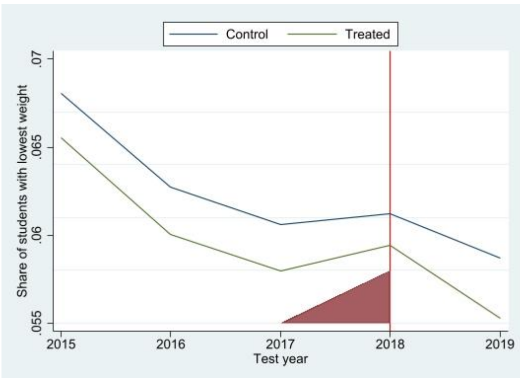
Panel A: Effect of environmental zone for cars in Utrecht on share of students with highest weight



Panel B: Effect of environmental zone for truck in Arnhem on share of students with highest weight



Panel C: Effect on environmental zone for cars in Arnhem on share of students with highest weight



Panel D: Effect on environmental zone for cars in Rotterdam on share of students with highest weight