Natural Experiments in Environmental and Transport Economics

Klingen, Joris Johannes

2021

This thesis provides a collection of five natural experiments in environmental and transport economics. Chapter 1 introduces the topics and offers the methodological context. Chapter 2 tests the hypothesis that particulate matter has a direct effect on human decision-making. It uses chess games as a natural experiment and demonstrates that air pollution causes individuals to take less risk. Chapter 3 assesses whether ozone air pollution affects human physical activity. Findings show that ozone reduces cycling speed, even for concentrations below current air quality standards. Chapter 4 finds that public rental bicycles are a local net substitute for metro service and that these bicycles can alleviate time losses stemming from interruptions in public transport. Chapter 5 focuses on New York City and estimates the causal effect of protected bike lanes on traffic speed, flow, and road safety. Bike lanes seem to improve cyclists' safety both on streets and at junctions, while having no statistically significant effect on traffic speed and traffic flow. Chapter 6 investigates to what extent smartphones play a role in the number of road accidents. The results indicate that smartphone distraction can explain 10% of accidents and that phone-related accidents mainly happen on local urban roads.

Joris Klingen (1989) completed Future Planet Studies at the University of Amsterdam in 2012, Spatial Transport and Environmental Economics at the Vrije Universiteit Amsterdam in 2014, and Economics at the Tinbergen Institute in 2017. Furthermore, he worked as a junior lecturer at the Institute of Interdisciplinary Studies of the University of Amsterdam, and as PhD researcher at Vrije Universiteit. He currently works as a data scientist for the City of Amsterdam.

Natural Experiments in Environmental and Transport Economics

Joris Klingen

778

tinbergen institute

Joris Klingen

Natural Experiments in Environmental and Transport Economics

Vrije Universiteit Amsterdam

NATURAL EXPERIMENTS IN ENVIRONMENTAL AND TRANSPORT ECONOMICS

ISBN 978 90 361 0645 0

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **778** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back

VRIJE UNIVERSITEIT

NATURAL EXPERIMENTS IN ENVIRONMENTAL AND TRANSPORT ECONOMICS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de School of Business and Economics
op donderdag 25 maart 2021 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Joris Johannes Klingen

geboren te Doorn

promotoren:	prof.dr. J.N. van Ommeren
	prof.dr. E.T. Verhoef

Contents

1 Introduction

Sometimes, a determined researcher wants to test a hypothesis for which no accepted theory exists. This was the case in 1848 in London when physician John Snow hypothesised that cholera was transmitted through contaminated water. Despite his conviction and the urgency to reduce cholera cases, John Snow could not test his hypothesis for several years. First of all, his conjecture was not supported by any major theory. At the time, the dominant thought was that diseases transmit through inhaling bad air. Also, running an actual experiment was considered unethical. It would require him to define a control group with participants drinking freshwater, and, more problematically, a treated group that would have to drink contaminated water.

After six years, in 1854, John Snow found a way to test his hypothesis, using the local cholera outbreak in the neighbourhood of Soho. He realised that this outbreak was a large scale *natural experiment* and he wrote:

> No fewer than three hundred thousand people of both sexes, of every age and occupation, and of every rank and station, from gentlefolks down to the very poor, were divided into groups without their choice, and in most cases, without their knowledge; one group being supplied with water containing the sewage of London, and amongst it, whatever might have come from the cholera patients; the other group having water quite free from such impurity (Snow, 1855, p. 75).

The cholera outbreak in Soho allowed John Snow to compare the naturally emerged treated and control groups. Thereby, he could show that it was indeed contaminated water that transmitted the disease. This finding was not only a breakthrough in the science of epidemiology, but it is also seen as the birth of natural experiments as a scientific method to identify causal effects from observational data (Freedman, 1999).

Researchers now widely use natural experiments for causal inference. The increasing availability of data, advances in statistical knowledge, and ever-increasing computer power, make analysing natural experiments an effective tool in case a classic experiment is unethical or simply too expensive. In addition, because the size of a laboratory does not bound the scale of natural experiments, they also allow researchers to study more subtle effects, that require large statistical power, i.e. many observations.

This thesis is a collection of studies in the spirit of John Snow. Not because their research questions are as urgent as understanding cholera transmission, but because they too rely on natural experiments with large groups of 'participants'. The experiments analysed in this thesis are *natural* because the researcher did not induce the identifying variation, nor assigned participants randomly over the treatments. Instead, the variations come from: weather conditions (chapters 2, 3), technical failures (chapter 4), or a change in policies (5 and 6). These settings are then applied to understand the immediate effects of air pollution, road safety, road congestion, and substitution between metro and public rental bicycles.

Chapter 2 tests the hypothesis that particulate matter has a direct effect on human decision making. Medical research suggests that particulate matter (PM) increases stress hormones, therefore increasing the feeling of stress, which has been hypothesised to induce individuals to take less risk. As a natural experiment, this study focuses on whether PM increases the probability of drawing in chess games using information from the Dutch club competition. The ideal experimental setup to estimate the causal effect of pollution on the probability to make a draw is to examine the outcome of games of players that are randomly assigned to play against each other at different locations with different levels of pollution. The Dutch chess competition comes close to that setup, because teams belonging to the same league play at different locations on the same day, and all teams play each other during a season. The results provide evidence of a reasonably strong effect: A 10μg increase in PM_{10} (33.6% of mean concentration) leads to a 5.8% increase in draws. This chapter, therefore, demonstrates that air pollution causes individuals to take less risk.

Chapter 3 estimates how air pollution in general, and ambient ozone in particular, affects human physical activity through impaired lung functioning. The study provides novel evidence of the immediate impact of air pollution on time delays in urban outdoor activities. This effect is estimated on cycling speeds in London using several estimation strategies. The results show that ozone reduces speed for concentrations above 20 ppb, which is far below the minimum threshold suggested by other studies. A 10 ppb increase in ozone concentration leads to a 0.3-0.4% reduction in cycling speed, despite that most cycling trips are short so that exposure to ozone tends to be short. It seems plausible that ozone induces time losses of similar magnitude of other outdoor activities, such as walking.

Chapter 4 studies how public transport and cycling are related in a dense urban network. Specifically, it focuses on how demand for public rental bicycles is affected by local and temporary metro interruptions in Paris. A unique dataset is constructed by linking metro interruptions announced on Twitter by the Parisian metro operator to usage data on the Vélib' public rental bicycles. The results show that, as a direct consequence of a metro interruption, the consumption of bicycles within 100 metres of metro stations increases by 0.72 bicycles per hour per docking station on average, and with 1.54 bicycles per hour per docking station during the first 20 minutes; an increase of approximately 11% and 22% respectively. Due to their effects on demand, metro interruptions increase the probability of empty stocks at docking stations with 15%. The findings highlight that cycling is a local net substitute for metro service, and that public rental bicycles can alleviate time losses stemming from interruptions in public transport.

Chapter 5 focuses on Manhattan in New York City and estimates the causal effect of bike lanes on congestion and road safety. Because a street-level analysis is prone to biases due to treatment-induced rerouting, aggregated observations of streets in the same direction within narrowly defined areas on Manhattan are used. Thereby, Manhattan's elongated shape and grid-structured street network are exploited to assure causality, but also to obtain policy-relevant area-level estimates. Bike lanes are found to improve safety for cyclists both on streets and at junctions. Once an area can be completely traversed on a protected bike lane, accidents with cyclists involved are estimated to be reduced by 34%. The results further indicate that bike lanes have no statistically significant effect on overall road safety at junctions, but reduce accidents away from junctions by 59% for *all modes* in the whole area. Using taxi trips as an accurate proxy for traffic indicators, bike lanes appear to have no statistically significant effect on traffic speed of traffic flow at the area-direction level. However, traffic speed on streets with bike lanes is 1.3% lower compared to streets in the same direction in the same area. At the same time, the results indicate that streets with a bike lane accommodate a 2.3% higher throughput.

Chapter 6 investigates to what extent smartphones play a role in the number of road accidents. The study exploits variation in phone usage fees in the Netherlands following the European Union (EU) roaming regulations in 2017, which abolished all roaming surcharges for EU residents. This change is used to estimate a difference-in-differences model where non-Dutch drivers from the EU are treated, while Dutch drivers serve as a control group. Phone use patterns show that the growth rate of mobile calls, texts, and particularly data usage increased substantially after the change in roaming regulations, making roaming phone use more in line with usage in home countries. While actual phone use of drivers is not directly observed, the overall phone use is likely to (partly) carry over to phone use while driving. The results then suggest that 10% of road accidents can be explained by smartphone distraction. Under plausible assumptions, the preferred estimate implies a crash risk odds ratio of around 3.8, which indicates that that within-vehicle smartphone use makes an accident almost four times more likely to occur. The findings further indicate that phone distraction increases accidents of *all* severity levels by a similar magnitude, and that phone-related accidents mainly happen on local urban roads.

Chapter 7 concludes the thesis with a summary of the studies and a brief discussion on why natural experiments are particularly useful in answering the research questions in this thesis.

2

Risk-taking and air pollution: Evidence from chess

2.1 Introduction

Particulate matter (PM) is found to increase stress hormones and blood pressure, therefore increasing the feeling of stress, which has been hypothesised to induce individuals to take less risk (Duflo and Banerjee, 2011). To examine this, we study the effect of PM on risk outcomes of the game of chess, i.e. the probability to draw.

The literature mainly focuses on long-term health effects of PM and other air pollution (see e.g. Chappie and Lave, 1982 and Beach and Hanlon, 2018). In contrast, we focus on the immediate effect of air pollution, for which there is growing attention (Graff Zivin and Neidell, 2018). We now know that pollution also has an *immediate* detrimental effect on physical health and therefore on economic and social activities

This chapter is based on Klingen and van Ommeren (2020a). The authors like to thank Koos Stolk of the Royal Dutch Chess Federation for help with the data. Moreover, we would like to thank Hans Koster, Erik Verhoef, Francis Ostermeijer, Devi Brands, Jesper de Groote, and seminar participants at University of Birmingham and VU Amsterdam.

which depend on physical health (e.g. labour productivity, cycling to work).[1]

It is less well-known that the immediate effects of PM are more subtle and widespread. PM affects cognitive ability, and therefore reasoned judgement and decision-making (see e.g. Hamanaka and Mutlu, 2018). Medical studies show that PM increases stress hormones (such as cortisol) as well as blood pressure (Li et al., 2017; Barbosa et al., 2012). In other contexts, PM negatively affects important activities which require cognitive performance, including educational achievement (Ebenstein et al., 2016), high skill work (Kahn and Li, 2019) and investment decisions (Huang et al., 2017). Traders on Wall Street have lower returns on days with higher PM concentrations (Heyes et al., 2016), while baseball referees underperform given higher levels of PM (Archsmith et al., 2018). Finally, Künn et al. (2019) find that chess players make more meaningful errors due to PM, especially when under time pressure.

Several recent studies show individuals' decision-making effects of PM that point at the possibility that PM reduces risk-taking (Lu, 2019). For example, Heyes et al. (2016) argue that one possible interpretation of their findings for lower returns for Wall Street traders is that PM induces these traders to take less risk.[2] This is in line with papers on PM and crime that suggest that anxiety increases with PM. (Herrnstadt et al., 2016; Burkhardt et al., 2019). There is also evidence that daily higher PM levels increase the probability of buying health insurance (Chang et al., 2018), and reduce the sales of lottery tickets (Bondy et al., 2019).[3]

These studies estimate PM effects that are likely the result of several behavioural factors (notably skills, discounting, and risk-taking). It is still unclear which behavioural mechanisms underlie previous findings. More specifically, we do not know whether PM directly affects risk attitude. In contrast to existing studies, we study the effect of PM on an indicator of risk-taking, using the game of chess. Thereby, we can provide field evidence on the often stated hypothesis that PM air pollution induces individuals to take less risk, and thereby reduces the expected pay-offs of the strongest

[1] Zivin and Neidell (2012) show that agricultural workers are less productive on days with high ozone levels. Lichter et al. (2017) identify a small effect of PM on some productivity indicators of professional players in football. Chang et al. (2016), Chang et al. (2019a) study productivity of pear packers and call centre employees and find adverse effects of PM pollution on productivity. Klingen and van Ommeren (2020b) show that ozone reduces cycling speed.

[2] An alternative explanation is that these traders lose or adapt their discount rate. Furthermore, Chew et al. (2019) also suggest that PM makes individuals more risk averse. PM increases car accidents (Sager, 2019) as well as crime (Bondy et al., 2019), but these studies show that these results are unlikely due to higher levels of risk-taking.

[3] The benefits of health insurance and lottery gains are in the future, so an alternative explanation is that PM affects the discount rate. Projections bias may also play a role. Projection bias is the tendency for individuals to exaggerate the degree to which their future tastes will resemble current tastes, which is likely affected by pollution.

player.

Risk-taking is essential to the game of chess. The main advantage of focusing on chess is that it offers a direct measure of risk: the variance of game outcomes reflects risk-taking, as many games end in a draw. Furthermore, the time horizon of a chess game is short (a few hours), so estimates are not affected by the effects of PM on discount rates or projection bias (Heyes et al., 2016; Bondy et al., 2019). Consequently, the effect of PM on risk-taking can be examined by analysing its effect on the probability of making a draw.[4]

The ideal experimental setup to estimate the causal effect of pollution on the probability to make a draw is to examine the outcome of games of players that are randomly assigned to play against each other at different locations with different levels of pollution. We come close to that set up by analysing games played in the Dutch team club competitions, where teams belonging to the same league play at different locations and all teams play each other (as is common in most national team sports competitions).[5] Ideally, one would also analyse the exact chess moves for each game. We only observe moves for a subsample from the highest league. However, due to the subtlety of the effect, the full sample is required to obtain sufficient statistical power.

All games take place at the same time (on Saturdays at 1 pm), and are scheduled in advance, so that our results are *not* driven by any extensive margin decisions, as would for example be the case for online chess games. Because pollution levels do not vary randomly over time and space, we control for time-specific as well as location-specific unobserved factors using a fixed-effects strategy.[6]

In our estimation approach, we pay special attention to measurement error in PM due to the distance between the pollution monitor and the chess location. Measurement error in pollution levels usually causes attenuation bias. One way to deal with this is to use instrumental variables, which is the preferred strategy in the literature. In particular, it is common to use temperature inversion as an instrument (Jans et al., 2018). Although this strategy is attractive, there are also disadvantages with its use.[7] We follow a different route. We focus on chess locations close to PM monitoring

[4]We will discuss alternative explanations for finding an increased number of draws due to PM, such as the length of a game or reduced cognitive performance.

[5]For that reason, we concentrate on the Dutch national competition, but ignore information from other countries (e.g. Germany, UK), where players tend to play at the same location, hence there is little or no spatial variation in those contexts.

[6]The probability of a draw depends on the strength of the players. Therefore, we can improve the efficiency of our estimates by controlling for Elo rating, which very accurately measures player strength at the time of playing (Regan and Haworth, 2011).

[7]It is plausible that the instrument affects a range of pollutions, and not only PM, so it is difficult to interpret the IV estimate as a causal estimate of PM. In addition, confidence intervals of the IV estimates are much larger (and tend not to differ from OLS estimates using Hausman tests).

stations. Furthermore, we will show how the PM effects decrease with distance to the pollution monitor.

There may be alternative explanations of our finding of increased draws due to PM. Most notably, one may think that because PM negatively affects the cognitive performance of chess players, this would increase the probability of making a draw. We show that this alternative hypothesis does not hold by demonstrating that weaker players make fewer draws, which implies that reduced cognitive performance can only induce a negligible downward bias. Furthermore, as shown by Künn et al. (2019), PM seems to only increase meaningful errors (i.e. blunders), which make draws even *less* likely. Another possible explanation is increased fatigue due to PM, which may induce players to offer or accept draws earlier in the game. We cannot completely rule out this explanation, but we note that players can also end games earlier by resignation, which is very common in chess. Therefore, a preference for shorter games does not per se imply more draws, as it is plausible that both resignations and draw-offerings increase. Overall, our preferred interpretation for finding more draws is, therefore, a reduction in risk-taking.

In conclusion, we will provide evidence that PM_{10} reduces risk-taking.[8] A 10μg increase in PM_{10} (33.6% of mean concentration) leads to a 5.8% increase in draws. We do not find any effect of PM when measured at the location of the visiting club, or of effects of PM on previous days, which implies that the PM effect is immediate. This finding supports and complements other studies that show the effect of pollution on decision taking, but which offer several explanations to explain their findings or projection bias (Heyes et al., 2016; Bondy et al., 2019).

This paper proceeds as follows. Section 2.2 explains the methods employed. Section 2.3 describes the data and descriptive statistics. Section 2.4 presents results. Section 2.5 concludes.

[8]We use a daily measure of PM_{10} rather than a measure of $PM_{2.5}$ observed during the game, which may be a slightly better measure from a theoretical point of view. However, $PM_{2.5}$ is roughly 1/30 of the diameter of a human hair, it may go through walls, and as result outside and inside concentration levels are usually the same. We observe the latter only for a short period. This is not problematic, as the attenuation bias of using PM_{10} rather than $PM_{2.5}$ is small, as the correlation between those two measures is 0.90.

2.2 Empirical method

2.2.1 Identification

Chess is a zero-sum perfect information game between two players, with two possible outcomes: either one of the players wins, or there is a draw. In chess, the players' moves are strongly related to level of risk they take (these determinants are discussed later on). For example, players choose between safe and risky openings, which affects the probability of a draw. Players can also choose risky moves, i.e. moves that reduce the probability of a draw.[9]

In the (financial and economics) literature on risk-taking, a common measure of risk-taking outcomes is the standard deviation (e.g. the standard deviation of the return of a portfolio), and therefore the outcome variance. The variance of chess outcomes is a one-to-one linear negative function of the proportion of draws.[10] Let D_{ict} be a dummy indicator of whether a game i in location c on day t ends in a draw. The level of PM in location c on day t is denoted by PM_{ct}. We aim to estimate the causal effect of PM_{ct} on D_{ict}. Because draws are common (32% in our sample) we use a linear probability model. [11]

The first main econometric issue when aiming to estimate a causal effect of PM on the probability of a draw, is that PM does not randomly vary over time, but there are strong time trends in levels of PM (as air pollution tends to decrease over this time). Furthermore, PM is not randomly allocated across space but is concentrated in certain cities. It is also possible that certain cities attract players with different propensities of making a draw.

The ideal way to address these issues is to compare outcome of games of players that are randomly allocated to other players at different locations for different time periods. By using the universe of chess games of a national competition for longer periods, combined with a day and location fixed effect regression design, we approach this ideal setup. In the national competition, players play half of all games at their home location and the other half at another location. Hence, in essence, we use variation in PM at different locations within the same day. For reasons of efficiency, we include two game-specific control variables: the difference in ELO rating between the

[9] Players are often categorised as those with a high risk attitude (e.g., the 1985-2000 world champion Kasparov) or with a less risky attitude (e.g., the 1963-1969 world champion Petrosian).

[10] The outcome variance is equal to (1 - proportion of draws)/4.

[11] The estimates results hardly change when estimating using similar specifications with a logit model.

players, as well as the average rating of the players.[12] We will control for weather conditions that potentially confound the effect of PM, as studies such as Wang (2017) and Heyes and Saberian (2019) show that temperature has an effect on decision making.

Consequently, we will estimate the following fixed-effects regression:

$$D_{ict} = \alpha_c + \alpha_t + \beta \cdot PM_{ct} + \gamma \cdot \mathbf{X}_{ict} + \delta \cdot z_{ct} + \varepsilon_{ict}, \tag{2.1}$$

where α_c and α_t denote location and day fixed effects. Here, $\mathbf{X}_{ict}$ denotes control variables that capture players' strength and z_{ct} denotes (time-varying) location-specific control variables, such as weather conditions.

We have not yet been specific about the type of location fixed effects used. To elaborate on this, we use three types of location fixed effects. We use one fixed effect for the club of the home player, one fixed effect for the club of the visitor player, and one fixed effect for the PM monitor. In the analysis, we will cluster the standard errors at the level of the PM monitor as well as day t.

One of the strengths of the design is that we will see that inclusion of the location fixed effects as well as the weather control variables does not affect the results, which makes it more plausible that the variation in PM is exogenous.[13] This also makes sense, as the Netherlands is a geographically small country, hence the distance between these locations is small (the average distance is only 70km, where we weigh by number of games per location).[14] Hence, by including day fixed effects, we already almost perfectly control for differences in *weather* conditions (e.g. differences in temperature and sunshine are negligible).

The second econometric issue is that PM reduces the ability of players to play well, resulting in more mistakes (Ebenstein et al., 2016; Künn et al., 2019). This does *not* imply that this will induce more draws. An important feature of chess, for which we will provide ample evidence, is that the probability of making a draw is a *non-decreasing* function of players' ability level. Even better, we show that the probability to make a draw does *not* depend on the level of the player, except for very strong

[12]It is straightforward to show that this specification implies that we control for the rating of the strongest player and the rating of the weakest player, where we allow the effects of these variables to differ.

[13]Because our results also hold without location fixed effects, our estimates hold with one-way fixed effect models, we do not have the issue that two-way fixed effects models have difficulties addressing heterogeneity of estimates, resulting in inconsistent estimates (de Chaisemartin and d'Haultfoeuille, 2020).

[14]For example, the distance between Amsterdam and Rotterdam, the two largest cities of the Netherlands, is only 65 km, whereas a number of cities, such as the Hague, Delft and Leiden, are located in between.

players (who are rare in our dataset) who make more draws.[15] Hence, the effect of PM on the probability of draws through its effect on ability is negligible. Furthermore, we will demonstrate that even if we assume that PM has (unreasonably) large effects on the ability of both players (i.e. an unreasonable large drop in their Elo ratings), then this assumption cannot explain our findings.

The third econometric issue is measurement error, as the spatial density of PM monitor stations is usually rather low, which causes attenuation bias. To avoid this, we use to our advantage that in the Netherlands many chess clubs, and in particular large clubs with many players, are located in larger cities which have several monitoring stations. Subsequently, we will focus on chess games within a maximum distance of 5 km of a monitoring station, so the average distance between the chess location and the monitoring station is small and slightly more than 2 km. In our sensitivity analysis, we will show that increasing the maximum distance indeed results in lower, but still statistically significant, estimates, whereas reducing the maximum distance results in somewhat higher point estimates but larger confidence intervals.[16]

The fourth econometric issue is whether the effect of PM is dynamic. The medical literature does not answer the question whether the effect of PM is immediate or also with a delay.[17] The latter is theoretically possible, because PM remains in the blood circulation. For that reason, we will also measure PM on the day before the match, as well as at the location of the visiting club. The idea of the latter PM measure is that chess players typically live close to their club, and hence visitor players might be treated in the night or morning before travelling to the game. As a placebo check, we will additionally examine the effects of PM measured on the day after the game.

2.3 Data

We observe the universe of outcomes of (classical) chess games for the Dutch national team competition from 2002 until 2018, played at locations as shown on the map in Figure 2.1.[18] Each year, there are about 15 different leagues, in which between

[15]This makes sense. Stronger players are better able to calculate the consequences of their moves, and therefore have more control over the game outcome.

[16]It is not an issue that we do not measure PM inside buildings, as environmental policies use information from outside monitoring stations, so the preferred measures, from a policy point of view, is the measure used by us.

[17]Our main source of effect of PM in this literature is Li et al. (2017). In this study, participants are treated with PM for a number of days, but dynamic treatment effects are not investigated.

[18]The Dutch league follows the rules of the World Chess Federation FIDE: players receive 90 minutes for the first 40 moves, and an additional 30 minutes for the rest of the game. For each move played, the player receives an additional 30 seconds. A player who exceeds the time limit loses the game.

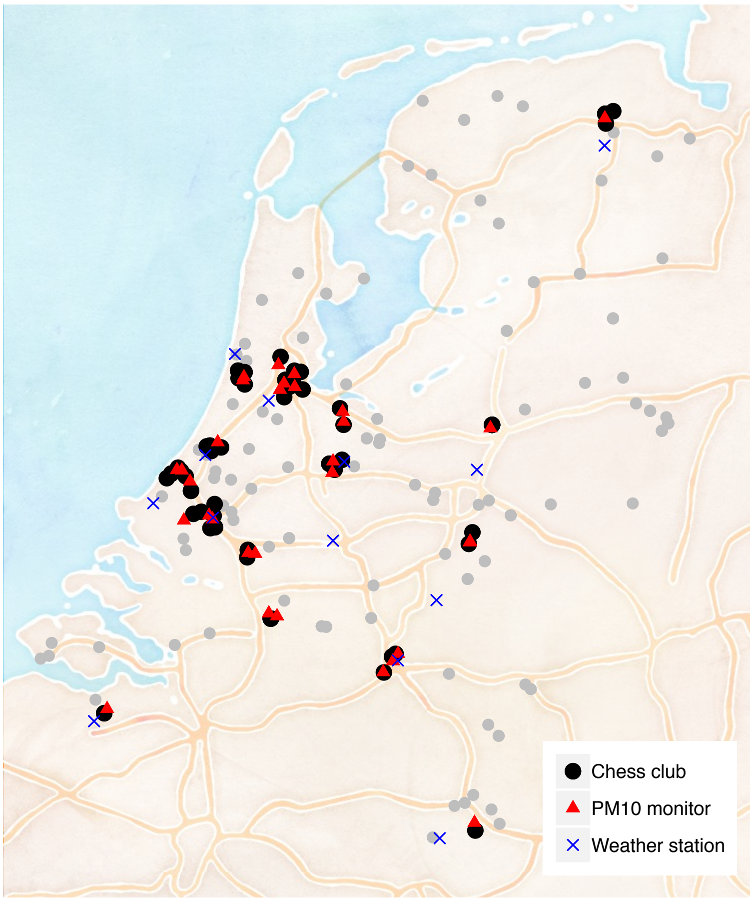


Figure 2.1: Locations of chess clubs, weather stations and PM monitoring stations.

Notes: Greyed out chess locations are excluded because of a too large distance to PM or weather monitoring stations. Some of these locations are used for sensitivity analyses.

8 and 10 teams compete. For games played in the highest league, we also observe the moves of the full game. Teams have 8 or 10 players and play either at home or away (similar to, for example, soccer). Although chess players play for a team, individual chess outcomes are relevant for players, as the outcome influences their Elo-rating. A competition year contains 9 rounds, played on Saturday afternoons (from September until May). Typically, there is one month between two consecutive rounds. In a round, each player plays one game (a game takes about 4 hours).[19]

To reduce measurement error in PM measurements, we focus on games within 5km

[19]When chess clubs have several teams in the national competition, and the team plays at home, then all games are played at the same location.

Table 2.1: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
Draw	4,415	0.32	0.47	0.00	1.00
Mean rating game (100 points)	4,415	21.06	1.56	16.80	26.26
Abs. rating difference game (100 points)	4,415	1.11	0.94	0.01	8.08
Distance to PM monitor (km)	4,415	2.33	1.03	0.27	4.96
Distance to weather station (km)	4,415	10.86	6.95	0.65	36.77
Distance between home and visitors (km)	4,415	68.75	43.06	20.22	266.54
PM10 (10 μg/m^3)	4,415	2.91	1.41	0.57	13.77
Temperature (Celsius)	3,897	7.71	4.72	−7.80	20.10
Radiation (Watt/m^2)	3,897	0.68	0.54	0.02	2.32
Air Pressure (1000 hPa)	3,887	1.02	0.01	0.98	1.04

of a PM station (we come back to this in the sensitivity analysis). Furthermore, we exclude a limited number of games further than 50km of a weather station. We also make another selection, which is not essential, but improves interpretation. To reduce correlation between PM observations measured at the home club and the visitors club, we concentrate on games with a minimum distance between home location and visitors location of 20km. We also require the presence of a PM station within 20km of the visitor club's location. Given these restrictions, the average distance to the PM monitor is slightly above 2 km, hence the distance to the nearest weather station is small and about 11 km.[20] The share of draws is 0.32. The average PM level is about 30 μg/m^3. Given these restrictions, we have almost 20k games played by 3,326 players at 81 different locations (see Table 2.1). We do not always observe weather conditions. When we focus on games for which we observe weather conditions, we observe more than 17,000 games for more than 1,000 clusters, defined as unique PM location-day observations.

For each player, we observe the so called Elo rating at the time of playing, which is an accurate numerical representation of a player's strength (Regan and Haworth, 2011). The average rating is about 2100, with a standard deviation of 157. Almost all players have a rating between 1800 and 2400. In Figure 2.A.1a in the Appendix, we provide the rating distribution. Taking risks may be perceived differently by two players who play a game depending on the difference in strength. Figure 2.A.3 shows a histogram of the (absolute) difference in rating as well as the probability that the player with the highest rating will win, draw, or lose. The difference in ratings is usually less than 300 points (less than two standard deviations), and up to that level, the weaker player

[20]The correlation between PM measurement stations for a sample with the same average distance as our main sample, is about 0.77, suggesting that attenuation bias will be about 40%. Here we use the formula $1 - \rho^2 = 1 - 0.77^2 = 0.40$, derived from Cameron and Trivedi (2005), where ρ is the correlation between PM at the measured location and PM at the chess location.

still has a reasonably high chance of winning the game.

We use daily PM_{10} measured at 63 locations provided by Netherlands National Institute for Public Health and the Environment (2019), see Figure 2.1. In addition we use daily weather observations from Royal Netherlands Meteorological Institute (2019), which include temperature, solar radiation, rain and atmospheric pressure.

2.4 Results

2.4.1 Main results

We show in Table 2.2 the estimated effects of PM on draws for a range of specifications based on equation (2.1). In specification (1), we show the effect when we control for day fixed effects, the average rating (per game) and the difference between the rating of the players. We find a positive effect of PM. The point estimate is equal to 0.015 (with a standard error of 0.005), which implies that one standard deviation increase in PM increases the probability of a draw with 2.4 percentage points. Furthermore, we find a weak (but positive) effect of the mean rating (later on we will see that this effect is entirely due to the games with higher ratings), whereas we find a rather strong effect of the difference in rating.

One criticism of this specification is that it does not control for unobserved characteristics. We deal with this in specification (2), where we include PM locations and club fixed effects. The results become somehow more pronounced.[21] We have additionally estimated models with player fixed effects. This does not change the estimates (as we control for club fixed effects, and for rating.) Another criticism is that it may provide an underestimate of the overall effect of PM, because visitors are treated for a shorter period (i.e. only during the game) than home players, who are treated before they arrive, because they tend to live locally. In line with that, we find a slightly stronger effect when we control for PM at the visitor's location, see specification (3).[22]

In specification (4), which is our preferred specification, we also control for weather conditions. As the Netherlands is small, it appears that these additional control variables do not have any effect on the estimated effect of PM. In the last two specifications, we further investigate the effect of PM on the previous day, as well as on the next day. In specification (5), we find a small positive effect of lagged PM (about half),

[21]We have also estimated models with different restrictions on the distance between home and visitors team locations. The results are not sensitive to that.

[22]Consistent with this reasoning, the point estimate of visitor's PM is negative (but not statistically significant).

Table 2.2: Regression results.

	Draw					
	(1)	(2)	(3)	(4)	(5)	(6)
PM_{10}	0.0151***	0.0183***	0.0186***	0.0197***		0.0193***
	(0.0045)	(0.0050)	(0.0051)	(0.0053)		(0.0064)
PM_{10} (visitors)			-0.0035	-0.0052		-0.0037
			(0.0048)	(0.0051)		(0.0067)
PM_{10} lag					0.0121*	0.0029
					(0.0062)	(0.0071)
PM_{10} lag (visitors)					-0.0036	-0.0004
					(0.0061)	(0.0076)
PM_{10} lead						-0.0020
						(0.0075)
PM_{10} lead (visitors)						-0.0031
						(0.0074)
Mean rating (100 pts)	0.0088***	0.0081***	0.0081***	0.0106***	0.0107***	0.0107***
	(0.0022)	(0.0027)	(0.0027)	(0.0030)	(0.0030)	(0.0030)
Abs. diff. rating (100 pts)	-0.0488***	-0.0507***	-0.0507***	-0.0489***	-0.0488***	-0.0489***
	(0.0031)	(0.0033)	(0.0033)	(0.0036)	(0.0036)	(0.0036)
Loc. & club FE		Yes	Yes	Yes	Yes	Yes
Weather dummies				Yes	Yes	Yes
Time FE	Day	Day	Day	Day	Day	Day
Clusters	1137	1137	1137	1029	1028	1028
Observations	19,763	19,763	19,763	17,327	17,310	17,310
R^2	0.0172	0.0378	0.0378	0.0432	0.0428	0.0432

Notes: PM_{10} variables are rescaled to $10\mu g/m^3$. Location fixed effects are at the level of a monitoring station. Club fixed effects contain separate controls for playing at home or as visitor. Robust standard errors in parentheses are clustered at the level of day$\times$monitoring station. ***, **, * indicate significance at 1%, 5%, and 10%.

but no effect for lagged visitors PM when we do not control for PM on the day itself. In specification (6) we show that this lagged PM effect is spurious (and entirely due to positive autocorrelation of PM). Specification (6) includes two additional placebo variables, PM and visitor PM on the next day, which are both highly statistically insignificant.

One may argue that our results are driven by reduced cognitive ability and not by reduced risk-taking. This could be the case if PM reduces the playing strength of chess players (which is very plausible) and the probability of a draw depends on players' strength. To examine the latter, we first show in Figure 2.A.1b in the Appendix the probability of a draw as a function of the players' rating. It clearly shows that the probability of a draw does *not* depend on the rating level, except for low (below 1800) and high ratings (above 2350), which occur infrequent in our data (in less than 14% of the data). This figure is slightly misleading as it ignores that the players' probability of a draw does not only depend on the player's rating, but also on the opponent's rating (and the opponents' ratings are positively correlated). We therefore show Figure 2.A.2 in the Appendix, where we show the effect of the rating on the probability of a draw, while controlling for the difference in rating between opponents. This figure confirms the previous message and even shows that there is only an effect of rating for players with a rating above 2350, which occur seldom in our data (less than 9 percent). Hence, our estimates may be slight underestimates. Notice however, that the effect of mean rating is very small, implying that even if the stronger player would play much weaker, the underestimate is still negligible.

For policy, an important question is whether the marginal effect of PM is constant, as implied by the linear specification.[23] We investigate this in several ways. An analysis using dummy indicators suggests a linear response. This is confirmed by an analysis using polynomials, see Table 2.B.1 in Appendix B. These estimates imply that the marginal effect is constant (i.e. the quadratic term is highly insignificant, whereas the linear term remains statistically significant). Hence, we do not reject linearity.

In conclusion, we find robust evidence of the effect of PM for a range of specifications, whereas placebo tests confirm that these results are unlikely by chance. Furthermore, we have demonstrated that this effect captures a reduction in risk-taking and cannot be explained by the alternative hypothesis that players make more draws because of weaker play. If anything, our estimates are underestimates of the true effect.

2.4.2 Sensitivity analysis

2.4.2.1 Distance to the PM monitor

We perform several other analyses to examine the robustness of our results to measurement error induced by the distance between the chess location and PM monitor. In particular, we have examined how the results in Table (2.2) change when we depart from 5km as maximum distance between chess location and PM monitor stations. The 5km maximum distance implies an average distance of about 2.3km, see Figure 2.2. It shows that the PM point estimate becomes smaller if we increase the maximally allowed distance, and therefore the average distance. Conversely, the coefficient increases if we reduce the maximum distance, but the confidence interval also increases

[23]On theoretical grounds, one may expect a convex function, for example as PM has to surpass a certain threshold, or a concave function, for example because a saturation level of PM kicks in.

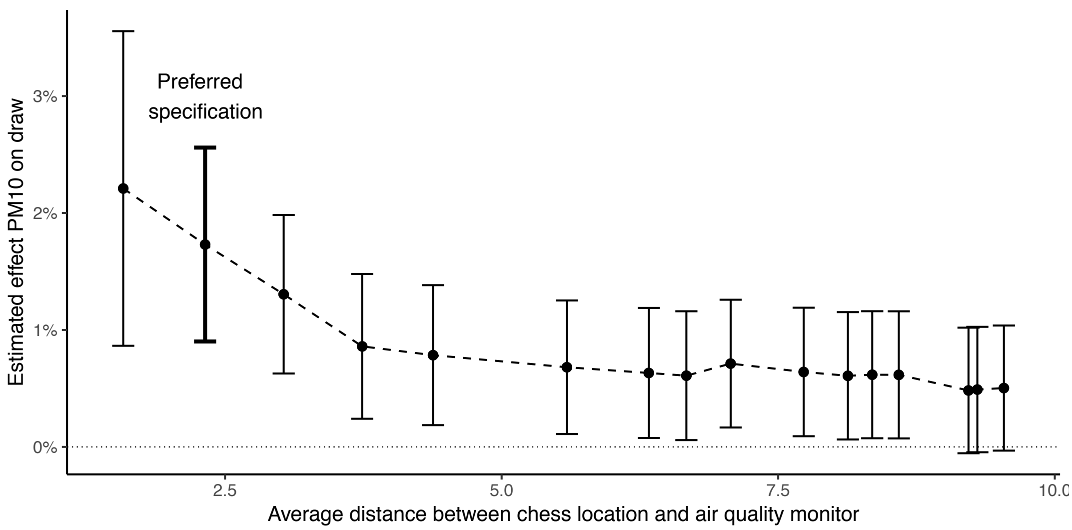


Figure 2.2: Sensitivity to distance between chess game and air quality monitoring station (error bars indicate 95% confidence intervals).

because of the reduction in observations. Hence, our preferred specification is a conservative estimate of the effect of PM on risk-taking.

2.4.2.2 Heterogeneity

One relevant sensitivity analysis is to distinguish between the effects of different rating levels. In Table 2.3, using our preferred specification, it is shown that the point estimates are positive when we distinguish between three rating subgroups (see the first three specifications), but we do not have enough power to distinguish between the PM effect of these subgroups. These estimates also confirm that, except for the subgroup of strongest players, there is no effect mean rating, and therefore of players' strength, on the probability to make a draw. Consequently, one may argue that the estimates based on samples excluding the strongest players subgroup are more accurate if one is interested in the effect of PM on risk outcomes. If one accepts this view, then the estimated effect in Table 2.2 is a slight underestimate of the effective PM on risk-taking. We come to the same conclusion if we do not control for visitors PM (see the last three specifications).

Another form of heterogeneity that may be interesting to examine, is whether the effect of the PM varies between players, because PM exacerbates existing stress levels. We cannot directly test this. However, it may be the case that stress levels are related to the difference in ratings between players. Additional analysis indicates that the

Table 2.3: Regression results using ELO subsamples.

	Draw					
	(1)	(2)	(3)	(4)	(5)	(6)
PM_{10}	0.0132	0.0266***	0.0100	0.0132	0.0255***	0.0101
	(0.0115)	(0.0081)	(0.0096)	(0.0115)	(0.0081)	(0.0097)
PM_{10} (visitors)	0.0006	-0.0149*	0.0042			
	(0.0095)	(0.0082)	(0.0114)			
Mean rating (100 pts)	-0.0023	0.0049	0.0419***	-0.0023	0.0046	0.0419***
	(0.0128)	(0.0103)	(0.0111)	(0.0128)	(0.0103)	(0.0111)
Abs. diff. rating (100 pts)	-0.0414***	-0.0460***	-0.0698***	-0.0415***	-0.0460***	-0.0697***
	(0.0067)	(0.0058)	(0.0081)	(0.0067)	(0.0058)	(0.0081)
Loc. & Club FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather dummies	Yes	Yes	Yes	Yes	Yes	Yes
ELO subsample	<2000	2000-2200	>2200	<2000	2000-2200	>2200
Clusters	883	1009	760	883	1009	760
Observations	4,942	8,191	4,322	4,942	8,191	4,322
R^2	0.1000	0.0680	0.1118	0.1000	0.0677	0.1117

Notes: PM_{10} rescaled to $10\mu g/m^3$. Location fixed effects are at the level of a monitoring station. Club fixed effects contain separate controls for playing at home or as visitor. Robust standard errors are clustered at the day×monitoring station. ***, **, * indicate significance at 1%, 5%, and 10%.

marginal effects of PM does not depend on the Elo rating difference.[24]

2.4.2.3 Other dependent variables

In Table 2.4 we perform consistency checks by analysing the effect of PM on the probability that the stronger player wins and on the probability that the weaker player wins using linear probability models.[25] Given the reasonable assumption that players maximize expected outcome (and hence their rating) the stronger players win *less* due to PM (because the stronger player reduces the outcome variance by taking *less* risk, which comes at the cost of having fewer wins). Additionally, the weaker player cannot win *more* due to PM (because a draw exceeds the expected outcome for this player). This assumption also implies that the stronger player's effect on the proba-

[24] We have also estimated logit models using the same specification. The average marginal effects are almost identical to those in the linear model. Because the difference in Elo ratings between players strongly reduces the probability of a draw, the relative effect of PM becomes stronger when the absolute difference in Elo ratings increases.

[25] We have also estimated a multinomial logit models with three outcomes (stronger player wins, draw, weaker player wins). Results are almost identical to the results in Table 2.4.

Table 2.4: Regression results for game outcomes.

	Draw	Stronger wins	Weaker wins	Home wins	Away wins
	(1)	(2)	(3)	(4)	(5)
PM_{10}	0.0183***	-0.0118***	-0.0062	-0.0172***	-0.0011
	(0.0050)	(0.0044)	(0.0040)	(0.0056)	(0.0061)
Mean rating (100 points)	0.0081***	0.0000	-0.0080***	-0.0111***	0.0030
	(0.0027)	(0.0028)	(0.0022)	(0.0028)	(0.0028)
Abs. diff. rating (100 points)	-0.0507***	0.1265***	-0.0732***	0.0407***	0.0100**
	(0.0033)	(0.0036)	(0.0026)	(0.0044)	(0.0041)
Loc. & Club FE	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes
Clusters	1137	1137	1137	1137	1137
Observations	19,763	19,763	19,763	19,763	19,763
R^2	0.0378	0.0797	0.0551	0.0405	0.0327

Notes: ..***, **, * indicate significance at 1%, 5%, and 10%.

bility of winning must be stronger than the weaker player's effect on the probability of winning in absolute value, i.e. increased number of draws should mainly come at the cost of the strongest player's wins (otherwise taking less risk as the stronger player would be a dominant strategy i.e. less risk and more wins).[26] Columns (2) and (3) confirm these predictions and support our claim that PM induces less risk-taking. It suggests that people are willing to trade off a lower expected pay-off with a safer approach, in line with Duflo and Banerjee (2011). It thus appears that players take less risk than what they would prefer without PM.

In columns (4) and (5) we test whether there is a difference of the PM effect for home and visiting players. It appears that home players are affected more strongly by PM pollution. This makes sense and suggests that PM exposure prior to the game (but on the same day) has a detrimental effect in addition to the exposure during the game itself.

Finally, for games played in the highest league (about 10 percent of our sample), we know the moves. This offers alternative ways of doing a sensitivity analysis, because if PM induces players to make more draws, then it must be true that they either play less risky moves or they are more likely to agree to a draw (which nullifies the risk of losing), which will result in shorter games (i.e. games with less moves), given higher

[26] Conversely, the weaker player should not win more often due to PM. It is however possible that there is *no* effect on the number of wins of weaker player, as increased draws are *favourable* for the weaker player.

levels of PM. We find evidence of both mechanisms.[27] However, the results are not robust with respect to specification (e.g. controlling for weather) and sample selection, which is not surprising given that we have a small subsample. Most importantly, for all specifications, we either find no effect (due to large standard errors), or we find statistically significant results that support that PM reduces risk-taking.

2.5 Conclusion

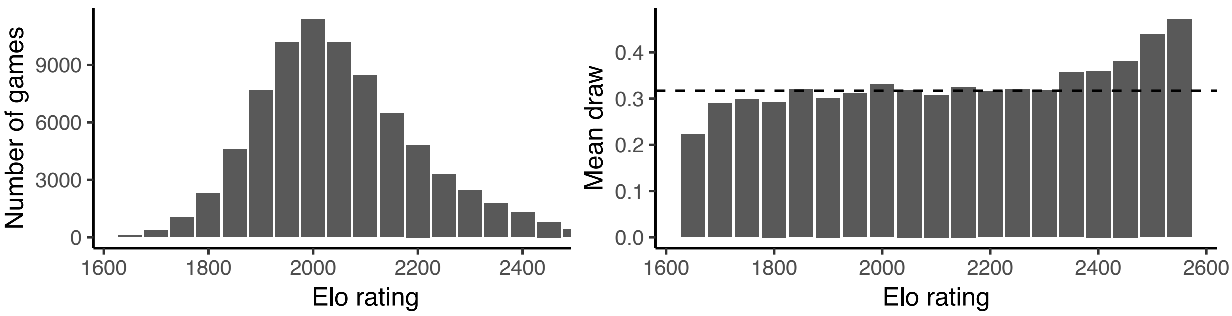
Air pollution has been shown to affect cognitive ability and is hypothesized to decrease an individual's risk-taking. This hypothesis emerged from earlier literature that finds detrimental effects of particulate matter on composite decision outcomes (e.g. stock market returns in Heyes et al., 2016). Because it is unclear which mechanism drives these results, in this paper, we focus specifically on risk-taking using the game of chess as a natural experiment.

We estimate the effect of PM on the probability that chess players make a draw, which directly reflects the variance of game outcomes, and is a clean indicator for risk-taking. We use information from the Dutch national team league, where games are played at the same time in different locations. This setting comes close to the ideal experimental setup, as air pollution levels vary over time and space.

Our results show that PM induces chess players to make more draws. We find that A 10μg increase in PM_{10} (33.6% of mean concentration) leads to a 5.8% increase in draws. We do not find any effect of PM when measured at the location of the visiting club, or of effects of PM on previous days, which implies that the PM effect is immediate. Our results demonstrate that air pollution reduces risk-taking.

[27] We have classified risky play using several measures distinguishing between opening risk, where risk is based on the opening's share of draws, opposite castling, and white plays G4 in the opening. We also demonstrate that these measures are valid measures of risk-taking as they are strongly related to the probability of making a draw. Results can be received upon request.

Appendix 2.A Additional descriptives



(a) Histogram of Elo rating.

(b) Draw as a function of rating.

Figure 2.A.1: Frequency of Elo rating and draw per rating.

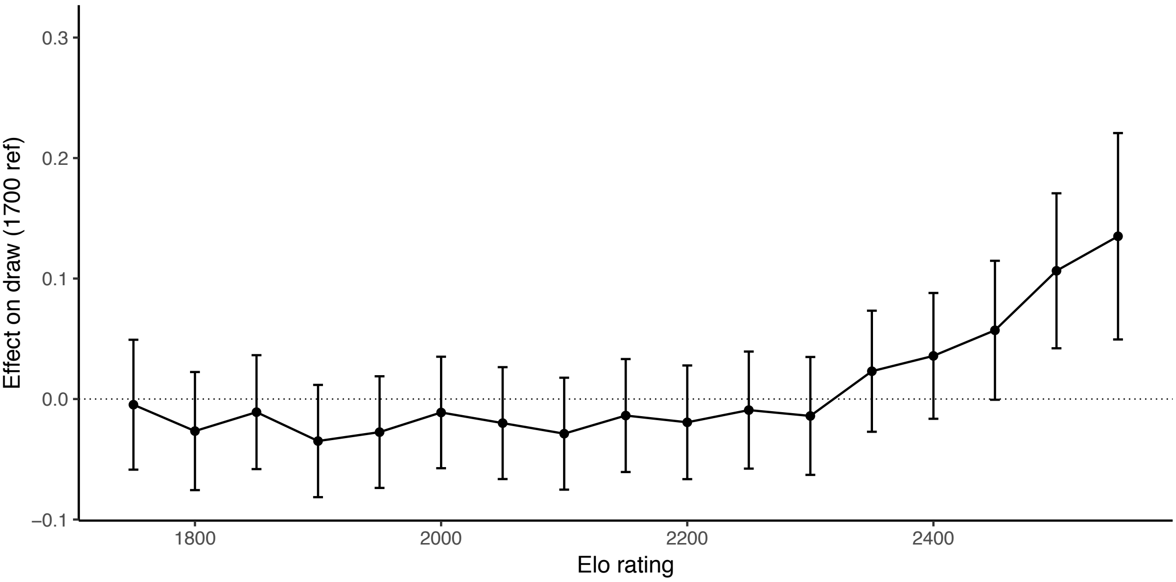
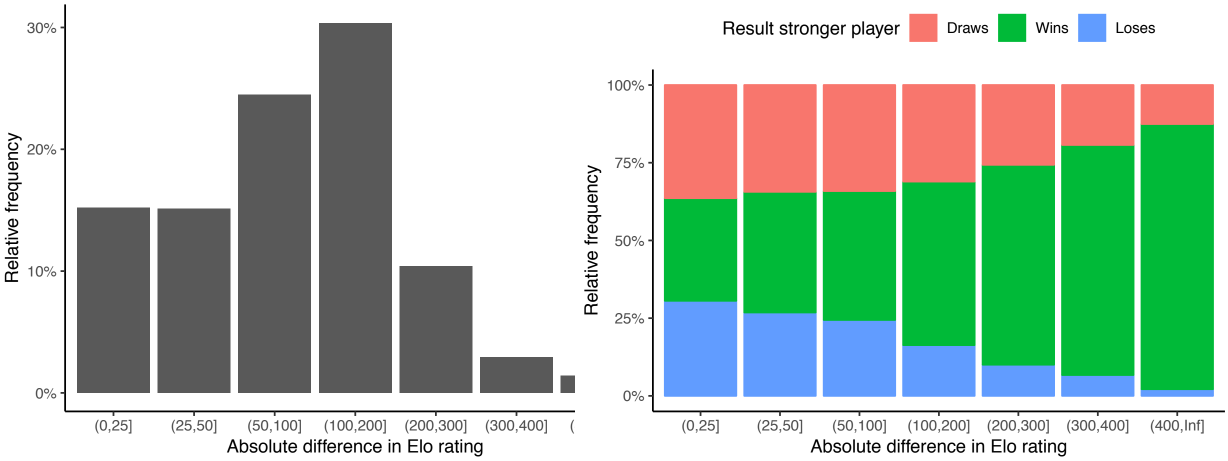


Figure 2.A.2: Effect of players strength (mean rating per game) on draw, conditional on difference in rating.



(a) Histogram of difference in rating (b) Results player with highest rating

Figure 2.A.3: Histogram difference in rating and results stronger player.

Appendix 2.B Additional results

Table 2.B.1: Regression results with higher order polynomial terms.

	Draw		
	(1)	(2)	(3)
PM_{10}	0.0183***	0.0243**	0.0173
	(0.0050)	(0.0115)	(0.0252)
$(PM_{10})^2$		-0.0006	0.0009
		(0.0009)	(0.0045)
$(PM_{10})^3$			-0.0001
			(0.0002)
Mean rating (100 points)	0.0081***	0.0081***	0.0081***
	(0.0027)	(0.0027)	(0.0027)
Abs. diff. rating (100 points)	-0.0507***	-0.0507***	-0.0507***
	(0.0033)	(0.0033)	(0.0033)
Loc. & team FE	Yes	Yes	Yes
Weather dummies	Yes	Yes	Yes
Day FE	Yes	Yes	Yes
Clusters	1137	1137	1137
Observations	19,763	19,763	19,763
R^2	0.0378	0.0378	0.0378

Notes: PM_{10} variables are rescaled to $10\mu g/m^3$. Location fixed effects are at the level of a monitoring station. Club fixed effects contain separate controls for playing at home or as visitor. Robust standard errors in parentheses are clustered at the level of day×monitoring station. ***, **, * indicate significance at 1%, 5%, and 10%.

3

Urban air pollution and time losses Evidence from cyclists in London

3.1 Introduction

Urban air pollution is of growing concern for policy makers in cities across the globe. Nitrogen oxides, NO_x, particulate matter, PM, and ambient ozone, O_3, are pollutants of primary focus. NO_x are predominantly emitted by motorized traffic, whereas ambient ozone forms more slowly after complex interactions between NO_x, volatile organic compounds, solar radiation and heat (Krupa and Manning, 1988; Sillman, 1999).

Ambient, or tropospheric, ozone (henceforth ozone) is present at ground level and is linked to negative health outcomes (Goldberg et al., 2001; Fann et al., 2012).[28] NO_x are

This chapter is based on Klingen and van Ommeren (2020b) as published in *Regional Science and Urban Economics*. For this chapter I would like to thank Erik Verhoef, Paul Koster, Devi Brands, Yann Delaprez, Gerben de Jong, Francis Ostermeijer, Jiska Klein, Hans Koster, Thomas de Graaff, Felipe Carozzi, Matthew Neidell, Gabriel Ahlfeldt (editor), and two anonymous referees for their helpful comments. I also thank seminar and conference participants in Amsterdam (Eureka, Cycling Research Board), New York City (UEA), Düsseldorf (UEA), Paris (ITEA), and Nicola Franks from TfL for providing the traffic counter data

[28]Ozone pollution as analysed in this paper should not be confused with stratospheric ozone—known as the ozone layer—that is linked to positive health outcomes due to its capacity to absorb ultraviolet radiation.

precursors of ozone and therefore core contributors to poor health outcomes induced by pollution from road traffic (Hamra et al., 2015; Rice et al., 2015).

Much of the literature focusses on long-term health consequences of ozone and other air pollution (see e.g. Chappie and Lave, 1982 and Beach and Hanlon, 2018). In contrast, we focus on the immediate effect, for which there is growing attention in the literature (Graff Zivin and Neidell, 2018). Recent controversy surrounding emission tests of diesel vehicles and NO_x inhalation studies further highlights that the immediate effects of air pollution are at the centre of policy and scientific attention (i.e. 'dieselgate', see e.g. Anenberg et al., 2017 and Ewing, 2018).[29] Medical research using laboratories shows that ozone affects the respiratory function of lungs, such that exercise capacity is temporarily reduced (Cakmak et al., 2011; Adams et al., 1987), especially in a hot environment (Gong et al., 1986). It is therefore plausible that ozone pollution impedes physical activity of individuals recently or currently exposed to ozone.[30]

One immediate effect of ozone is on labour productivity. We are aware of two studies that assess the effect of ozone pollution on productivity in the field (see the review of Neidell, 2017). Zivin and Neidell (2012) show that agricultural workers in California are less productive on days with high ozone levels. They estimate that a 10 ppb increase in ozone leads to a 5% reduction in productivity for higher ozone levels (above 42 ppb). Lichter et al. (2017) do not find an effect of ozone, but identify a small effect of PM pollution on some productivity indicators of players in the German football league.[31]

We hypothesise that ozone pollution slows down all outside activities which require a minimal physical effort, for example when walking or cycling is involved. The effect of ozone would then apply to activities such as pedestrian shopping, which is in many countries the dominant walking activity (Koster et al., 2019), but in some cities also to commuting, as walking or cycling are an important part of the commute. For example, in London about 20 percent of commuters exclusively walk or cycle to and from work. For non-commuting trips, the share of walking and cycling is even

[29]One important question is the total burden of NO_x pollution. Because NO_x are precursors for ozone, the immediate effect of ozone is of interest when estimating this burden.

[30]In addition, it is well known that for elderly, walking speed is a very strong predictor of survival rates (Studenski et al., 2011), but also for younger persons, walking speed is an important indicator of health (Martin et al., 1992). Therefore, in addition to time losses in themselves, the lower cycling speeds that we find might also be indicative of general health losses caused by ozone.

[31]There are related studies that focus on PM pollution. Chang et al. (2019b; 2016) study productivity of pear packers and call centre employees and find adverse effects of PM pollution on their productivity.

higher (Transport for London, 2016, 2018c).[32]

Consequently, one expects that the ozone effects are much more generic then suggested by studies focusing on labour productivity of outside workers, which applies to a limited number of workers in modern economies. There are also other important differences. In contrast to workers, individuals involved in most outdoor activities (such as commuters who cycle to work) are not exposed to ozone for a full working day, and exert physical effort for a much shorter amount of time. As ozone effects seem to increase in exposure and exercise duration, it is plausible that the size of the effect of ozone on travel time losses of most outside activities is (much) smaller than for workers.

The current paper sheds further light on effects of air pollution, by inferring a causal effect of ozone on cyclists' speed in London for the years 2013-2017, while using hourly measures of air quality and over 40 million trips made on public rental bicycles.[33] About 40% of these trips are made for commuting purposes.[34] Our study contributes to the existing air pollution literature in four ways.

First, we provide non-experimental field evidence of health effects of local air pollution. By focusing on time losses during cycling, we estimate the effect of air pollution on the outcome of a common, non-specialized, and rather light, type of effort (with plausible generalization to other physical activities). The question is here whether the large effect sizes reported by productivity studies also hold for other physically less-demanding activities.[35]

Second, to our best knowledge, this paper is the first to investigate the immediate effect of ozone pollution in an explicitly urban context, where due to high levels of air pollution and large concentrations of persons, negative externalities related to air

[32]These shares generally underestimate the total amount of walking and cycling in cities. For instance for London, where distances between underground stations are large, many public transit commuters walk quite extensively, whereas in cities such as Amsterdam, many public transit users cycle from their home to public transit stations.

[33]We use other pollutants, such as NO_x and PM, as control variables, but we do not aim to estimate the causal effect these pollutants, because we are not aware of medical research showing a direct effect of other pollutants than ozone on physical effort of healthy individuals. Furthermore, it remains unclear how one convincingly may estimate a causal effect of NO_x or PM, due to possible confounding effects from car traffic.

[34]This estimate is based on information in Transport for London (2018b,c) which indicate that 60% of trips are made by members, of which 72% uses the rental bicycles for commuting.

[35] The effect of ozone on cylists' power output has been shown in the laboratory by Gong et al. (1986), who find a 8% reduction in output power given exposure to 200 ppb. The external validity of that result to other physical activities may be limited, because it focuses on 17 elite cyclists who exercise intensively for more than an hour and who were exposed to ozone concentrations of 120 and 200 ppb, more than twice the maximum level observed in London. Remarkably, when we extrapolate our results to 200 ppb, we also find a point estimate of around 8% reduction in cycling speed.

pollution are most urgent.

Third, in contrast to above-mentioned studies that identify effects using daily observations, our hourly data allow for a wider range of identification strategies. We exploit within-day and between-day variation, as well as spatio-temporal variation in ozone levels. Having hourly data not only allows for many robustness checks, it is also informative about the impact of within-day pollution peaks. The latter is especially relevant for environmental regulation, which may concentrate either on daily averages or on peak pollution.

Fourth, due to our large sample, we identify an effect of ozone at much lower levels of ozone than suggested by previous studies. We are able to find a minimum ozone threshold at concentrations as low as 20 ppb, far below EU air quality standards set above 50 ppb (European Environmental Agency, 2019).[36] These figures make it plausible that previous studies severely underestimate the extent of the effect of ozone.

This paper relates to a rich urban economics literature that investigates the relation between density and air pollution (Ahlfeldt and Pietrostefani, 2019). Theoretical and empirical contributions highlight that inhabitants of compact cities usually face higher local air pollution levels (Borck and Tabuchi, 2018; Carozzi and Roth, 2020; Borck and Schrauth, 2019). Related recent work suggests that a shift towards public transport can alleviate pollution levels in cities (Borck, 2019).

Our paper also fits within an urban transport economics literature, where the focus is on speed, which is (optimally) chosen by travellers (Verhoef and Rouwendal, 2004; Fosgerau, 2005; Van Ommeren and Dargay, 2006; Couture et al., 2018). In line with that literature, we assume that cyclists choose speed by maximising their utility with respect to travel time, conditional on travel distance.[37]

Utility is then defined by the (negative) of the sum of effort cost, $c(t, z)$, a decreasing and convex function of travel time, which in turn is an increasing and convex function of ozone, and travel time costs, $(w + p)t$, where w denotes the value of time, p denotes the rental fee per unit of time, z denotes ozone, and t denotes the individual's travel time. The cyclists that we study thus have two incentives to increase their cycling

[36] Our threshold is also much lower than $100\mu\text{g/m}^3$ ($\approx$ 50 ppb) set as air quality guideline by World Health Organization (2006). These guidelines do mention, however, that some health effects may be visible for lower concentrations, which are not further specified. In here, we use an approximate conversion rate of 1 ppb $\approx 2\mu\text{g/m}^3$, which depends on temperature and pressure (Levy et al., 2005).

[37] One might argue that ozone directly affects cycling decisions, for instance because of individuals' health concerns related to exposure. However, we find no evidence that ozone affects the decision to cycle, or trip distance. See Section 3.3 for further details. Therefore, it seems reasonable to abstract from health concerns.

speed: through the rental fee p and through their value of travel time w.[38] These incentives highlight that our context is comparable to studies based on work-related observations, although the exact incentive schemes differ. It seems further reasonable to assume that the marginal effort cost of travel time is an increasing function of ozone, denoted by z, so $dc/dz > 0$, because of decreased lung capacity. It is then straightforward to show that $dt^*/dz > 0$, so optimally chosen travel time, t^*, is an increasing function of z. Consequently, it follows that speed falls in ozone.

This paper proceeds as follows. Section 3.2 describes the data and presents descriptive statistics. Section 3.3 explains the methods employed. Section 3.4 discusses the results and robustness checks. Section 3.5 concludes.

3.2 Data

We combine data on air quality, weather, traffic, and cycling trips from Inner London for five years, from January 2013 to December 2017. The map in Figure 3.1 indicates where the data is collected, and shows the locations of the air quality monitor sites, bicycle docking stations, and automatic traffic counters (ATCs).

3.2.1 Cycling data

We use publicly available data from London's rental bicycle system *Santander Cycles*, provided by Transport for London (2019).[39] The pricing scheme works as follows. Individuals can choose between an annual subscription (which costs £90) or a daily subscription (which costs £2), that both allow for unlimited bike hires under 30 minutes. For trips exceeding 30 minutes, users pay additionally £2 for each additional 30 minutes or less. There is no price differentiation over time or space, and users cannot pause a trip other than by docking a bike in a station.

About 60% of trips are made by members, who get a special key to unlock bikes. Most of these members are men (78%) and between 25 and 54 years old (84%). About

[38] We are not aware of any study estimating the value of time for cyclists in London, w, but a Swedish study suggests it is about £10 per hour for cycling commuters (Börjesson and Eliasson, 2012). Presumably, the value of time is less for leisure trips. In London, the rental fee per unit of time is not constant. For trips shorter than 30 minutes, the marginal fee is zero, as renters only pay an initial fee (usually £2) regardless of the trip duration, provided it is under 30 minutes. 11.56% of trips last longer than 30 minutes. For those trips, the marginal fee p is approximately £4 per hour. See Section 3.2.1 for further details.

[39] This is the only public bicycle system in London that has docked bikes, and the only major bicycle rental scheme active in London.

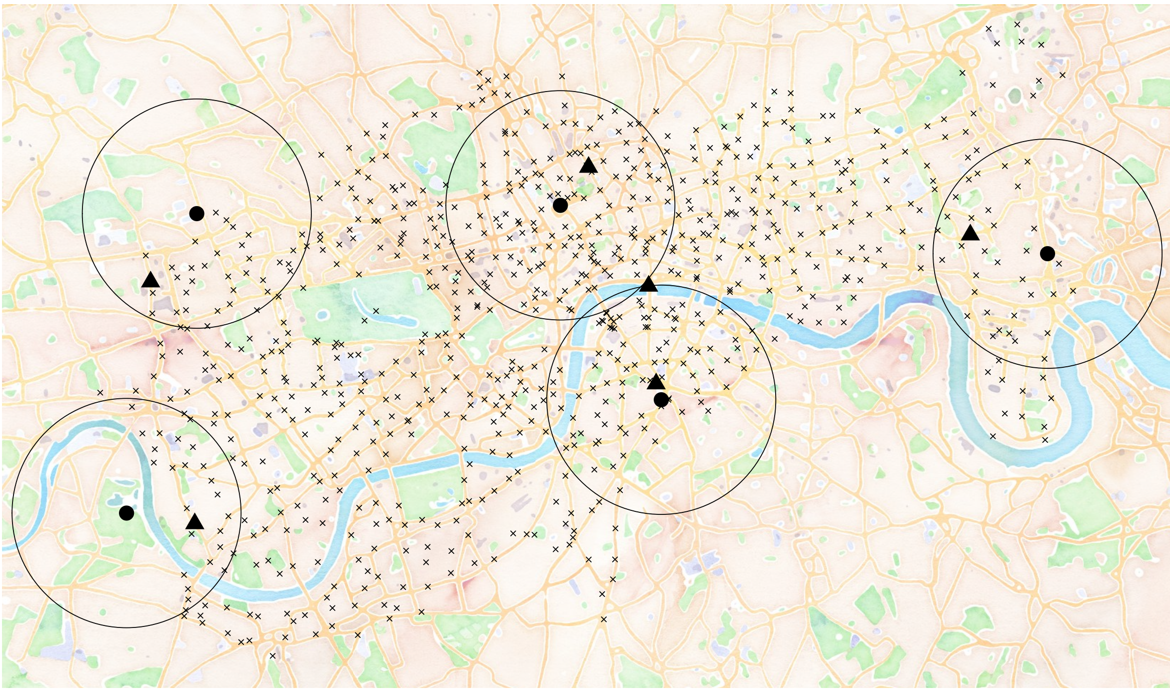


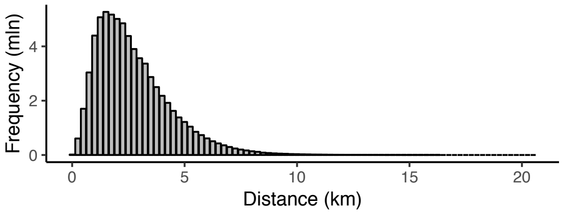
Figure 3.1: Location of bicycle docking stations and air quality monitor sites.

Notes: ×, ●, and ▲ represent bicycle docking stations, air quality monitoring stations, and automatic traffic counters respectively. The circles surrounding the monitoring stations demarcate the five zones with 2km radius used in the spatio-temporal analysis (see Section 3 in Appendix B). Map tile by Stamen Design, under CC BY 3.0, base layer data by OpenStreetMap, under CC BY SA.

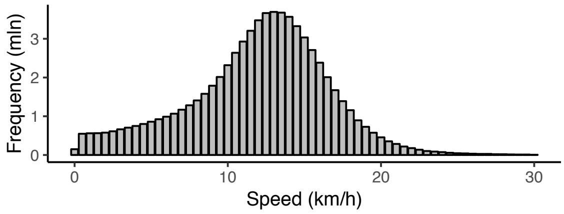
70% of members use the public bicycle scheme mainly for commuting, and about half use a rental bicycle for more than three days a week (Transport for London, 2018b,c). About two thirds of non-members, i.e. 'casual users', live in London, which highlights that only a minority of cyclists are tourists (maximally 15%). Roughly a quarter of cyclists start their trip around a rail or metro station, which makes it plausible that cycling is often part of larger trip that includes public transport (Transport for London, 2018a).

We observe 73.23 million time-stamped origin-destination pairs, covering the universe of cycling trips made on rental bicycles from 2013 to 2017. Because the exact routes are not observed, we infer a measure of speed for each trip by using the trip duration and distance, calculated as the shortest path over the road network.[40]. Figure 3.A.2 in Appendix A shows the inferred routes, and highlights the central areas

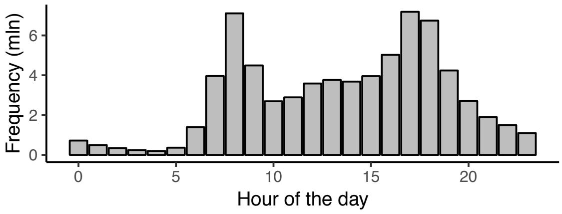
[40]We calculate the shortest paths over the bicycle infrastructure using the algorithm developed by Padgham (2019), that we use within our main analysis tool (R Core Team, 2019).



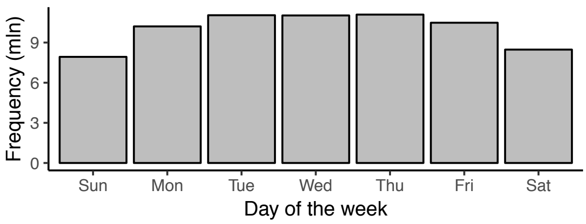
(a) Distribution of distances



(b) Distribution of speed



(c) Nr. of trips during the day



(d) Number of trips per day of week.

Figure 3.2: Public rental bicycle descriptives.

have the highest flows.

Because cyclists might not take the shortest route, our indicator for speed is an underestimate of the actual speed.[41] The estimated effect on speed in levels is therefore a conservative estimate (but the estimated effect on speed relative to mean speed is still accurate).

We exclude individual trips with zero distance, i.e. cyclists who return their bicycle to the same docking station (3.81% of observations).[42] Further, we exclude trips with a speed exceeding 30 km/h or less than 0.1 km/h (0.16% and 0.09% of observations). Our cycling data then contains 70.26 million observations (95.94% of initial amount).

In total, we observe 21.42 million cycle hours and 191.83 million cycle kilometres. The average trip distance is 2.73 kilometres, and the vast majority of trips is below 5 kilometres, see Figure 3.2a. Trips take on average 18.29 minutes, during which travellers

[41] In addition, a trip may include stopping at traffic light, so we underestimate cruising speed. This is non-problematic because most physical activities (e.g. outside physical labour) also include stops and breaks.

[42] This includes round trips and hires of bicycles with a technical failure that are returned immediately. 17.48% of zero-distance observations have a duration of less than 10 minutes.

have an average speed of 11.97 km/h.[43] Figure 3.2b shows that the distribution of speed is slightly skewed to the left and that observations with speeds above 25km/h are rare (0.33%). Figure 3.2c shows that there are clear peaks of bicycle activity during rush hours, while Figure 3.2d indicates that there are slightly more trips during weekdays.

For a robustness check using a spatio-temporal estimation (see Appendix 3), we construct five circular zones, each surrounding one of the air quality monitoring stations in London (see the map in Figure 3.1). To avoid overlap between the zones, we set the radius of the zones to two kilometres. We then assign a trip to a zone if *both* the origin and destination are within that zone. This leaves us with a subsample containing 8.51 million observations, 12% of the initial sample.

3.2.1.1 Constructing a panel of cyclists

The cycling data does not provide us with an identifier for individuals. Here, we show how we construct a panel of cyclists using regular patterns in departure times for each origin-destination pair (henceforth route). Our starting point is that most commuters make the same trip around the same time on different days of the week (Noland and Small, 1995). Identifying individual cyclists is possible because most routes in our data are very 'thin': for roughly 75% of all trips there is not even *one* other cyclist who cycles the same route within the same hour. For longer distances, this percentage is even higher (see Figure 3.A.3 in Appendix A). This makes it plausible—but not certain—that trips observed around the same time on different days within a time window, e.g. a month, are made by the same individual.

To identify individuals, we construct a panel of cyclists by first selecting trips for which no other cyclist is observed who takes the same route within the same hour (only in these cases it is possible to identify individuals). We then make several further data selections based on the idea that, for a given route, trips on different days with departures around the same time (i.e. within the same hour of the day), are more likely to come from the same cyclist than from a random group of several cyclists. To implement this idea, we apply the following data selection steps.

First, for each route we select sets of recurring trips in one month: trips that are made on different days with departures within the same hour of the day. For instance, such set may consist of three trips, made on three different days, with departures on 9:12, 9:14, and 9:16, in March 2017, from Kennington Station to Tower Gardens.

[43] This is well within the range of 11-14 km/h as observed by Jensen et al. (2010) for Lyon (using odometers). This suggests that most cyclists take a route that is at most slightly longer than the shortest path, which makes it plausible that observed speed is hardly an underestimate of the actual speed.

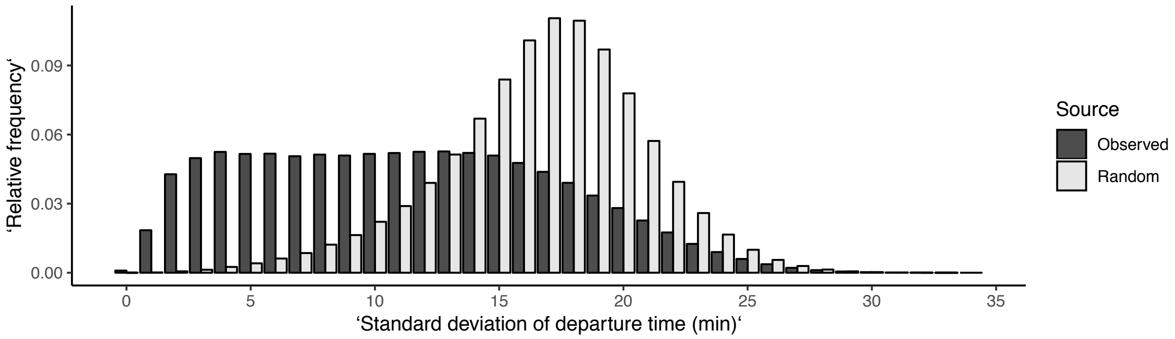


Figure 3.3: Distribution of standard deviations of departure time.

Second, for each of those sets we calculate the standard deviation of departure time. We compare this standard deviation with one from (simulated) random draws from the departure time distribution of all departures in the selected sample (see Figure 3.A.4 in Appendix A for the histogram of departure time).[44] Figure 3.3 shows that the distribution of observed standard deviations has much more mass to the left compared to the one based on random departures.[45] The average observed standard deviation is 11.9 minutes, rather than the 17 minutes standard deviation from random departures. This indicates that a large part of our observations comes from the same individuals' recurrent trips. Observations with a relatively large standard deviation (e.g. 20 minutes) are still likely to come from the same person—recall that we focus on very thin routes—but less likely than those with a small standard deviation.

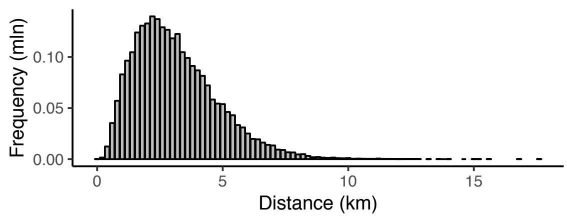
Third, we select sets with a small standard deviation. To be more precise, we select a subsample of sets that have a standard deviation below a threshold, that is set in such a way that maximally 5% of observations might be erroneously identified as part of a single-person set.[46] After this selection, the average standard deviation is only 4.6 minutes, much lower than the standard deviation of 17 minutes of a random group.

The resulting panel contains 2.58 million trips (3.7% of initial sample), made by 0.31

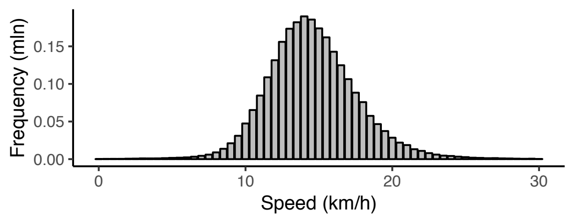
[44]This distribution resembles a uniform distribution and has a mean of 30 minutes and a standard deviation of 17 minutes. Both moments of the empirical distribution are the same as the theoretical moments of a uniform departure time distribution within a 60 minutes interval.

[45]For this figure, we have excluded sets with less than 4 trips per month, as these observations are not part of the resulting panel, because there is hardly any power to distinguish between random departures and those coming from the same cyclists for sets with less than 4 trips.

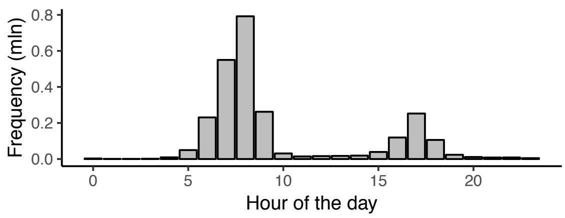
[46]The standard deviation depends on the number of trips in a set, so we calculate the cumulative distribution function for both the observed and random standard deviations separately for each number of trips. We then determine threshold values of standard deviations at which the ratio between random and observed mass exceeds the cut-off of 5%. Sets with an observed standard deviation above the threshold (i.e. cyclists with too much variability in their departure time) are excluded.



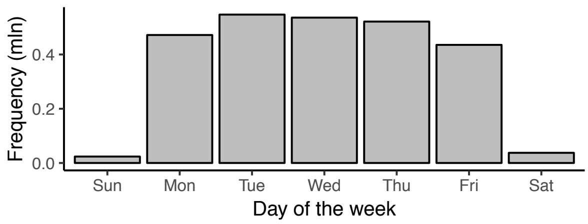
(a) Distribution of distances.



(b) Distribution of speed.



(c) Number of trips per hour of the day.



(d) Number of trips per day of week.

Figure 3.4: Descriptive statistics of cyclist panel.

million cyclists, who cycle at least four times, and on average more than eight times, per month. The cyclists in the panel differ from the overall population of cyclists in several respects. Figures 3.4a and 3.4b show that for this subsample, cycling speed and trip distances tend to be higher than for the full population. Figures 3.4c and 3.4d show that the large majority of trips takes place during peak hours on weekdays. We observe more trips in the morning than in the afternoon in our panel, which is not the case for the whole sample. The asymmetry stems from selection based on a small standard deviation of departure time, which is more common in the morning. This makes sense, as it is plausible that commuters have a relatively fixed schedule in the morning, whereas the timing and destination of the return trip is more variable. For instance, if commuters depart later due to overtime work, or go to a pub after work, their trips are excluded from the panel. Overall, these figures are highly suggestive of having a subsample that mainly consists of commuters, rather than recreational cyclists.

3.2.2 Air quality and weather data

We use hourly environmental data that is publicly available through London Air (2019). For each of the five monitoring sites we observe: ozone (O_3), particulate matter (smaller than $2.5\mu m$, $PM_{2.5}$), and nitrogen oxides (the sum of NO and NO_2, henceforth NO_x). We calculate the city average using the same weight for each station. In

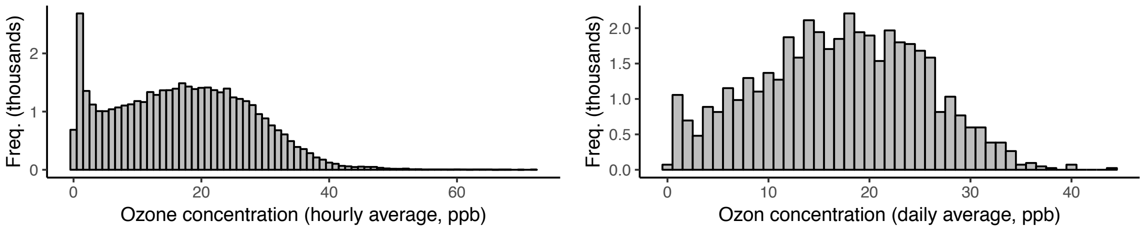


Figure 3.5: Histograms of ozone concentrations.

addition, we have city-wide hourly measures of the following weather characteristics: temperature, relative humidity, rainfall, wind speed, wind direction and atmospheric pressure.[47] We obtain solar radiation from two stations outside Inner London.[48]

Figure 3.5 shows that hourly and daily ozone concentrations rarely exceed 40 and 30 ppb respectively. These levels are much lower than in California during the summer months (Zivin and Neidell, 2012), but only slightly lower than in Germany during the football season (Lichter et al., 2017).[49] Appendix A presents further descriptives of the (spatial) variation of ozone.[50]

3.2.3 Traffic data

We observe hourly traffic flows for the full 2013–2017 period as collected by Transport for London (TfL). These flows are measured at six different automatic traffic counters (ATCs) and data is extracted from TfL's Patched ATC Database. The map in Figure 3.1 shows the locations of the ATCs. If a counter is located at bidirectional streets (all but one), then it reports traffic for both directions separately. We standardize the observations per counter and direction (using mean traffic intensity) and include the

[47]Because for most analyses we use city averages, it is non-problematic that we observe weather not for each monitoring station separately. Further, because for the spatio-temporal robustness check we use daily averages and our context is rather small geographically, it seems hardly restrictive to assume that weather conditions are comparable across zones.

[48]These stations are located in Romfort and Stanmore, see Figure 3.A.1 in Appendix A.

[49]Less solar radiation and stronger winds—ozone is transported vertically—explain lower ozone concentrations in London. Ozone forms after complex interactions between NO_x, volatile organic compounds, solar radiation and heat (Krupa and Manning, 1988; Sillman, 1999) and is transported horizontally by air currents (Lelieveld and Dentener, 2000). Therefore, ozone concentration varies over time and space.

[50]Table 3.A.1 in Appendix A shows spatial correlations for ozone measurements of different monitor sites. These correlations are high, which explains why identification based on spatio-temporal variation results in relatively large standard errors. The autocorrelation of ozone concentration is high within days, but moderate between days, see Table 3.A.2.

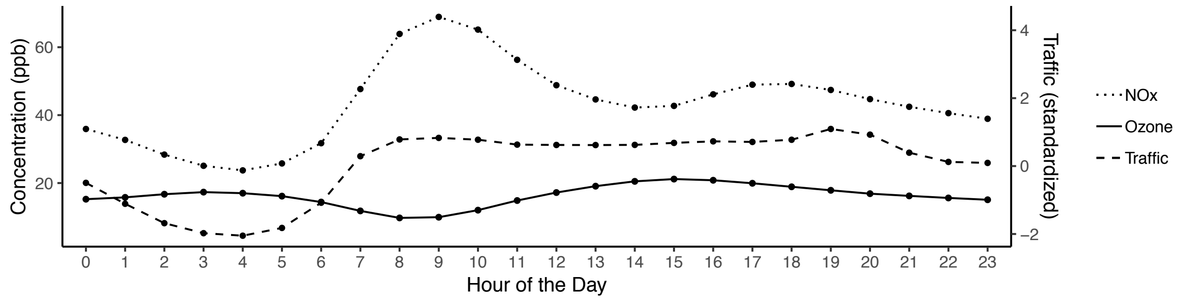


Figure 3.6: Ozone, NO_x, and car traffic levels per hour of the day.

(unweighted) city average as our measure for traffic.[51]

Figure 3.6 shows that traffic intensity is usually high between 8am and 8pm, and that NO_x levels tend to peak around rush hours (consistent with the idea that during peak hour, the density of motor vehicles is particularly high). In contrast, ozone concentration is generally high in the afternoon and low during mornings, which suggests that ozone is hardly related to traffic.

In Appendix 1 we analyse the relation between traffic and air pollution in a more rigorous way. We demonstrate that, conditional on controls, NO_x are directly related to traffic intensity, but ozone is not (see Tables 3.B.1 and 3.B.2 in Appendix B).

3.2.4 Hourly descriptives

Our analysis uses hourly and daily variables. We present descriptives at the hourly level in Table 3.1 (at the daily level are shown in Table 3.A.4 in Appendix A). The five-year sample provides us with a total of 43,831 hourly observations. We note that there is substantial variation in the number of trips (and bicycle traffic), which makes this variable an attractive indicator for sorting (discussed below).

3.3 Empirical methods

Our aim is to identify the causal effect of ozone pollution on cycling speed using observations of ozone (by hour and by zone) and cycling (location and timing of departure and arrival). Let us suppose one regresses cycling speed on ozone. There are then two main econometric challenges: omitted variable bias and sorting. Omitted

[51]We also run a robustness check where we include measurements from each ATC as a separate control.

Table 3.1: Descriptive statistics per hour.

Statistic	Mean	St Dev.	Min	Max
Number of trips	1065	1090	5	12188
Speed (km/h)	11.94	1.64	2.30	17.17
Duration (min)	18.41	6.66	6.83	329.00
Distance (km)	2.73	0.27	1.32	4.98
Bicycle traffic (km)	6167	3208	10	48211
Road traffic (vehicles per counter)	491	145	47	897
Ozone (O_3, ppb)	18.21	10.65	0.05	72.10
Nitrogen oxides (NO_x, ppb)	43.95	34.49	1.35	500.97
Particulate matter ($PM_{2.5}$, ppb))	13.05	11.02	0.00	114.86
Sulphur dioxide (SO_2, ppb)	10.61	9.70	0.00	912.00

Note: 43,831 hour observations, cycle variables are city wide hourly means of 70.26 million observations. Mean and standard deviation are weighted for the number of trips per hour.

variable bias will occur, for example, if ozone is positively correlated with intensity of motorized traffic, which may hamper cyclists, and therefore reduces cycling speed. Sorting will occur, for example, if ozone is positively correlated with nice weather, which induces additional, and likely different types of cyclists on the road. To overcome these challenges, we employ a within-day and a between-day estimation strategy.[52] We will also focus on a panel of cyclists that appear to be (predominantly) commuters.

3.3.1 Within-day estimation

Let us first ignore spatial variation in ozone and identify the effect of ozone using hourly averages in the city. We regress average cycling speed, $speed_{wdh}$, on ozone concentration, $ozone_{wdh}$, at hour h of day d in week w, while controlling for weather conditions, other pollutants, and several types of time fixed effects. We start with the

[52]We also applied two alternative strategies. In Appendix 3 we discuss a spatio-temporal strategy and show that the results are robust to such a specification. We do not show results of another alternative strategy which uses night-time thermal inversion as an instrumental variable for ozone, as used for other pollutants by Arceo et al. (2016). This approach works well for first-order pollutants (i.e. that are directly emitted) such as NO_x. However, because ozone forms after complex atmospheric interactions, thermal inversion appears a weak instrument in our data (with a first-stage F-test of about 3), despite having an observation period of five years.

following within-day specification:

$$speed_{wdh} = \beta ozone_{wdh} + \boldsymbol{\gamma} \boldsymbol{W}_{wdh} + \boldsymbol{\zeta} \boldsymbol{T}_{wdh} + \boldsymbol{\delta} \boldsymbol{E}_{wdh} + \kappa_{wd} + \eta_{dh} + \varepsilon_{wdh}, \tag{3.1}$$

where $w = 1, \ldots, W$, $d = 1, \ldots, 7$, $h = 1, \ldots, 24$ denote week, day of the week, and hour of the day respectively.[53] We include sets of control variables for weather conditions, $\boldsymbol{W}_{wdh}$, car and bicycle traffic, $\boldsymbol{T}_{wdh}$, and other environmental pollutants, $\boldsymbol{E}_{wdh}$ (all discussed in detail below). Furthermore, we include κ_{wd}, a day fixed effect, and η_{dh}, an hour-of-the-week fixed effect. Including day fixed effects has the advantage that we identify our parameters within a day, and thereby control for unobserved day characteristics. Given these controls, we will argue below (see sections 3.3.1.1 and 3.3.1.2) that ozone is uncorrelated with the disturbance ε_{wdh} and hence we can estimate (3.1) with ordinary least squares. Because we are interested in the effect of ozone on cycling speed *per trip* we use weighted least squares with the number of trips per hour as weights.

3.3.1.1 Omitted variable bias

Weather conditions may lead to omitted variable bias, because they correlate with ozone and also affect cycling speeds directly, for instance when higher temperatures reduce exercise capacity. To solve this issue, we use $\boldsymbol{W}_{wdh}$ that contains flexible controls for temperature (1°C dummy indicators), atmospheric pressure, relative humidity, rainfall, solar radiation (100 W/m^2 dummy indicators), solar radiation dummies interacted with temperature dummies, wind speed, wind direction (8 dummy indicators), and 8th hour lags of temperature and rain. Importantly, these flexible controls reduce biases that arise if cyclists take a detour or a break when the weather is nice. Because we do not observe the exact routes, nor stopping times, we cannot completely rule out this identification threat. However, flexible weather controls and day fixed effects should largely account for this issue.[54]

Air pollution and traffic intensity are in general positively related (Kelly et al., 2011; He et al., 2019). We emphasise that we will show that this does *not* hold for ozone as it is usually formed in the afternoon before the evening rush hour. Nevertheless, to avoid any omitted variable bias due to cyclists slowed down by traffic, we use several traffic controls in $\boldsymbol{T}_{wdh}$. For car traffic, we include road vehicle count data from automatic traffic counters, which we aggregate into a standardized city average. For bicycle traffic, we include the city-wide total distance cycled as control.

[53] For our sample of five years $W \times 7 \times 24 = 5 \times 52 \times 7 \times 24 = 43,800$ in case no observations would be missing. Observations of environmental data are sometimes missing, so that we have just under 40k observations.

[54] When we use a panel dataset which largely consist of commuters (discussed below), detours and stopping are much less likely an issue due to the nature of the trip.

Finally, we take into account that our road traffic controls may not cover all roads typically used by cyclists. Therefore, to further improve the control for car traffic, we also use environmental data, in $\boldsymbol{E}_{wdh}$, as a proxy for traffic intensity and congestion. For this purpose NO_x are especially suitable as they are—in contrast to ozone—directly emitted by combustion engines, and therefore strongly related to traffic intensity (Carslaw and Beevers, 2004; Farias and ApSimon, 2006). Vector $\boldsymbol{E}_{wdh}$ contains SO_2 and flexible controls for NO_x and $PM_{2.5}$ using 10 ppb dummy indicators. These indicators not only serve as additional controls for traffic intensity, but also ensure that β is not capturing effects of other pollutants.[55]

3.3.1.2 Sorting

Ozone forms when there is sufficient heat and solar radiation, and is, therefore, strongly correlated with 'nice weather', which affects the decision to cycle. For example, bicycle commuters are typically less sensitive to weather circumstances than recreational cyclists. Hence, during good weather the total number of cyclists increases, of which a higher proportion is recreational. As a consequence, the composition of cyclists changes (i.e. sorting), which may bias our estimates.[56]

We address a potential bias due to sorting as follows. First, we include a range of time fixed effects. Hour-of-the-week fixed effects control for typical differences in the population of cyclists during a week. Day fixed effects control for unobserved events that cause for a one-day change in the composition of the cycling population (e.g. national holidays or strikes). Second, flexible controls for weather and other pollutants, in $\boldsymbol{W}_{wdh}$ and $\boldsymbol{E}_{wdh}$, account for sorting that can arise if certain types of travellers decide to use a bicycle conditional on weather conditions. These controls include the 8th hour lag of rain and temperature, which account for sorting during afternoons induced by nice weather earlier on the day.

Finally, and importantly, we *test* for sorting by estimating the effect of ozone on the number of bicycle trips and average bicycle-trip distance. Because these regressions show that there is no observable sorting when we exclude weekends, we estimate

[55] Note that if β captures the effect of other pollutants, then this would be detrimental to our conclusion that *ozone* reduces speed, but would not affect our conclusion that air pollution reduces speed.

[56] In general, local air pollution (like particulates or SO_2) is found to directly affect extensive margins (Graff Zivin and Neidell, 2013; Aragón et al., 2017). There is, however, no evidence that this also holds for ozone (e.g. Zivin and Neidell (2012) do not find an effect on the extensive margin of labour supply). It is thus more likely that weather conditions correlated to ozone are the determining factor for sorting.

(3.1) using weekday observations.[57]

We note that our sorting tests may not completely rule out sorting. Even if ozone does not affect the trip distance or the number of cycling trips, one might argue that there may still be compositional changes in the cyclists population, for instance induced by nice weather. This argument is however implausible, as it implies that weather conditions that attract certain cyclists (e.g. recreational), repel other cyclists at the same time (and with the same magnitude). Instead, it seems more plausible that weather conditions favourable to some groups (e.g. commuters), will also be favourable to other groups (e.g. tourists). Nevertheless, we will also include an approach with cyclist fixed effects, which rules out sorting (see Section 3.3.3).

3.3.2 Between-day estimation

One disadvantage of the above approach is that it assumes that the effect of ozone is immediate. We will therefore also follow a strategy comparable to Zivin and Neidell (2012) where we employ daily variables to estimate β as the between-day estimator in the following model:

$$\overline{speed}_d = \beta \overline{ozone}_d + \boldsymbol{\gamma} \overline{\boldsymbol{W}}_d + \boldsymbol{\zeta} \overline{\boldsymbol{T}}_d + \boldsymbol{\delta} \overline{\boldsymbol{E}}_d + \alpha_m + \overline{\eta}_{\hat{d}} + \epsilon_d, \tag{3.2}$$

where a bar indicates a daily average (e.g. $\overline{ozone}_d = \sum_{h=0}^{23} ozone_{wdh}$), and $m = 1, \ldots, 5 \times 12$ and $\hat{d} = 1, \ldots, 7$ denote year×month and day of the week respectively. Here, α_m and $\overline{\eta}_{\hat{d}}$ are year-month and day-of-the-week fixed effects, whereas $\overline{\boldsymbol{W}}_d$, $\overline{\boldsymbol{T}}_d$, and $\overline{\boldsymbol{E}}_d$ refer to daily averages of controls for weather, traffic, and other pollutants (as discussed above). As before, to test for sorting, we use a variant of (3.2) where we estimate the effect of ozone on average cycling distance and on number of bicycle trips per day. We will see that this approach gives similar result as the within-day approach, but tends to give (much) larger standard errors.

3.3.3 Cyclists panel estimations

A remaining disadvantage of the previous approaches is that we can not completely rule out a bias from sorting. To address this, we exploit recurrent patterns in the cycling data to construct a panel of cyclists, which allows us to use cyclists fixed

[57] We find a borderline significant (at the 10 percent level) positive effect of ozone on number trips and distance when including weekends. This makes sense, because during weekends the population consists mainly of recreational cyclists who's decision to cycle is highly dependent on weather conditions. When we include weekends in the analysis of (3.1), the estimates hardly change.

effects.[58] As discussed above, this sample appears to consist mainly of commuters (see Section 3.2.1.1). We consider the following model:

$$speed_{ih} = \beta ozone_h + \boldsymbol{\gamma} \boldsymbol{W}_h + \boldsymbol{\delta} \boldsymbol{E}_h + \boldsymbol{\phi} \boldsymbol{T}_h + \eta_{\hat{d}} + \kappa_i + \varepsilon_{ih}, \tag{3.3}$$

where i refers to a cyclist and h to an hour in the sample. We use the same controls as above and include a day-of-the-week fixed effect, $\eta_{\hat{d}}$ and a cyclist fixed effect, κ_i. As individuals are defined based on the chosen trip (origin and destination) and hour of the day, an alternative interpretation of κ_i is that of a separate fixed effect for each hour-of-the-day for each origin-destination pair.

The main advantage of this approach is that we can rule out sorting issues. One issue—which is a disadvantage if one is interested in cyclists per se—is that this approach yields estimates only valid for this specific group of commuters, who differ from the general population of cyclists, for instance by cycling slightly faster and longer distances (see Section 3.2). Furthermore, commuters usually have an arrival-time target and may therefore increase their speed when they realise that they are late (due to high ozone levels). Hence, the results from this approach may be an underestimate of the effect of ozone on cycling speed for individuals that do not have a specific target. However, this group of commuters may be more representative for workers involved in other work-related outside activities.

3.3.4 Sensitivity analyses and robustness checks

Next to the identification strategies discussed above, we perform various additional analyses to assess the robustness and sensitivity of the results. For the within-day and between-day estimations we perform the following robustness checks. First, we exclude bicycle traffic as a covariate. Next, we check robustness against flexibly controlling for road traffic, by using individual road traffic measurements as separate controls, rather than relying on a city-wide average. To check if outliers drive the results, we also run an estimation where we use the hourly and daily median speed as dependent variable, instead of mean speed.

As another robustness check, we exploit spatial and temporal variation in ozone levels by dividing London into zones, each surrounding an air quality monitoring station (see the map in Figure 3.1). A cycling trip is assigned to a zone if its origin and destination are both within the same zone (see Section 3.2 for details). As in equation

[58]Identifying individual cyclists is possible because routes are very thin, for most of the trips there is no other person cycling the same route within the same hour.

(3.2), we use daily averages.[59] This method should (and does) give similar results to the between-day estimation, but relies on completely different identification, using variation over space rather than over time. See Appendix 3 for further details of the estimation method and the results.

We also provide a range of robustness and sensitivity checks for the cyclists panel estimations. First, we select subsamples where we vary the minimum distance cycled. Second, we put restrictions on the minimum number of trips (i.e. days) per cyclists per month. Third, we include ozone as 5 ppb dummy indicators, so that we can assess whether the assumption of a linear effect holds. This approach also allows us to identify the minimum ozone threshold, i.e. the level of ozone at which speed is affected. Fourth, we examine if there are within-day lag or lead effects of ozone. Fifth, we assess whether the effect of ozone is heterogeneous across cyclists with different fitness levels, as measured by their monthly average speed.

3.4 Results

3.4.1 Results using within-day and between-day estimations

Table 3.2 shows the estimation result using within-day and between-day estimations, specified in equations (3.1) and (3.2). In both specifications we find a negative and statistically significant effect of ozone on cycling speed. Column (1) shows that a 10ppb increase in ozone concentration during a certain hour, reduces speed by 53 metres per hour. At the sample average of 11.97 km/h, this is equivalent to a 0.4% drop in mean speed. Column (2) shows that using between-day rather than within-day estimation, yields a very similar point estimate, despite using a very different identification strategy. It shows that a 10 ppb increase in ozone on a certain day reduces speed by 63 metres per hour on that day.[60]

Columns (3)-(6) show the sorting tests: they highlight that ozone does *not* induce sorting as it does not affect the extensive margins (conditional on controls). Importantly, in these columns, the standard errors of the estimated ozone effects are small, imply-

[59] Horizontal ozone transportation (through air) makes identification using hourly data unreliable. This is because there is measurement error when assigning a trip to an hour of observation. For identification over time this is non-problematic as the same measurement error applies to the environmental variables. However, with identification across zones, horizontal transportation of ozone makes the assignment of trips to ozone levels unreliable due to overlap of trips per hour. At daily averages this is problem disappears because there is no measurement error in the timing of trips within a day.

[60] Note that the coefficients for traffic intensity in column (1) and (2), are not robust to the two different specifications and should not be interpreted as causal.

Table 3.2: Results using within-day and between-day variation.

	Speed		Distance		log(Trips)	
	Within	Between	Within	Between	Within	Between
	(1)	(2)	(3)	(4)	(5)	(6)
Ozone (10ppb)	-0.053***	-0.063***	-0.001	0.009	0.006	0.020
	(0.015)	(0.023)	(0.003)	(0.007)	(0.005)	(0.014)
SO_2 (10ppb)	0.008	0.009	0.003*	-0.004	-0.010***	0.002
	(0.006)	(0.011)	(0.002)	(0.003)	(0.003)	(0.007)
Car traffic	-0.030	0.795***	-0.003	0.095***	0.220***	0.358***
	(0.026)	(0.101)	(0.006)	(0.032)	(0.009)	(0.056)
Bicycle traffic	0.017***	-0.004***				
	(0.003)	(0.0004)				
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
NO_x and $PM_{2.5}$ dummies	Yes	Yes	Yes	Yes	Yes	Yes
Cycling controls	Yes	Yes	No	No	No	No
Day FE	Yes	No	Yes	No	Yes	No
Week hour FE	Yes	No	Yes	No	Yes	No
Month FE	No	Yes	No	Yes	No	Yes
Week day FE	No	Yes	No	Yes	No	Yes
Observations	28,828	1,220	28,828	1,220	28,828	1,220
R^2	0.926	0.903	0.913	0.668	0.955	0.848

Notes: Estimated using weighted least squares, with no. of trips (per hour or per day) as weights for columns (1)-(4). Weather controls: temperature (2.5°C dummy indicators), atmospheric pressure, relative humidity, rainfall, solar radiation (100 W/m^2 dummy indicators), solar radiation dummies interacted with temperature dummies, wind speed, wind direction (8 dummy indicators), 8th hour lag of temperature (1°C dummy indicators), 8th hour lag of rain. PM_{10} and NO_x dummies: 10 ppb dummy indicators. Cycling controls: log(number of trips), average trip distance.***, **, * indicate significance at 1%, 5%, and 10%.

ing that these tests have a lot of power. The results of these sorting tests imply that it is unlikely that our estimates in columns (1) and (2) are affected by ozone-induced sorting of cyclists.

Table 3.B.3 in Appendix 2 presents the results of three alternative specifications for the within-day and between-day estimations. It shows that omitting bicycle traffic as control leads to similar estimates. Similarly, it shows that including flexible controls for car traffic (a separate indicator for each of the eleven counters) hardly changes the results. Furthermore, when we include speed using the median (instead of mean), we again obtain very similar results. The latter highlights that outliers do not drive our results. Finally, also the results of the spatio-temporal estimation, as discussed in

Table 3.3: Estimation results with cyclist fixed effects.

	Speed					
	(1)	(2)	(3)	(4)	(5)	(6)
Ozone (10ppb)	-0.026***	-0.025***	-0.031***	-0.030***	-0.028**	-0.029
	(0.005)	(0.005)	(0.006)	(0.006)	(0.011)	(0.029)
$PM_{2.5}$ (10ppb)	0.007***	0.008***	0.005*	0.004	0.005	-0.011
	(0.002)	(0.002)	(0.003)	(0.003)	(0.006)	(0.014)
SO_2 (10ppb)	0.005*	0.004	0.003	0.006	0.012	0.003
	(0.003)	(0.002)	(0.004)	(0.004)	(0.008)	(0.022)
Car traffic	-0.135***	-0.123***	-0.114***	-0.096***	-0.076***	-0.104**
	(0.010)	(0.010)	(0.012)	(0.013)	(0.021)	(0.047)
Bicycle traffic	-0.004	-0.008	-0.010	-0.010	-0.012	-0.021
	(0.006)	(0.006)	(0.007)	(0.008)	(0.015)	(0.033)
NO_x dummies	Yes	Yes	Yes	Yes	Yes	Yes
Day of the week FE	Yes	Yes	Yes	Yes	Yes	Yes
Cyclist FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Min. obs. per cyclist	4	4	4	8	8	16
Min. distance	0 km	2 km	4 km	4 km	6 km	6 km
Avg. s.d. of departure	4.9 min	4.9 min	4.8 min	5.7 min	5.7 min	5.4 min
Avg. speed	14.2 km/h	14.4 km/h	14.9 km/h	15.0 km/h	16.2 km/h	16.4 km/h
Observations	2,359,675	1,701,521	598,324	409,203	86,379	10,918
R^2	0.752	0.786	0.824	0.820	0.842	0.834

Notes: Standard errors in parentheses are clustered per hour. Weather controls are the same as in Table 3.2. ***, **, * indicate significance at 1%, 5%, and 10%.

Appendix 3, indicate that we find very similar estimates if we identify the effect on within-day variation across zones in the city.

3.4.2 Cyclists panel results

Table 3.3 shows the estimation results using a panel of cyclists who make the same trip (in terms of route and hour of the day) for several days within a month. In column (1), we obtain an effect of -0.026 with a standard error of 0.005. Columns (2)-(5) indicate that the effect is of similar size, or even stronger, when we exclude trips with short distances. There are several explanations for this. For instance, it may be that fatigue increases with distance, or that longer exposure to ozone during trips slightly increases the effect. Columns (4)-(6) indicate that the point estimates hardly change when we focus on cyclists who make more cycling trips per month. Only when we restrict the minimum number of trips to 16, as in column (6), the effect becomes statistically insignificant, as we now have less than 11,000 observations (out of 2.4 million).

Overall, Table 3.3 again shows evidence of a negative effect of ozone on cycling speed, albeit somewhat smaller compared to results without cycling fixed effects. Finding smaller effects makes sense, because this subsample is dominated by morning commuters, who have a preferred arrival time (at work), such that they may exert more effort to compensate for possible delays. We further note that the standard errors are substantially smaller than in previous tables, despite having fewer observations. Clearly, the cyclist fixed effects—which also control for the specific origin-destination pair on a specific hour of the day—vastly improve the efficiency of the estimator.

As a sensitivity check for this approach, we have re-estimated the model for groups of cyclists who differ in their average speed per month (see Table 3.B.5 in Appendix B). This monthly average speed may be interpreted as a proxy for fitness. We find that the ozone effect is very similar, despite substantial differences in average speed. This suggests that cyclists' fitness does not play a major role in the effect of ozone. Note, however, that differences in average speed are not just because of heterogeneity in fitness, but also due to differences in the grade of the road (London is not completely flat), number of traffic lights, direction (there is a dominant south-western wind), etc.

Because we can rule out confounding due to sorting, we can now also interpret estimates of traffic control variables in Table 3.3 as causal, which provides another consistency check.[61] We find that cyclists are slowed down by car traffic, but not by bicycle traffic. Both these findings make sense for London, because there is substantial car traffic, but there is no bicycle congestion, such as observed in Amsterdam and Copenhagen.

The above specifications assume that the marginal effect of ozone is constant, which is in line with a range of other studies (e.g. Folinsbee et al., 1988; Zivin and Neidell, 2012), but potentially restrictive. It also prevents us from identifying a minimum ozone threshold, i.e. the concentration at which an effect of ozone is observable. To allow for non-linear effects and to identify the minimum ozone threshold, we use 5 ppb ozone indicator dummies. The results in Figure 3.7 show that a linearly downward sloping effect would fit within the confidence intervals. This implies that we

[61] For particulate matter ($PM_{2.5}$), we find a very small (about one fifth of the ozone effect) but positive effect on cycling speed. This finding is likely spurious as it is not robust across sub samples, but it might be that because PM pollution is not odourless, cyclists increase their speed to reduce the duration of exposure. We do not find any effect of SO_2 (which has a strong odour), in line with its very low levels in Inner London. High SO_2 levels are mainly found around coal fired power plants, that are absent in Inner London.

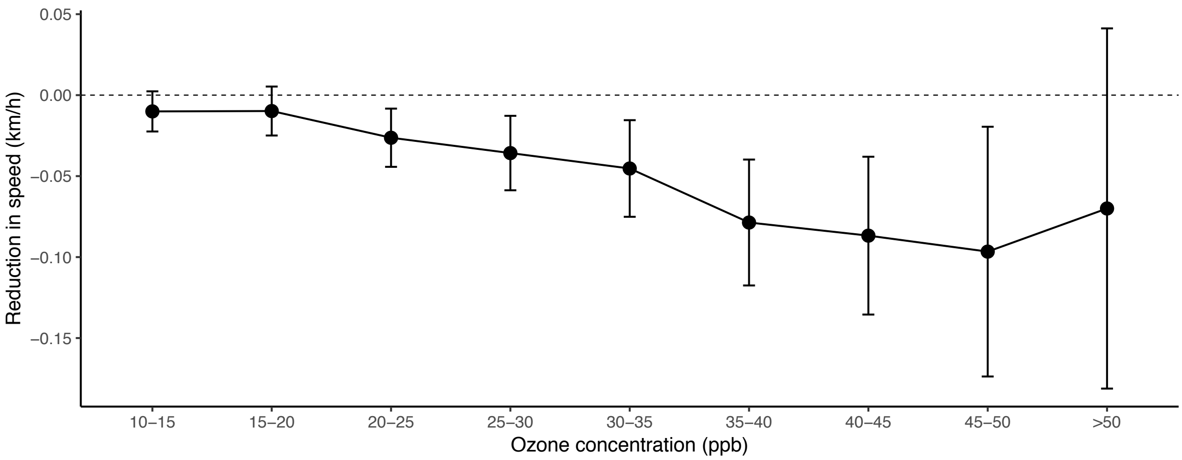


Figure 3.7: Effect of ozone on speed relative to concentrations between 0 and 10 ppb.

Notes: Error bars denote 95% confidence intervals. Ozone concentrations between 0 and 10 ppb are set as the reference category. Other than the ozone indicators, the specification is the same as in column (1) of Table 3.3.

cannot reject the null hypothesis of a constant marginal effect of ozone on speed.[62] We note that a linear effect is convenient for policy: it allows for straightforward environmental regulation based on constant marginal cost pricing and daily averages, and may ignore within-day variation in ozone.[63]

An important difference with previous literature is that we have a much larger sample. As a consequence, we are able to show a negative impact of ozone starting at concentrations as low as 20 ppb (see Figure 3.7), which is half the value reported by Zivin and Neidell (2012). This is particularly relevant because ozone levels in London are much lower than in California; the Zivin and Neidell (2012) results imply that for London ozone effects are largely absent.

So far, we have assumed that the ozone effect is contemporaneous. We test this assumption using a specification test which assesses whether the effect on speed is more strongly caused by hourly lags or leads than contemporaneous ozone. The results are presented in Figure 3.8. In terms of point estimates, contemporaneous ozone has the strongest effect. This makes it plausible that our previous results are due to an immediate effect of ozone. For leads, as shown in the figure, the first two hours hour have a statistically significant effect. This effect is not causal, but due to strong hourly

[62] We find similar results if we do this exercise for the between-day approach. The other approaches have too large standard errors to examine non-linearity.
[63] Note that regulation on daily averages still ignores hourly differences in exposure.

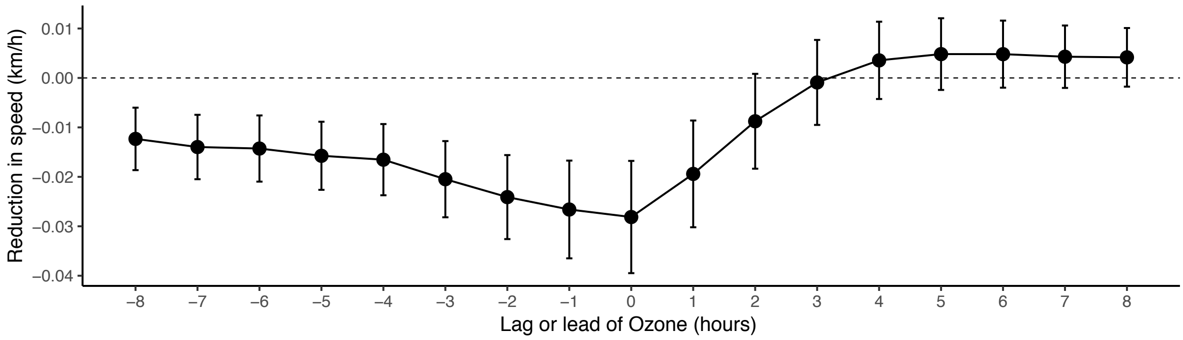


Figure 3.8: Regressions results for testing for lag or lead effects.

Notes: Error bars denote 95% confidence intervals.

autocorrelation in ozone levels (see Table 3.A.2 in Appendix A). For leads longer than two hours the effect disappears. For lags, we find statistically significant effects up to the 8th hour. Because the point estimates for lags are stronger than for leads, one may believe that there is a lagged effect of ozone on cycling speed *additional* to a contemporaneous effect. However, further analysis, where we include both lagged and contemporaneous ozone, shows that for all lag lengths the sum of the effects of lagged and contemporaneous ozone is exactly equal to the estimated contemporaneous effect as shown in column (4) of Table 3.3. Hence, there is no evidence that the contemporaneous effect underestimates the overall effect.

3.5 Conclusion

We provide novel evidence of the immediate negative impact of urban air pollution. We focus on the effect of ozone, which is known to reduce lung capacity. We hypothesise that ozone slows down all outside activities which require a minimal physical effort. We focus on cycling in London and demonstrate that the speed of cyclists is reduced due to ozone pollution. Our results imply that a 10 ppb increase in ozone induces an immediate 0.3% to 0.4% reduction in cycling speed. Our most important finding for policy on air quality is that that ozone pollution reduces speed for concentrations above levels as low as 20 ppb (observed 36% of the time in London). This is *far* below the minimum threshold found in previous field studies, and also far below current air quality standards.

Our effects are of the same magnitude as the ones found in the lab by Gong et al. (1986). In contrast, we find much smaller point estimates than Zivin and Neidell

(2012), who analyse the effect of ozone on productivity of agricultural workers in California. A plausible explanation for this difference is that, in contrast to these workers, cyclists in London are not exposed for a full working day, and exert physical effort for a much shorter amount of time.[64] An alternative explanation is purely statistical: our point estimate is just within the 95 percent confidence interval reported by Zivin and Neidell (2012), so that their estimates may not be so different from ours.

Because cycling is a light physical activity, which is usually done for a short period of time (about 20 minutes in our sample), we argue that the effect identified here plausibly carries over to most other outside activities involving physical effort, such as walking. For London, with ozone levels usually below 50 ppb, our results imply a moderate impact on physical activity, with a total negative effect not exceeding 2%. However, the marginal effects are sizeable, which becomes important when extrapolating to cities with higher ozone levels. For example, in Chinese cities, ozone levels have increased over the past decades (Verstraeten et al., 2015) and often exceed 100 ppb (Wang et al., 2017). If we assume a linear effect—in line with our results in Figure 3.7, but also implied by combining the estimates from our study and laboratory studies—then speed is reduced by more than 5% due to ozone. If such cities would improve air quality to daily concentrations found in London, then cycling speed would increase by about 4%. Our results therefore indicate that, besides the well-known health hazards, urban air pollution has a sizeable immediate negative effect on travel time and other activities that require physical effort.

[64] An additional explanation is that, as shown by Gong et al. (1986), ozone reduces lung capacity especially in hot climates. In contrast to California, London has a relatively cool climate and temperatures rarely exceed 30°C (less than 0.5% of the time). To test for this we ran a sensitivity check with temperature interacted with ozone, but this does not yield results different from Table 3.2. In addition, estimations using only 'hot' days (e.g. more than 20°C) results in similar results, but with (much) larger standard errors.

Appendix 3.A Additional descriptives

3.A.1 Map with radiation monitors and sites names

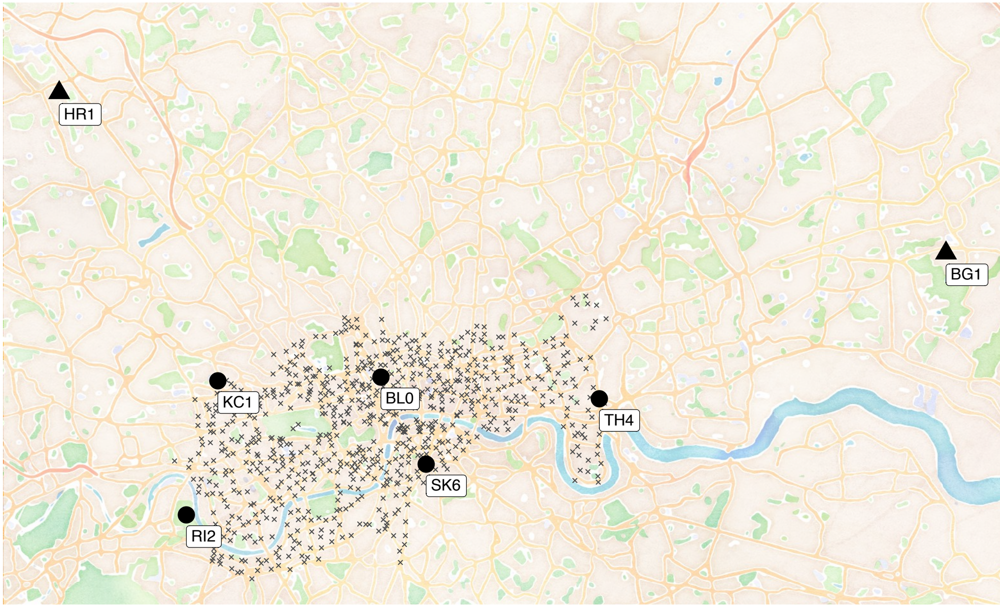


Figure 3.A.1: Location of docking stations (×), air quality monitors (●), and radiation monitors (▲).

Notes: Map tile by Stamen Design, under CC BY 3.0, base layer data by OpenStreetMap, under CC BY SA.

3.A.2 Spatial and temporal autocorrelation of ozone

Table 3.A.1: Spatial correlation of ozone.

	BL0	KC1	RI2	SK6	TH4
BL0	1.00	0.94	0.92	0.93	0.82
KC1	0.94	1.00	0.96	0.91	0.87
RI2	0.92	0.96	1.00	0.90	0.89
SK6	0.93	0.91	0.90	1.00	0.80
TH4	0.82	0.87	0.89	0.80	1.00

Table 3.A.2: Autocorrelation of ozone per hour (a) and per day (b).

	t=0	t=-1	t=-2	t=-3	t=-4	t=-5	t=-6	t=-7	t=-8
cor	1.00	0.96	0.89	0.80	0.71	0.63	0.56	0.50	0.46

(a)

	t=0	t=-1	t=-2	t=-3	t=-4	t=-5	t=-6	t=-7	t=-8
cor	1.00	0.73	0.54	0.48	0.45	0.42	0.41	0.41	0.37

(b)

3.A.3 Daily descriptives

Table 3.A.4: Descriptive statistics per day.

Statistic	Mean	St Dev.	Min	Max
Number of trips	25559	10642	3159	111694
Speed (km/h)	11.94	1.33	6.43	14.10
Duration (min)	18.41	4.39	11.99	42.60
Distance (km)	2.73	0.13	2.19	3.74
Bicycle traffic (km)	82688	31300	7373	406608
Road traffic (vehicles per counter)	491	43	279	663
Ozone (O_3, ppb)	18.21	9.02	0.33	48.29
Nitrogen oxides (NO_x, ppb)	43.95	29.03	7.14	307.67
Particulate matter ($PM_{2.5}$, ppb))	13.04	9.29	2.37	84.96
Sulphur dioxide (SO_2, ppb)	10.59	7.89	1.35	171.25

Note: 1,826 daily observations, cycle variables are city wide daily means of 70.26 million observations. Mean and standard deviation are weighted for the number of trips per day.

3.A.4 Flows on inferred routes

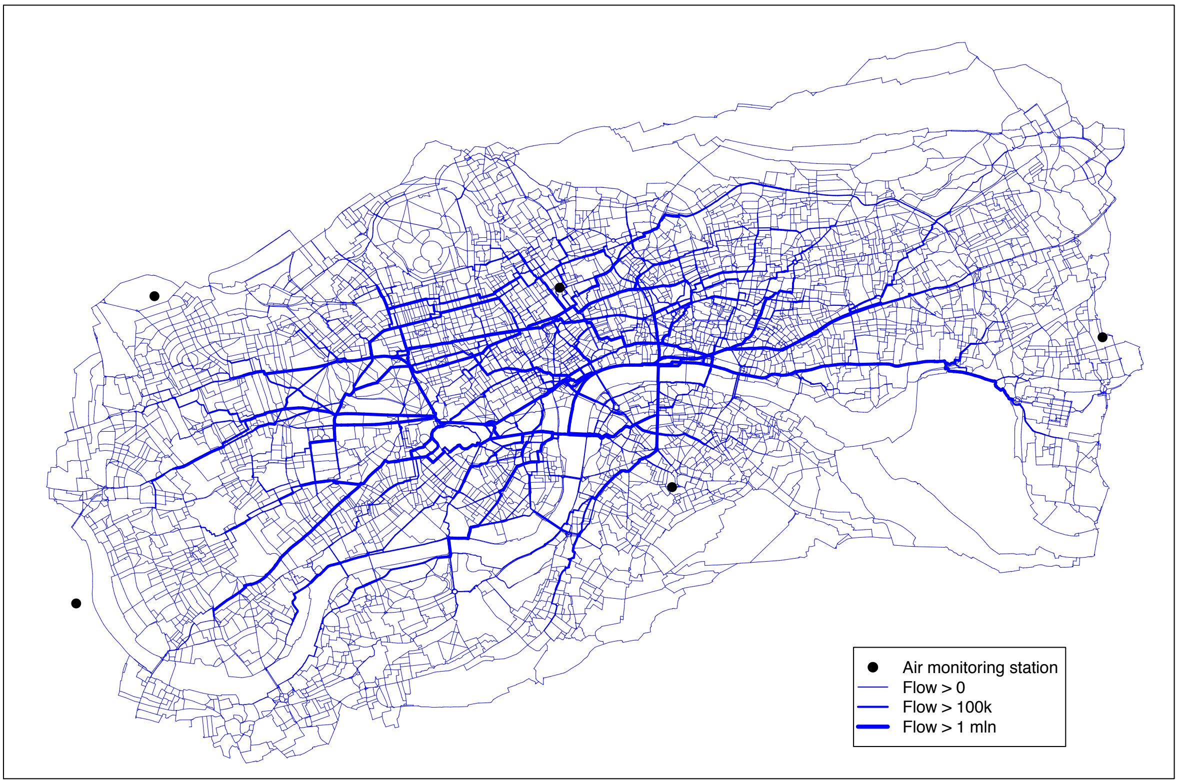


Figure 3.A.2: Map with flows and location of air quality monitoring stations.

3.A.5 Panel characteristics

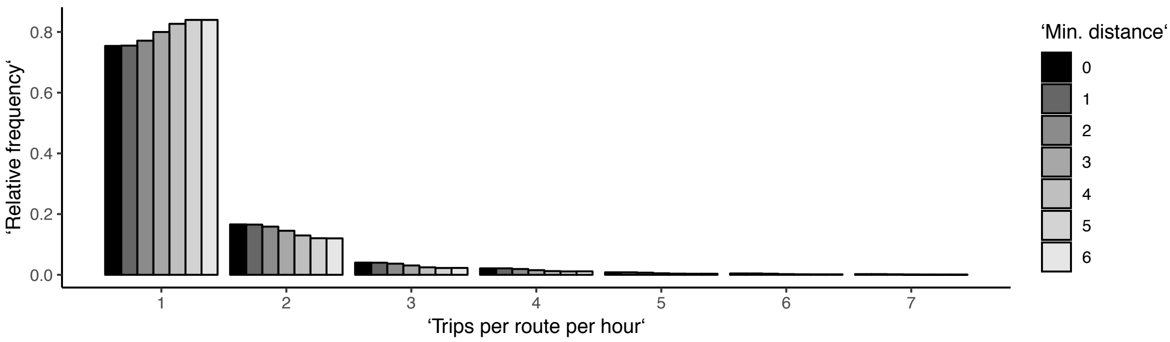


Figure 3.A.3: Histogram of number of trips per route per hour for all observations.

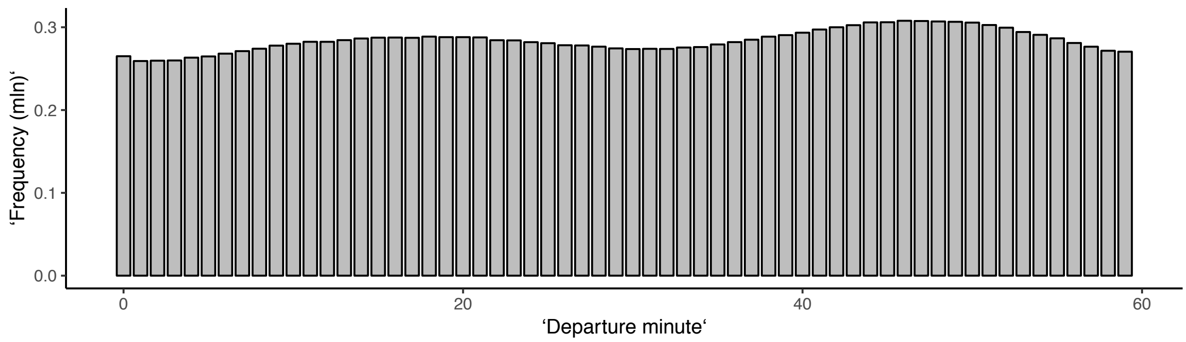


Figure 3.A.4: Histogram of departure time of panel observations.

Appendix 3.B Additional analyses

3.B.1 Traffic and air pollution

Table 3.B.1: Results of the effect of traffic on air pollution using within-day variation.

	Ozone				NOx			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Car Traffic	0.198***	-0.131***	0.007	0.038*	1.185***	1.081***	0.664***	0.707***
	(0.015)	(0.022)	(0.020)	(0.022)	(0.025)	(0.076)	(0.064)	(0.103)
NO_x (10ppb)			-0.108***					
			(0.002)					
$PM_{2.5}$ (10ppb)			-0.160***				1.690***	
			(0.007)				(0.020)	
SO_2 (10ppb)			0.111***	-0.009			0.893***	0.908**
			(0.006)	(0.008)			(0.018)	(0.381)
Baseline ctrls.	No	No	No	Yes	No	No	No	Yes
Day FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Week-hour FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	31,295	31,295	31,173	30,178	31,295	31,295	31,173	30,178
R^2	0.006	0.778	0.819	0.925	0.069	0.736	0.812	0.852

Notes: Baseline controls as in Table 3.2.***, **, * indicate significance at 1%, 5%, and 10%.

We explore the relation between car traffic and air pollution by estimating the within-day and between-day model, as in equations (3.1) and (3.2), with ozone and NO_x as dependent variables. The within-day results in Table 3.B.1 (columns (1) and (5)), show that, without controls, car traffic correlates positively with both ozone and NO_x, but much stronger with the latter (as evidenced by a larger coefficient, and a much higher R^2). Adding day fixed effects and hour-of-the-week fixed effects (columns (2) and (6)), shows that within a day, ozone has a weak negative, and NO_x still a strong positive relation with car traffic. In column (3) and (7) we add controls for other pollutants, this nullifies the effect of car traffic on ozone, while its effect on NO_x stays roughly the same. Finally, for columns (4) and (8), we include the same controls as in Table 3.2. This specification now includes flexible controls for other pollutants and weather. Again we find that ozone is unrelated to traffic. In contrast, car traffic has a strong—and arguably causal—causal effect on NO_x. These results are in line with the notion that NO_x is directly emitted by cars' combustion engine, while ozone is a second-order pollutant whose concentrations are only related to traffic *through* NO_x.

Table 3.B.2: Regression results of the effect of traffic on air pollution using between-day model.

	Ozone				NO_x			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Car Traffic	-0.303***	-0.579***	-0.234**	-0.146*	2.775***	2.461***	1.837***	1.291***
	(0.104)	(0.116)	(0.099)	(0.077)	(0.371)	(0.424)	(0.337)	(0.347)
NO_x (10ppb)			-0.130***					
			(0.007)					
$PM_{2.5}$ (10ppb)			-0.093***				1.400***	
			(0.019)				(0.055)	
SO_2 (10ppb)			0.088***	-0.015			0.868***	0.840***
			(0.023)	(0.015)			(0.075)	(0.066)
Baseline ctrls.	No	No	No	Yes	No	No	No	Yes
Month FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Day of W. FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	1,826	1,826	1,823	1,797	1,298	1,826	1,823	1,797
R^2	0.005	0.519	0.652	0.883	0.041	0.368	0.604	0.760

Notes: Baseline controls: same controls as in Table 3.2.***, **, * indicate significance at 1%, 5%, and 10%.

Table 3.B.2 shows the same effect of car traffic on NO_x using between-day variation, see columns (5)-(8). Columns (1)-(3) show that between days, traffic is negatively related to ozone. This is mainly driven by not controlling for weather conditions, which correlate with car traffic and ozone. Therefore, adding controls for weather, as in column (4), leads to an insignificant effect of car traffic on ozone.

3.B.2 Further robustness checks for within-day and between-day estimations

Table 3.B.3: Robustness checks for within-day and between-day estimations.

	Average Speed				Median Speed	
	Within	Between	Within	Between	Within	Between
	(1)	(2)	(3)	(4)	(5)	(6)
Ozone (10ppb)	-0.042***	-0.069***	-0.038***	-0.045**	-0.030**	-0.048*
	(0.013)	(0.023)	(0.012)	(0.020)	(0.012)	(0.025)
SO_2 (10ppb)	0.008	0.004	0.007*	0.0005	0.006	0.002
	(0.005)	(0.013)	(0.004)	(0.011)	(0.005)	(0.015)
Car traffic	-0.209***	0.373***			-0.185***	0.462***
	(0.027)	(0.085)			(0.032)	(0.094)
Bicycle traffic			0.018***	-0.015***	0.009**	-0.017***
			(0.004)	(0.001)	(0.004)	(0.001)
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
NO_x and $PM_{2.5}$ dummies	Yes	Yes	Yes	Yes	Yes	Yes
Sep. traf. controls	No	Yes	No	No	Yes	No
Cycling controls	Yes	Yes	No	No	No	No
Day FE	Yes	Yes	Yes	No	No	No
Week hour FE	Yes	Yes	Yes	No	No	No
Month FE	No	No	No	Yes	Yes	Yes
Week day FE	No	No	No	Yes	Yes	Yes
Observations	28,760	1,699	28,760	1,699	28,760	1,699
R^2	0.925	0.948	0.931	0.961	0.906	0.951

Notes: Estimation using weighted least squares, with number of trips per hour or per day as weights. The same controls as in Table 3.2 are included. ***, **, * indicate significance at 1%, 5%, and 10%.

3.B.3 Spatio-temporal estimation

For the robustness check exploiting spatio-temporal variation, we consider the following two-way fixed effects model:

$$\overline{speed}_{zd} = \beta\overline{ozone}_{zd} + \boldsymbol{\delta}\overline{\boldsymbol{E}}_{zd} + \boldsymbol{\zeta}\overline{\boldsymbol{T}}_{zd} + \kappa_d + \lambda_z + \overline{\eta}_{zt} + \overline{\varepsilon}_{zd}, \tag{3.B.1}$$

where $z = 1, \ldots, Z$ denotes a zone, and t denotes either a week or a month.[65] Here, κ_d refers to a day fixed effect, and λ_z to a zone fixed effect. To control for sorting, we also include $\overline{\eta}_{zt}$, a zone-time fixed effect, which accounts for unobserved patterns across time and zones. We estimate two specifications: one with zone-month fixed-effects, and another with zone-week fixed effects. These controls account for temporary local changes (e.g. local events, changes in infrastructure etc.). As before, we test for sorting using distance and number of trips per zone per day as dependent variable.

Table 3.B.4: Results using spatio-temporal variation.

	Speed		Distance		log(Trips)	
	(1)	(2)	(3)	(4)	(5)	(6)
Ozone (10ppb)	-0.050**	-0.030	0.002	0.001	0.007	0.004
	(0.023)	(0.028)	(0.003)	(0.004)	(0.005)	(0.006)
Car traffic	0.007	0.010	-0.005	-0.005	-0.036***	-0.031***
	(0.036)	(0.044)	(0.005)	(0.006)	(0.008)	(0.010)
Bicycle traffic	-0.062***	-0.061***				
	(0.011)	(0.013)				
NOx dummies	Yes	Yes	Yes	Yes	Yes	Yes
Zone FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Zone-Month FE	Yes	No	Yes	No	Yes	No
Zone-Week FE	No	Yes	No	Yes	No	Yes
Observations	6,079	6,079	6,079	6,079	6,079	6,079
R^2	0.931	0.948	0.968	0.976	0.997	0.998

Notes: Estimated using observations from weekdays. ***, **, * indicate significance at 1%, 5%, and 10%.

Table 3.B.4 shows the results for two specifications using spatio-temporal variation.

[65]In this specification, we include NO_x, but not $PM_{2.5}$ and SO_2 because at the zonal level there are substantial amounts of data missing in the measurements of $PM_{2.5}$ and SO_2. As other specifications do not show that controlling for these indicators is relevant, this is a non-problematic issue. Similarly, we do not include weather controls here, as we only observe weather at the city level. Arguably, the differences in weather conditions within a day do not vary much across zones in the city.

In the first specification we use zone-month fixed effects, whereas in the second we use zone-week fixed effects. Again, for both specifications, the effect of ozone on cycling speed is negative, but only statistically significant in the first one. These effect sizes are similar, but with larger standard errors, to those in Table 3.2. Columns (3)-(6) indicate that ozone does not induce bicyclist sorting.[66]

[66]Recall that we exclude weekend days. When we include weekend days, the ozone effect is similar but somewhat stronger, which may be due to sorting, because the sorting tests show then that ozone is weakly positively related with cycling distance and number of trips.

3.B.4 Heterogeneity in panel estimations

Table 3.B.5: Results with cyclist fixed effects varying by average speed.

	Speed					
	(1)	(2)	(3)	(4)	(5)	(6)
Ozone (10ppb)	-0.030	-0.021***	-0.023***	-0.029***	-0.026***	-0.028*
	(0.018)	(0.008)	(0.006)	(0.006)	(0.009)	(0.015)
$PM_{2.5}$ (10ppb)	-0.012	0.0003	0.006**	0.004	0.012**	0.029***
	(0.009)	(0.004)	(0.003)	(0.003)	(0.005)	(0.008)
SO_2 (10ppb)	-0.009	0.009**	0.008**	0.004	0.007	-0.004
	(0.009)	(0.004)	(0.004)	(0.005)	(0.006)	(0.012)
Car traffic	-0.022	-0.124***	-0.129***	-0.134***	-0.147***	-0.184***
	(0.040)	(0.017)	(0.014)	(0.013)	(0.019)	(0.027)
Bicycle traffic	-0.060**	-0.015	-0.009	0.003	0.008	0.018
	(0.025)	(0.010)	(0.008)	(0.009)	(0.013)	(0.021)
NO_x dummies	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Cyclist FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Avg. s.d. dep.	4.4 min	4.6 min	4.7 min	4.6 min	4.5 min	4.4 min
Speed (km/h)	$<$10 km/h	10–12 km/h	12–14 km/h	14–16 km/h	16–18 km/h	$>$18 km/h
Avg. spd. (km/h)	8.8 km/h	11.2 km/h	13.0 km/h	14.9 km/h	16.9 km/h	20.0 km/h
Observations	82,355	359,242	728,453	655,705	325,967	208,001
R^2	0.437	0.143	0.149	0.132	0.107	0.430

Notes: Standard errors in parentheses are clustered per hour. Weather controls are the same as in Table 3.2. ***, **, * indicate significance at 1%, 5%, and 10%.

4
Do metro interruptions increase the demand for public rental bicycles? Evidence from Paris

4.1 Introduction

Urban policy makers in many cities are promoting cycling as part of a sustainable and reliable transport system. Cycling has been linked to several positive urban outcomes. Cycling is associated with health benefits (De Hartog et al., 2010; Celis-Morales et al., 2017), does not generate pollution (Gössling and Choi, 2015), improves connectivity (Bullock et al., 2017), and bicycles make more efficient use of road capacity compared to cars (Wang et al., 2008). One of the ways through which cities encourage cycling is by providing a public bicycle system (PBS).[67] While the first PBS

This chapter is based on Klingen (2019) as published in *Transportation Research Part A*. I am grateful to Jos van Ommeren, Erik Verhoef, Paul Koster, Erik Plug, Stuart Donovan, Jiska Klein, Simon Mayer, Gerben de Jong, and Devi Brands for their comments, to Etienne Côme from IFSTTAR for guiding me to the Vélib' data, and to three anonymous referees who provided very useful comments and suggestions on a previous version of this paper. Map tile Fig 4.1 by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under CC BY SA.

[67] Here, *public* refers to the fact that the bicycles are readily and flexibly available in a city, this is irrespective of whether the operator is a private or public organisation. In Paris, the context studied here, the PBS was initiated by the mayor but a concession won by private operator JCDecaux.

was introduced in Amsterdam in 1965, the launch of Lyon's *Velo'v* system in 2005 marks a turning point in the deployment of large scale urban bicycle systems (DeMaio, 2009). Today, more than thousand public bicycle schemes worldwide host over two million rental bicycles (Metro bike, 2017).

The main goal of this paper is to examine the role of a PBS in relation to public transport, by analysing the effect of metro interruptions on the local demand for public rental bicycles within the Vélib' sharing scheme in Paris. Thereby I will use Twitter accounts of the metro operator to construct a dataset of metro interruptions. Focusing on Paris for this study offers, more than any other city with a PBS, the possibility to analyse net substitution of transport modes within an urban system as a whole. This is because of all PBS, Velib' is the most comprehensive system, spanning both central and more peripheral parts of the city. In contrast, other cities' PBS often only cover the inner city (e.g. London), or have disconnected clusters of docking stations (e.g. Los Angeles).

There is a wide literature on vulnerability of transport networks, often defined as "...a susceptibility to incidents that can result in considerable reductions in [...] network serviceability" (Berdica, 2002, p119). A common method to assess the vulnerability of a transport network is to analyse the impact of removing individual links (Mattsson and Jenelius, 2015). Using this method, von Ferber et al. (2012) find that the structure of the Parisian public transport network is less vulnerable compared to London's system. When certain routes are interrupted, the load transferred to remaining links is better absorbed in Paris compared to London. As a result, it can handle more interruptions before overall capacity is exceeded. Similarly, Cox et al. (2011) find that during a city-wide public transport interruption in London, cycling was used to absorb excess travel demand. This study relates to this literature by analysing to what extent cycling is used to absorb capacity constraints following local metro interruptions. However, instead of looking at a city-wide shut down, I analyse local interruptions which enables us to study the effect of removing only part of the transport network.

As a by-product, I can identify whether cycling in general complements metro service or serves as a substitute. The literature on this subject is still in development. Based on surveys, Shaheen (2012) finds that people in bigger cities who start using rental bicycles take fewer (light)rail trips:

> *In larger cities, bicyclesharing appears to draw from public transit use, freeing up capacity and perhaps serving as a faster connection to intraurban locations than previously provided by bus and rail systems* (Shaheen, 2012, p3).

Looking instead at trains rather than the metro, previous studies suggest that cycling is more of a complement than a substitute. Kager et al. (2016) argue that a bicycle-train connection benefits from complementarity between speed and accessibility, while Rietveld (2000) finds that train station's market potential strongly depends on surrounding cycling infrastructure. To my best knowledge, empirical studies on the interaction between public transport and cycling are based on survey data only, and therefore understanding of this interaction can be further improved by analysing data derived from the use of a PBS.

The introduction of modern PBS with electronic docking stations creates opportunities for further analysis of urban cycling due to the large amount of data that is produced and collected. Using these data, several studies investigate the (local) demand for cycling. One branch of literature focuses on the operational side of a PBS. Han et al. (2014) and Côme and Oukhellou (2014) develop models to predict local demand at station level, based on network and usage characteristics of the Vélib' PBS. Using similar data for London, Adham and Bentley (2015) provide a method to optimize the redistribution of bicycles over the network, by optimally addressing the spatial variation in local demand over time. Another branch of literature focuses on the type of users of a PBS. Bordagaray et al. (2016) and Martin-Moral and Fonzone (2017) develop tools to classify the type of rental bicycle usage using cluster analysis based on trip characteristics in the London PBS. Finally, Wang et al. (2015) assess determinants of local demand for rental bicycles in the Minneapolis–St. Paul Metropolitan Area in Minnesota and find positive correlation between rental bicycle usage and, among others, socio-economic characteristics of the neighbourhood, proximity to a central business district, and local economic activity in general.

In this study I try to bridge the survey based results on interaction between a PBS and public transport on one hand, with the studies based on PBS data on the other hand. To that end, I link usage data of the Parisian PBS Vélib' with data on metro interruptions that I construct using the transport operator's Twitter accounts. The metro interruptions serve as quasi-natural experiments to measure the effect of negative shocks in metro supply on the demand for rental bicycles.

The rest of this paper is structured as follows: Section 4.2 gives the theoretical framework; Section 4.3 describes the data collection and gives descriptive statistics; Section 4.4 presents the empirical methods; Section 4.5 discusses the results and implications; Section 4.6 concludes.

4.2 Theoretical framework

For this study I use a simple partial equilibrium framework to analyse the demand and supply of rental bicycles. Transport modes can be substitutes or complements, and in a model with two markets this translates into a change in demand in one market induced by a price change in the other market. For this study, I analyse how demand for rental bicycles is affected by a negative supply shock in the market for metro trips. Demand is studied at the level of a bicycle docking station (hereafter referred to as 'docking station'), denoted by i, and during a 20 minute time window, denoted by t. Rental bicycles at time-location pair i, t are modelled as distinct commodities, such that demand captures the willingness to pay for bicycles during one period. In this way I have a one-to-one connection with the empirical results. Demand for rental bicycles is described by $q^d_{i,t} = \tilde{D}(p_{i,t})$, where $q^d_{i,t}$ denotes the number of newly demanded bicycles at price $p_{i,t}$ in the 20 minutes preceding t. Demand is assumed to be downward sloping, i.e. $\tilde{D}'(p) < 0$. Supply is piece-wise constant and (fully) represented by $\overline{q}_{i,t}$, the number of bicycles available at a docking station. Supply is fully elastic for quantities within the available stock, and fully inelastic when the stock of bicycles at a station is depleted. This gives the following price scheme in equilibrium:

$$p_{i,t} = \begin{cases} c & \text{if } q^\star_{i,t} \leq \overline{q}_{i,t}, \\ \infty & \text{if } q^\star_{i,t} > \overline{q}_{i,t}, \end{cases} \tag{4.1}$$

where c is the rental price.

Here, I assume that metro interruptions can induce a change in the demand for rental bicycles at docking stations in the neighbourhood of affected metro stations. Let $q^\star_{i,t}$ and $q^\#_{i,t}$ denote the quantity of bicycles that are rented in a normal situation and during a metro interruption nearby respectively. If the quantity of bicycles rented changes from $q^\star_{i,t}$ to $q^\#_{i,t}$ due to a metro interruption, then the sign of $\beta_{i,t} := q^\#_{i,t} - q^\star_{i,t}$ indicates if bicycles are a net substitute or a net complement to metro trips, provided $q^\star_{i,t} < \overline{q}_{i,t}$ and $q^\#_{i,t} < \overline{q}_{i,t}$. If $\beta_{i,t} > 0$ then rental bicycles are a net substitute for metro trips, and if $\beta_{i,t} < 0$ then they are a net complement. The empirical part of this paper deals with estimating β, the average observed change in the number of rental bicycles consumed, that can be ascribed to a ceteris paribus change in the generalized price of metro trips. Hence, the estimate of β identifies the shift in demand, such that I can empirically test whether rental bicycles are a net substitute or a net complement to metro trips.

4.3 Data and descriptive statistics

For this study I observe data from Paris for one year, from July 2016 to June 2017. The map in Figure 4.1 shows that the infrastructure of the Vélib' PBS and the metro network are both dense and cover a similar area in the city.

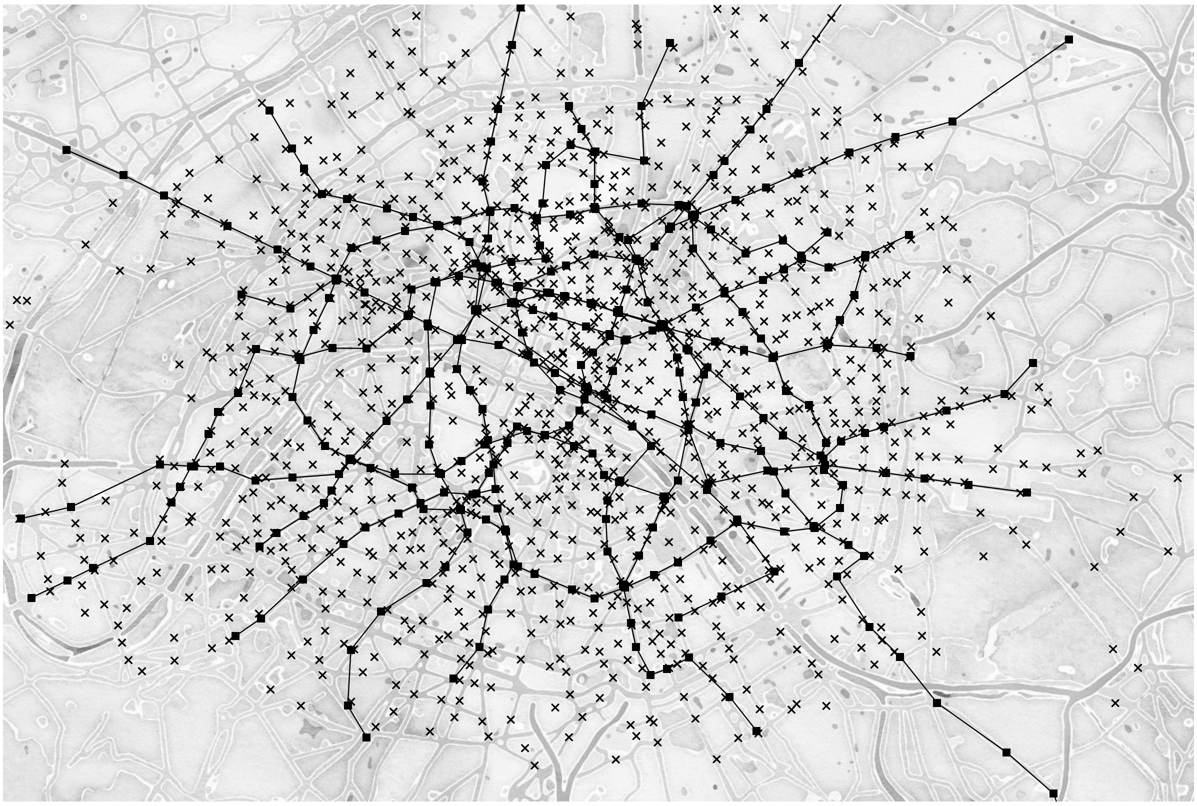


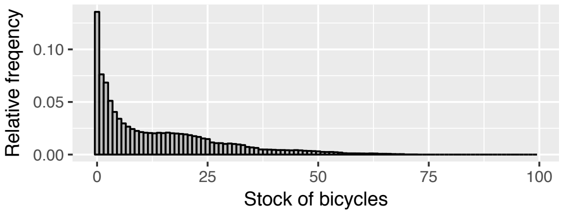
Figure 4.1: Map of Paris with Vélib' docking stations ($\times$) and metro stations ($\blacksquare$).

4.3.1 Vélib' public bicycle system

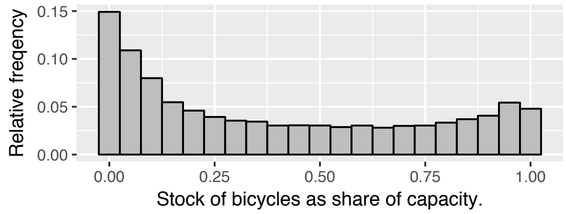
The Vélib' PBS has roughly 20,000 bicycles and 1,225 docking stations located across Paris. It was launched in 2007 and, at that time, Vélib' was the largest PBS in the world. The goal was to provide bicycles within 300 metres from every location within Paris in order to promote cycling (JCDecaux, 2008).

Prices are €1.70 for a day, €8.00 for a week, or €29 per year, with discounts for students and adolescents. Within these subscriptions, usage is free of charge for the first 30 minutes. Exceeding the free usage period is priced €1 for the first hour, €2 for the second, and €4 per hour for any subsequent hours (Vélib', 2017).

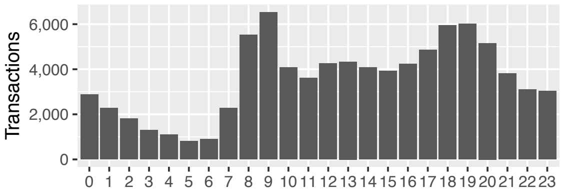
Marie de Paris (2016) estimates that roughly 5% percent of within-city trips in Paris are made by cyclists. Of all bicycle trips, Vélib' bicycles are used 35% of the time, which amounts to just over 100,000 trips per day. Trips take on average about 20 minutes—such that for the majority of the trips the price is zero —and 30% of trips are estimated to be commutes.



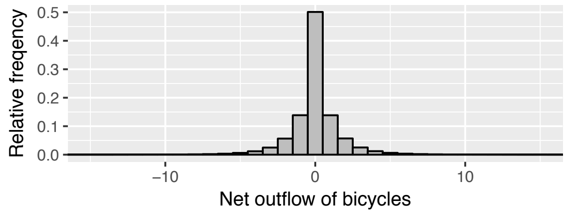
(a) Histogram of stock.



(b) Histogram of stock relative to capacity.



(c) No. of transactions during the day.



(d) Histogram of net outflow of bicyles per 20 min.

Figure 4.2: Stocks and transactions of bicycles at docking stations.

4.3.1.1 Data

I observe historical series from docking station occupancy as reported every 20 minutes by the operator JCDecaux, and collected by IFSTTAR (2017).[68] The histogram in Figure 4.2a shows that docking stations are empty 13% of the time and Figure 4.2b indicates that docking stations are more often empty than fully occupied (10% of time). The average daily pattern of all transactions over the full sample shown in Figure 4.2c shows two local maxima at peak hours, suggesting usage by commuters.

Using the definitions from Section 4.2, the Vélib' data gives us $\overline{q}_{i,t}$, the stock of bicycles at docking station i at time t. However, for this study I am ultimately interested in the gross outflow $q_{i,t}$, i.e. the number of bicycles that leave docking station i during the interval $[t-1,t]$. To get a proxy for the outflow, I first calculate the change in stock at each docking station, which gives the net inflow: $\Delta\overline{q}_{i,t} := \overline{q}_{i,t} - \overline{q}_{i,t-1}$. Then I take the negative value to get the net outflow that I define as

$$\tilde{q}_{i,t} := -\Delta\overline{q}_{i,t} = q_{i,t} - \hat{q}_{i,t}, \tag{4.2}$$

where $q_{i,t}$ denotes the gross outflow and $\hat{q}_{i,t}$ denotes the gross inflow, both at docking station i in interval $[t-1,t]$. Figure 4.2d depicts the histogram for this indicator and shows that during most time periods there is only a small change in the stock of bicycles, if any at all. For the regression analyses, I use the net outflow $\tilde{q}_{i,t}$ as a proxy

[68] The data is available at: `vlsstats.ifsttar.fr/rawdata`

for gross outflow $q_{i,t}$, which implies that I treat $\hat{q}_{i,t}$ as measurement error. This seems reasonable, as most interruptions considered in this paper are short, such that is not likely that inflow is affected, especially not in the first 20 minutes of an interruption. To what extend this affects my results is discussed in greater detail in section 4.4.

4.3.2 Metro interruptions

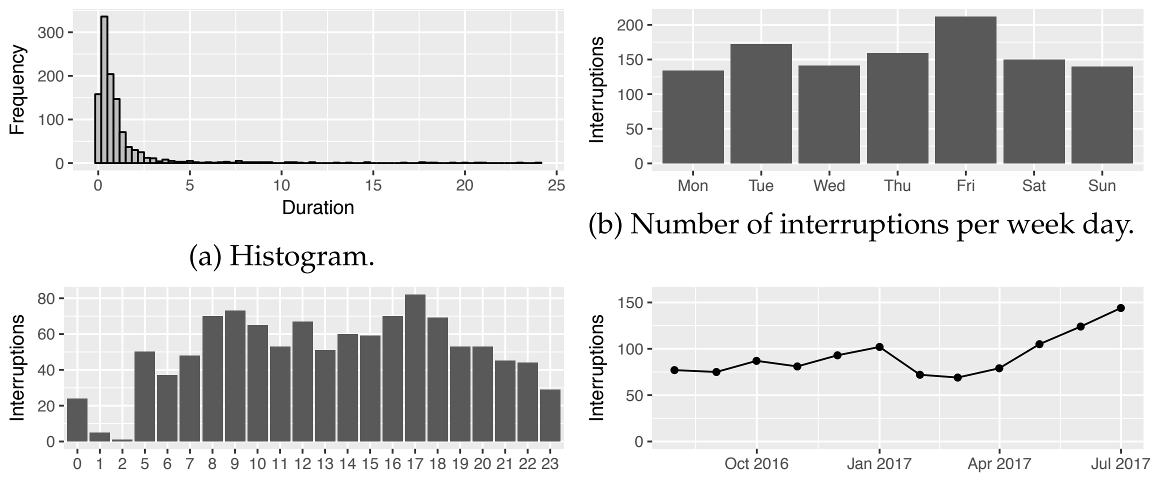
One of the contributions of this study is the construction of a unique data set based on Twitter messages sent by Régie Autonome des Transports Parisiens (RATP), the Parisian metro operator. Each of the fourteen metro lines has a separate Twitter account that is used for communication with travellers and provides information on interruptions.[69] When a metro line is interrupted, RATP sends out a *tweet*: a short, publicly available, online text message containing a maximum of 140 characters. In these messages the operator announces metro interruptions, and also states which part of the line is affected. Some interruptions affect only one station (e.g. when a station is closed for a security check), while other interruptions affect multiple stations or cause a full metro line to be temporarily shut down.

From the *tweets* posted by RATP, I infer the spatial characteristics of interruptions by connecting the names of the metro stations mentioned in an announcement to the (spatial) structure of the metro network. See Appendix A for further details on the data collection. In the end, the data then contain time stamped information on when the metro service is interrupted and which metro stations are affected.

For this study, I focus on local interruptions in the metro network with a duration longer than 20 minutes but less than a day.[70] These interruptions last 1.79 hours on average and the large majority (95.4%) take less than 5 hours (see Figure 4.3a). The distribution over the day exhibits a morning and afternoon peak (see Figure 4.3c). Across the week, Tuesdays and Fridays seem to have slightly more interruptions, while Sundays have less (see Figure 4.3b). Figure 4.3d shows that the data contains between 50 and 100 monthly interruptions for most months, while the months May till June 2017 have more interruptions.

[69] The Twitter accounts are consistently named: `twitter.com/ligne`[*line number*]`_ratp`. For instance, the url of the account for line 1 is `https://twitter.com/ligne1_ratp`.

[70] The lower bound of of 20 minutes is set to match the time interval of the data on rental bicycles, in order minimise measurement error in that respect. For longer interruptions measurement error increases in the length of the interruptions (see Appendix A), and based on manual inspection it turns out that errors longer than 24 hours were often incorrectly measured. Faulty exclusion of errors will have a negligible effect on the results due to very low number of metro interruptions compared to the number of observations. Instead, including metro interruptions that are in fact measurement errors can have a substantial downward bias. Therefore, it seems justified to be on the safe side and omit short and long interruptions.



(a) Histogram.

(b) Number of interruptions per week day.

(c) Interruptions per hour of the day.

(d) Interruptions per month.

Figure 4.3: Characteristics of metro interruptions.

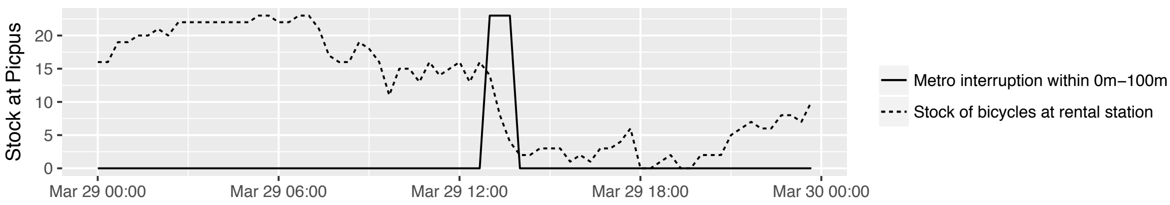


Figure 4.4: Example of docking station ('Picpus') with metro interruption within 100 metres.

4.3.3 Descriptive statistics

For each docking station, metro stations that lie within 100, 200, and 300 metres are identified. I then calculate whether there is an interruption within these bands to analyse the variation over time and across docking stations in relation to metro interruptions. Figure 4.4 provides an example of a metro interruption in the vicinity of a docking station. At first glance the stock of available bicycles seems to fall when the supply of metro drops nearby.

Figure 4.5 shows the net outflow of rental bicycles just before and during a metro interruption. In general all histograms show a distribution with fatter tails compared to the one of the full sample in Figure 4.2d, which indicates that there is more activity in this sub sample. This is likely because interruptions happen during busier hours and at locations with more activity at docking stations. The histograms indicate that there is more activity during interruptions (fewer zero flows) and also suggest that net outflow is higher during metro interruptions (rightward shift of mass). These

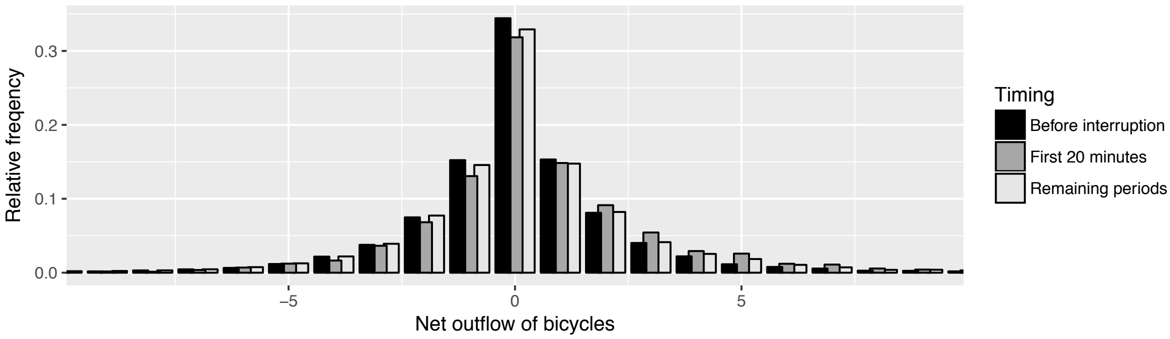


Figure 4.5: Histograms of net outflow conditional on metro interruption within 100m.

effects are most visible during the first 20 minutes of an interruptions. I will employ statistical models that will be discussed in Section 4.4 to test whether this pattern is structural and can be ascribed to metro interruptions.

The descriptive statistics in Table 4.1 indicate that at docking stations, on average, 14 out of the total of 33 stands are occupied by a bicycle. The mean of net outflow ($\tilde{q}_{i,t}$) is roughly zero, as is to be expected due to the bounded stock levels, and the large number of observations. The maximum and minimum net outflow are unrealistically large, either due to data collection errors, or due to occasional relocation of bicycles by the operator. Because there are only very few of these outliers (see the histogram in Figure 4.2d) I do not exclude them, such that the panel remains balanced. The distance to the centre ranges from 200 metres to 10 kilometres, this will be used as interaction with the time of day to control for daily spatial rental patterns.[71] Across the whole dataset, metro interruptions are rare. During the observed period, metro interruptions within 0-100 and 100-200 metres from a docking station occur in only 0.1% of the time for both bands, and in 0.2% of the time for interruptions within 200-300 metres.

4.4 Estimation methods

4.4.1 Baseline models

For the analysis I compare different linear cross sectional and panel estimation techniques to estimate the effect of metro interruptions on the demand for rental bicy-

[71]The centre is set to be Île de la Cité, with coordinates $(48.85341, 2.34880)$.

Table 4.1: Descriptive Statistics for data from July 2016 to June 2017.

Statistic	Mean	St. Dev.	Min	Max
Capacity (bicycle stands)	33.42	13.96	6	99
Stock of bicycles	13.67	14.12	0	99
Net outflow of bicycles	0.0001	1.99	−69	92
Distance to centre (in km)	3.93	1.82	0.22	9.77
Metro interruption within 0m-100m	0.001	0.03	0	1
Metro interruption within 100m-200m	0.001	0.03	0	1
Metro interruption within 200m-300m	0.002	0.04	0	1

Note: 31,154,200 observations accross 1225 docking stations, time interval of 20 minutes.

cles.[72] The unit of observation is set to be a bicycle docking station.[73]

4.4.1.1 Measurement error

Consider the following model

$$q_{i,t} = \beta_0 + \sum_{k=1}^{3} \beta_k x_{k,i,t} + \varepsilon_{i,t}, \tag{4.3}$$

where $q_{i,t}$ denotes the local demand for rental bicycles and $x_{k,i,t}$ denotes discrete variables capturing the presence of a metro interruption in the vicinity within three bands (0m-100m; 100m-200m; 200m-300m). As explained in Section 4.3, I do not observe $q_{i,t}$ but instead $\tilde{q}_{i,t} = q_{i,t} - \hat{q}_{i,t}$, with $\hat{q}_{i,t}$ inflow of bicycles at docking stations, such that there is measurement error in the dependent variable. Substituting the measurement error into equation (4.3) gives

$$\tilde{q}_{i,t} = \beta_0 + \sum_{k=1}^{3} \beta_k x_{k,i,t} + \varepsilon_{i,t} - \hat{q}_{i,t} = \beta_0 + \sum_{k=1}^{3} \beta_k x_{k,i,t} + \epsilon_{i,t}, \tag{4.4}$$

[72] All parameters in this study are estimated using the linear regression function in software package R (R Core Team, 2013).

[73] Alternatively one could aggregate the change in stock of bicycles nearby a metro station, i.e. to make the metro station the unit of observation. This seems to have the advantage that it yields a direct estimate for the change in demand for rental bicycles in an area surrounding a metro station. However, the metro network in Paris is dense (e.g. the distance between some metro stations is less than 300 metres, also see the map in Figure 4.1), such that several docking stations lie nearby multiple metro stations than can simultaneously face interrupted metro service. Therefore, aggregating the observations geographically leads to double counting of the effects at these docking stations, and hence, I choose to use the docking station as the unit of observation.

with $\epsilon_{i,t} = \varepsilon_{i,t} - \hat{q}_{i,t}$. If x is uncorrelated with ϵ then vector β can be consistently estimated using ordinary least squares (OLS).

For x to be uncorrelated with ϵ, the inflow of rental bicycles at docking stations should not be affected by metro interruptions, such that x is uncorrelated with $\hat{q}_{i,t}$. This assumption can be violated when travellers who planned to complement a cycling trip with taking the metro find out about the interruption and decide not to dock their bicycles next to the interrupted metro line.

Although it is in general not unreasonable that cyclists bypass docking stations that are nearby interrupted metro lines, it is unlikely that this happens immediately—within 20 minutes—after the start of an interruption, for the following reason. First, one should note that the usual routine of using a PBS is to dock a bicycle directly upon arrival at any (intermediate) destination. This is also incentivised by the zero marginal price for rentals below 30 minutes, which is renewed with every rental. Moreover, since cycling trips take on average 20 minutes, to be warned about a metro interruption within those 20 minutes requires gathering information *whilst* cycling, for instance by checking information on a smartphone. Given that interruptions are rare in general, and that cycling in a dense city like Paris requires cyclists to pay close attention to the road and traffic, I regard it safe to assume that cyclist find out about an interruption, if at all, only after they stopped cycling and docked their bicycle.

Further, it seems safe to assume that if the estimates are prone to the bias as described above, it is growing over the course of an interruption due to increased awareness about the local shut down of the metro service. Arguably, the longer a metro interruption lasts, the more likely it is that travellers are informed about it. For instance they get informed by checking the schedule for the metro online, even before starting with cycling. However, I find that the estimated coefficients *decrease* over the duration of interruptions rather than increase, see Section 4.5. Hence, this suggests that, even during later periods of an interruption, there is most likely only a small upward bias, if any at all.

Finally, even if some cyclists bypass docking stations as a result of interruptions, then the coefficient is still unbiased when interpreted as a slightly broader measure of net substitution. That is, if cyclists find out about the interruption whilst cycling, decide not to take the metro as planned, and also not dock their bicycle as planned, but instead continue their trip cycling, then they still use the bicycle as a substitute for the metro. Therefore, in the unlikely case that the estimates of actual outflow of bicycles at docking stations are upward biased due to reduced inflow, the estimates still only capture substitution between metro and rental bicycles.

4.4.1.2 Controlling for fixed effects and spatial patterns

The map in Figure 4.1 showed that there is large variation in the centrality of the docking stations. In addition, there are flows towards the centre in the mornings and towards the periphery in the afternoon, see Randriamanamihaga et al. (2014). These flows are similar during weekends and on weekdays, but during weekends they start later (earlier) in the morning (afternoon). To control for this effect I construct a variable that is the interaction between the distance to the centre, defined as D_i, and the time of the day divided in six blocks of four hours, denoted by H_{p_t}, with $p_t := t \bmod 4$.[74] By using four-hour blocks I control for potential correlation between these flows—both on weekdays and during weekends—and interruptions in the metro network. Including these controls to specification (4.4) gives

$$\tilde{q}_{i,t} = \beta_0 + \sum_{k=1}^{3} \beta_k x_{k,i,t} + \sum_{p=1}^{6} \gamma_p D_i H_{p_t} + \epsilon_{i,t}. \tag{4.5}$$

Next, to control for unobserved spatial heterogeneity and temporal trends, I add docking station and hour fixed effects (FE):

$$\tilde{q}_{i,t} = \sum_{k=1}^{3} \beta_k x_{k,i,t} + \sum_{p=1}^{6} \gamma_p D_i H_{p_t} + \alpha_t + \eta_i + \epsilon_{i,t}, \tag{4.6}$$

where η_i and α_t denote docking station and hour specific constants.[75] This model accounts for docking station specific characteristics that can affect the change in stock, for instance due to a favourable location. The hour FE control for temporal effects like (in)favourable weather conditions and other factors that temporarily affect the demand for rental bicycles in the city. I use two ways to estimate equation (4.6). First, mean-differences are calculated per hour and per docking station, this gives

$$\begin{aligned} \tilde{q}_{i,t} - \bar{\tilde{q}}_i - \bar{\tilde{q}}_t + \bar{\tilde{q}} = & \sum_{k=1}^{3} \beta_k \left(x_{k,i,t} - \overline{x}_{k,i} - \overline{x}_{k,t} + \overline{x}_k \right) \\ & + \sum_{p=1}^{6} \gamma_p \left(D_i H_{p_t} - \overline{D_i H_p} - \overline{D H_{p_t}} + \overline{D H_p} \right) + \epsilon_{i,t} - \overline{\epsilon}_i - \overline{\epsilon}_t + \overline{\epsilon}, \end{aligned} \tag{4.7}$$

[74] The four-hour blocks captured by H_{p_t} are set to the following blocks: 1am-4am; 5am-8am; 9am-12am; 1pm-4pm; 5pm-8pm; 9pm-0am. Thereby the morning peak and afternoon peak have distinct controls both on weekdays and during weekends.

[75] Note that there is slight abuse of notation using subscript t, as the time interval is 20 minutes, while I use hour fixed effects instead.

where a bar denotes the average within a group, specified by a subscript, e.g. $\bar{\tilde{q}}_i$ is the average net outflow at station i. I will refer to this estimation method as the FE model. Second, starting again with equation (4.6) I mean-difference per hour and then take first-differences to get

$$\Delta\left(\tilde{q}_{i,t} - \bar{\tilde{q}}_i\right) = \sum_{k=1}^{3} \beta_k \Delta\left(x_{k,i,t} - \overline{x}_{k,i}\right) + \sum_{p=1}^{6} \gamma_p \Delta\left(D_i H_{p_t} - \overline{D_i H_p}\right) + \Delta\left(\epsilon_{i,t} - \overline{\epsilon}_i\right), \quad (4.8)$$

which will be referred to as the first-differences (FD) model.

An advantage of using a FD model over a FE model is a relaxation of the weak exogeneity assumption. That is, the FE model requires for consistency that $\mathrm{E}[\epsilon_{i,t}|x_{i,t}, \eta_i] = 0$, meaning that the error term should be independent from the regressors and the station FE. Instead, a FD model only requires $\mathrm{E}[\epsilon_{i,t} - \epsilon_{i,t-1}|x_{i,t} - x_{i,t-1}] = 0$. However, as the FD estimator is estimated on changes in x, the number of observations on which β is effectively identified is substantially reduced due to interruptions that take longer than one period. As a result, the FD model might not be as efficient as the FE model. Finally, there can be an issue with feedback between $\epsilon_{i,t}$ and future instances of the regressor, denoted here as $x_{k,i,t+j}$, for $j > 1$. As time periods with metro interruptions are serially correllated, the FD estimator can have a bigger bias than the FE estimator, see Wooldridge (2015). Overall, it is difficult to decide ex-ante which model performs best for the problem at hand and I report results for both models.

4.4.2 Duration and low stocks

Next, I analyse two extensions to the baseline model in (4.6). First, I recalculate the regressors for the metro interruptions in such a way that I can distinguish the duration of the interruption. That is, I calculate separate dummy variables for consecutive time periods during a metro interruption. With this routine I can assess the path of substitution over time. Analysing this pattern also sheds further light on the properties of the FE and FD models, because their main difference stems from serial correlation in the metro interruptions.

Second, the supply of rental bicycles in the time window preceding t is bounded by the number of bicycles in stock. Therefore, it might be that metro interruptions next to rental bicycle stations increase the probability of stock depletion. In such cases no bicycle can be taken, while there might still be latent demand. To assess this effect, I estimate a linear probability model using station and hour FE, and centrality-time controls on an indicator variable for stocks below 1, 2, and 3 respectively.

Table 4.2: Estimation results of baseline models.

	Net outflow of rental bicycles			
	OLS	OLS	FE	FD
	(1)	(2)	(3)	(4)
Metro Interruption 0m-100m	0.244^{***}	0.230^{***}	0.241^{***}	0.415^{***}
	(0.028)	(0.027)	(0.027)	(0.041)
Metro Interruption 100m-200m	0.056^{***}	0.046^{**}	0.055^{***}	0.072^{**}
	(0.021)	(0.021)	(0.021)	(0.035)
Metro Interruption 200m-300m	0.015	0.004	0.014	0.075^{***}
	(0.014)	(0.014)	(0.014)	(0.025)
Constant	-0.0002^{***}	-0.0004^{***}		
	(0.00004)	(0.0001)		
Centrality-time controls	No	Yes	Yes	Yes
Hour FE	No	No	Yes	Yes
Observations	31,154,200	31,154,200	31,154,200	31,153,010
R^2	0.00001	0.001	0.001	0.00002

Note: ***, **, * indicate 1%, 5% and 10% significance levels. Standard errors in parentheses are robust and clustered by docking station. Columns (1)-(4) refer to Equations (4)-(7) respectively.

4.5 Results

4.5.1 Estimation results

4.5.1.1 Baseline

Table 4.2 shows that for the baseline model all estimation methods yield a positive and significant impact of metro interruptions on $\tilde{q}$, the net outflow of bicycles at docking stations within 200 metres from a metro station. At docking stations within 100 metres the average net ouflow of bicycles is between 0.230 and 0.415 bicycles per period higher. As expected, the effect seems to decay with distance and is not significantly different from zero for docking stations within 200m-300m in the OLS models and the fixed effects (FE) model.

Comparison of column (1) and (2) shows that controlling for centrality-time patterns reduces the coefficients by roughly 8%, which suggests that there is correlation between the location and time of metro interruptions on one hand, and peak hour usage of the more centrally located docking stations on the other hand. Accounting for

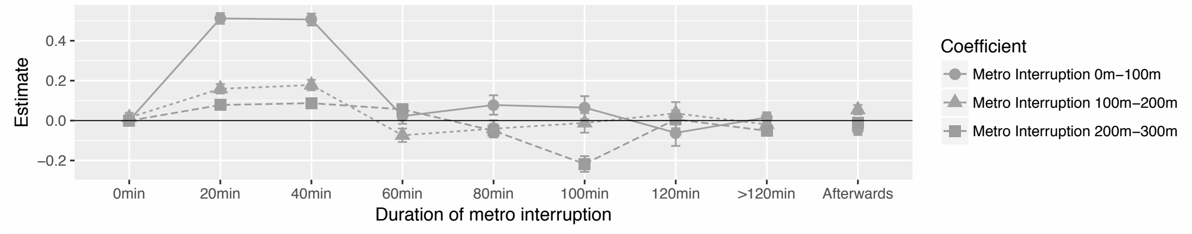


Figure 4.6: Estimations over the course of an interruption using FE model.

docking station FE and hour FE results in only a slight increase compared to the OLS model with additional controls, whereas the FD model finds a substantially higher effect.

As discussed in Section 4.4, there can be various mechanisms that drive the difference between FE and FD estimations. The robust standard errors show that the FD model is indeed less efficient as it uses effectively fewer observations. This is because it is identified only on changes in metro interruptions, which also means that only the start and end points of interruptions matter for the FD model. In contrast, the FE model averages out the effect over the full length of an interruption. The fact that the FD estimates are larger than the FE estimates thus suggests that there is a non-linear and convex decay of substitution during interruptions. The next section further explores this suggestion.

4.5.1.2 Substitution pattern during an interruption

Figure 4.6 shows the estimated substitution pattern during an interruption using the FE model, see Appendix B for the table with underlying results. There are four main lessons to be learned from these results.

First, during the time interval when the interruption starts (indicated by period 0), I do not find a significant effect. These estimates provide a good placebo test and finding a null result here highlights that location and time specific effects do not drive the overall results.

Second, the observed pattern indeed shows a convex decay, suggesting that the FD model overestimates the average results compared to the FE model. This is because the average effect as measured by the FE model does take the rapid decline in the substitution into account, whereas the results from the FD model are driven by the higher substitution at the start of interruptions. Therefore, I regard the FE model as the preferred estimation method for measuring the average substitution.[76]

Third, I find significant substitution in all three distance bands within the first 40 minutes of an interruption. In line with the central estimations in Table 4.2, the substitution effects are larger the closer a docking station is to a metro station. After one hour the effect is roughly zero at all three bands. I identity five potential explanations for this drop. First, there might be measurement error in the timing of the end of an interruption. For instance, when the service is gradually recovering, the operator might still want to wait until the service is running at full capacity before sending the announcement that the interruption is over. Since most observations of interruptions are short (see Figure 4.3a) this then might suppress the coefficients estimated for longer interruptions. Second, the drop can indicate that bicycles are solely used during unexpected shut downs, such that after an hour people are informed about the interruption and seek for other alternatives. Third, it can also be that substitution towards bicycles is not necessarily lower after an hour, but just not observed in the vicinity of metro stations, as there is no need for renting a bicycle exactly nearby the metro station. Fourth, it might be that net outflow drops partially due to increased gross inflow. One explanation for such an effect would be that if interruptions are bi-directionally, then docking stations next to interrupted metro stations are more likely to be a destination for cyclists who rented a bicycle next to another metro station that faced the same interruption. It can also be that the operator redistributes bicycles to docking stations that ran out of bicycles following metro interruptions. In both of these cases I would underestimate the demand for bicycles, although only for later periods of metro interruptions, as it seems highly unlikely that increased inflow as described here takes place immediately at the start of metro interruptions. Fifth, stocks of bicycles might get depleted due to the increased demand.

4.5.1.3 Insufficient stock

Column (1) of Table 4.3 shows that metro interruptions increase the probability that a docking station is empty by 1.8 percentage points. Recalling from Figure 4.2a that stations are empty 13% of the time, it means that interruptions are associated with a roughly 15% increase in the probability that a station is empty. This suggests that there can be latent demand for rental bicycles during interruptions. Note, however, that for a given docking station the distance to the next nearest docking station is 369 metres on average. It is therefore likely that part of latent demand due to empty

[76] I also analysed the duration effect using the FD model, but this yields very similar results. The fact that there is little difference between the different techniques when estimating the interrupted periods separately resembles the equivalence between FD and FE estimators in two-period panels.

Table 4.3: The effect of metro interruptions on probability of low stock.

	Dependent Variable		
	P(Stock<1)	P(Stock<2)	P(Stock<3)
	(1)	(2)	(3)
Metro Interruption 0m-100m	0.018***	0.022***	0.030***
	(0.006)	(0.007)	(0.008)
Metro Interruption 100m-200m	0.009	0.010	0.012
	(0.007)	(0.008)	(0.008)
Metro Interruption 200m-300m	0.004	0.005	0.007
	(0.005)	(0.006)	(0.006)
Observations	31,154,200	31,154,200	31,154,200
R^2	0.001	0.002	0.003

Note: model calculated using station FE, hour FE and centrality-time controls. ***, **, * indicate 1%, 5% and 10% significance levels. Standard errors in parentheses are robust and clustered by docking station.

stocks is satisfied at other docking stations nearby.[77]

Columns (2) and (3) in Table 4.3 indicate that metro interruptions are associated with a 2.2 and 3.0 percentage points increase in the probability of stocks of one or two bicycles respectively. This is of interest because docking stations with some bicycles in stock might fail to satisfy demand for two reasons. First, if people do not travel alone, the availability of a single bicycle is not sufficient. Second, in case of low stocks, the bicycles left are more likely to be broken or to have a flat tire. This is because bicycles are not taken from docking stations randomly, and people will select non-defect bicycles to rent.[78]

[77]To see whether depleted stocks have an effect on the baseline estimations I re-estimated the baseline models using sub-sets in which cases are excluded with stocks below 1, 2, 3, and 4 bicycles respectively. Although this would provide potentially valuable insights in the extent to which there is latent demand, the analysis is problematic because the data is selected based on characteristics of the dependent variable (stock levels) that are correlated with the independent variable (metro interruptions). Nevertheless, when I estimate this model, I find only moderate increases in the estimated increase in demand for rental bicycles. Which might still suggests that the baseline results would not be much higher if sufficient bicycles were available at all times.

[78]Based on personal observations of docking stations in Paris, indeed often bicycles are registered as being available but actually are not usable due to defects.

4.5.2 Implications

In this section I calculate the total effects that the estimated coefficients imply, and relate these figures to annual statistics of the PBS and metro traffic.

4.5.2.1 Bicycle rental induced by metro interruptions

The average duration of metro interruptions in the sample is 1.79 hours. The total number of metro interruptions in the full set of Twitter data is 2,283.[79] During a metro interruption, 11.15 and 16.80 docking stations are on average affected for the closest band (0m-100m) and second band (100m-200m) respectively. Using these results I calculate that in the observed year (, 32,897 and 11,314 bicycles are used directly in response to metro interruptions for the closest band (0m-100m) and second band (100m-200m) respectively. That means that in total 44,211 bicycle trips were untertaken to alleviate time losses from metro interruptions. Assuming that the 39.4 million of Vélib' hires in Paris in 2015, as estimated by Marie de Paris (2016), can be extrapolated to the time span of the sample, metro interruptions account for 0.11% of all annual Vélib' trips.

Finally, because I observe stocks rather than flows, I need to approximate the gross outflow for situations without an interruption. That is, I have identified $q^{\#}_{i,t} - q^{\star}_{i,t}$ empirically, but not $q^{\#}_{i,t}$ and $q^{\star}_{i,t}$ separately, where $q^{\star}_{i,t}$ and $q^{\#}_{i,t}$ denote the quantity of bicycles that are rented in a normal situation and during a metro interruption nearby respectively. To still be able to have an indication of the relative increase in demand for rental bicycles, I approximate net outflow outside interruptions as the share of annual outflow from Marie de Paris (2016) using the ratio between observed transactions during metro interruptions and the total number of transactions. This yields a relative increase of rental bicycles outflow at stations within 100m of 11% on average and 22% during the 20 minutes of a metro interruption.

4.5.2.2 Average substitution during interruptions

Using annual metro traffic statistics from RATP (2017), I can relate the results to metro usage. Yearly 1.41 billion (non-unique) passengers enter one of the 303 metro stations within Paris. That is 177 passengers per metro station per 20 minutes on average. To account for non-service hours I adjust this average under the assumption that the metro runs 18 hours a day—roughly in line with the distribution of interruptions in Figure 4.3c—to get an average of 236 travellers entering a metro station during the

[79] This set is not trimmed because the number of interruptions can still be accurately inferred from the full data set.

active hours. I use this value as a rough indication of the number of travellers affected within a time interval of 20 minutes during an interruption.[80]

Next, recall that the baseline estimations in Table 4.2 indicate that for the closest two distance bands, metro interruptions are associated with an increase in bicycle rental of 0.241, and 0.055 respectively. Using these results I find that during metro interruptions 0.16% of the travellers that enter a metro station are estimated to switch to a public rental bicycle. Finally, if I repeat this exercise with estimations of the increased demand during the first 20 minutes of an interruption, then I find that 0.44% of the travellers that would have entered a metro station switch to a bicycle for their trip.

4.6 Conclusion

This paper analysed the interaction between a public bicycle system (PBS) and public transport by estimating the effect of metro interruptions on the local demand for rental bicycles within the Vélib' PBS in Paris. I find that, as a direct consequence of a metro interruption, the consumption of bicycles within 100 metres of metro stations increases by 0.72 bicycles per hour per docking station on average, and with 1.54 bicycles per hour per docking station during the first 20 minutes; an increase of approximately 11% and 22% respectively. Due to their effects on demand, metro interruptions increase the probability of empty stocks at docking stations with 15%. The findings highlight that cycling is a local net substitute for metro service, and that public rental bicycles can alleviate time losses stemming from public transport interruptions.

Based on the results, I identify several directions for further research. First, since a PBS is used as a substitute for public transport during interruptions, models of the optimal location for bicycle docking stations, like the one from Wuerzer and Mason (2016), can be extended to include these effects. That is, location decisions should seek to embed a PBS into the broader transport network so as to reduce to the system's vulnerability. Second, future research should assess whether and by how much public rental bicycles are used as long term substitutes for metro service, and not just during interruptions. This can provide relevant evidence on to whether the promotion of cycling can be used as a second-best policy instrument to address congestion in public transport, see Prud'homme et al. (2012), or to reduce car congestion, see Adler and van Ommeren (2016). Third, future research could analyse more in depth how cycling infrastructure and public transport interact. This in order to better understand

[80]Note that not all metro interruptions are bi-directional and many stations serve more than one line. Therefore, this number is likely an overestimate of the total number of travellers affected by an interruption.

determinants of using cycling and public transport as complements (as suggested for trains) or as substitutes (as found in this study). In addition, stock management of bicycles at docking stations might be further optimized to enhance shock-absorbing capacity of the combined systems. Finally, it would be valuable to complement this study by analysing other cities.

Appendix 4.A Data collection

Metro operator RATP uses a Twitter account for each line to give status updates for travellers. The accounts are consistently named with the following structure `twitter.com/ligne`[*line nr.*]`_ratp`. Twitter allows for downloading the last 3,200 tweets from any account. This was done on July 8 2017. A tweet announcing an interruption is for instance

> *15:35, le trafic est interrompu entre Esplanade de la Defense et La Defense (Grande Arche) (panne de matériel) # RATP # Ligne1.*
>
> (posted Fri Apr 14 13:38:24, by `Ligne1_RATP`)

These announcements are usually, but not always, followed by a clearing message, in this case

> *Retour à un trafic régulier sur l'ensemble de la # Ligne1 # RATP. Incident terminé.*
>
> (posted Fri Apr 14 14:47:24, by `Ligne1_RATP`)

Using these tweets, the timing and location characteristics of metro interruptions are inferred. The start time of an interruption is set to be the time stamp of the tweet, which was maximum a couple of minutes later than the time mentioned in the tweets and accounts for cases where no start time was mentioned in the message itself. The location is determined by connecting the stations mention in a tweet to the network characteristics. For instance from the example above I infer the all stations on line 1 between and including *'Esplanade de La Defense'* and *'La Defense (Grande Arche)'* face interruptions.

Based on manual inspection of the messages sent, a clear pattern of potential interruption announcing messaged emerges such that the following three types of announcements could be classified as the start of an interruption:

- *"le trafic est interrompu..."* (traffic is interrupted);
- *"le trafic est perturbé.."* (traffic is disrupted);
- *"la station* [name] *est fermée..."* (station is closed).

The end of interruptions is classified if there was a clearing message or if there was an announcement of delays. The latter is included because it indicates that the metro is running again, albeit with some delays. Because from the tweets one cannot infer whether these delays are severe or minor, it seems safer to assume that the metro is running and that the interruption is over.

It is important to note that missing an occasional interruption due slightly other phrasing used for an announcement does not affect the results of this study. It might slightly reduce the statistical power, but in the total average of non interrupted time-station pairs its effect on the estimates is negligible due to the rarity of interruptions (see Table 4.1). For this reason, I only include cases of interruptions that were followed by a clearing message (the large majority, at least 87%).[81] However, even for interruptions with a clearing message, finding the correct time for the end of interruptions in the RATP Tweets is not straightforward and prone to measurement error. This is because not all endings of interruptions are announced, and thus the different messages on interruptions cannot be linked (e.g. with an ID). The only way to determine the end of an interruption is through scanning through consecutive tweets that succeeded the announcement to look for a clearing message. In case no clearing message was sent—e.g. because the error interruptions was not solved within a day—this algorithm erroneously sets the ending of the error at a clearing message that belongs to a subsequent error. A consequence thereof is that measurement error increases in the duration of the errors. Although 95.4% of interruptions takes less than 5 hours—for which the end of interruptions could be accurately determined—I exclude interruptions longer than 24 hours to avoid measurement error in the timing of longer interruptions (see also Section 4.3).

[81] This percentage is calculated as the ratio between total number of clearing messages with total number of interruption announcements. Note, however, that this value is an underestimate of the actual ratio and should be treated as a lower bound. This is because sometimes interruptions are updated, where every update is flagged as a new interruption.

Appendix 4.B Additional results

Table 4.B.1: Substitution pattern during metro interruption, using FE model.

Metro Interruption within	Net outflow of bicycles
	FE
0m-100m0 Start	0.001 (0.027)
0m-100m1	0.512*** (0.027)
0m-100m2	0.506*** (0.030)
0m-100m3	0.022 (0.039)
0m-100m4	0.078 (0.049)
0m-100m5	0.065 (0.057)
0m-100m6	−0.062 (0.066)
0m-100m-Rest	0.015 (0.026)
0m-100m-Stop	−0.045* (0.027)
100m-200m0 Start	0.015 (0.023)
100m-200m1	0.159*** (0.024)
100m-200m2	0.178*** (0.026)
100m-200m3	−0.074** (0.034)
100m-200m4	−0.041 (0.042)
100m-200m5	−0.012 (0.049)
100m-200m6	0.036 (0.057)
100m-200m-Rest	−0.019 (0.021)
100m-200m-Stop	0.054** (0.024)
200m-300m0 Start	−0.001 (0.018)
200m-300m1	0.078*** (0.019)
200m-300m2	0.087*** (0.021)
200m-300m3	0.056** (0.026)
200m-300m4	−0.051 (0.034)
200m-300m5	−0.217*** (0.039)
200m-300m6	0.006 (0.045)
200m-300m-Rest	−0.051*** (0.017)
200m-300m-Stop	−0.012 (0.019)
Observations	31,154,200
R^2	0.001

Note: model calculated using station FE, hour FE and centrality-time controls. ***, **, * indicate 1%, 5% and 10% significance levels. Standard errors in parentheses are robust and clustered by docking station.

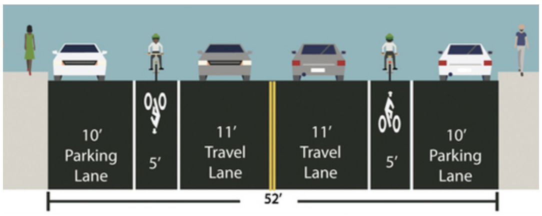
5 Bikes lanes, road safety, and congestion: evidence from New York City

5.1 Introduction

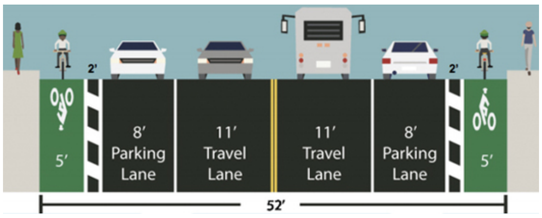
Urban policymakers around the globe increasingly see cycling as an essential part of sustainable urban transport. Urban cycling is associated with benefits for public health (De Hartog et al., 2010), reductions in air pollution (Gössling and Choi, 2015), and a reduction in congestion (Hamilton and Wichman, 2018). Cycling also has disadvantages. In particular, compared to other transport modes, cycling is associated with higher accident risk.[82] Hence, a policy that aims to induce a modal shift from cars to bicycles by increasing the cost of driving (e.g. a road charge) can lead to more (fatal) accidents (Schepers and Heinen, 2013). Therefore, many cities invest in cycling infrastructure, often with protected bike lanes, to accommodate an increase in urban cycling, while minimising road safety issues. In 2020, several cities sped up their cycling infrastructure investment, following tighter capacity constraints in

This chapter is based on a paper joint with Dan Graham, Jos van Ommeren, and Erik Verhoef. For this chapter I want to thank Anupriya, Laila Ait Bihi Ouali, Praj Xuto, Csaba Pogonyi, Prateek Bansal, Francis Ostermeijer, Hans Koster, Alexandros Dimitropoulos, Martin Adler, and seminar and conference participants in Amsterdam (Eureka), London (ICL), Toulouse (ITEA) for their helpful comments and suggestions.

[82] Nilsson et al. (2017) find that for injuries cyclists face a 29 times higher accident rate compared to car drivers and passengers, and a 10 times higher rate for fatal accidents.



(a) Standard bike lane.



(b) Protected bike lane.

Figure 5.1: Illustration of bike lane upgrades (adopted from NYC DoT, 2020).

public transport due to the COVID-19 pandemic (De Vos, 2020; Honey-Roses et al., 2020).

In this paper, we analyse the effect of bicycle infrastructure on road safety and traffic congestion, using methods similar to Mangrum and Molnar (2018). We focus on New York City (NYC) and exploit spatial and temporal variation from a recent expansion of protected bike lanes, that occurred mainly on Manhattan. Our paper contributes to a growing body of literature that examines the relationship between cycling infrastructure and road safety. Marshall and Garrick (2011) find improved safety due to bike lanes by analysing road safety trends in several cities in California. At a more disaggregated level, Li et al. (2017) find that installation of London's Cycle Superhighways did not increase collision rates.[83] For NYC, Gu et al. (2017) find that bike lanes are cost-effective due to their road safety improvements. In a related study, Wall et al. (2016) find that bike lanes decrease the severity of accidents involving cyclists. Further, bike lanes appear to yield most safety improvements close to intersections and on roads with high traffic volumes (Kondo et al., 2018).[84]

Here, we analyse effects of protected bikes in NYC, often as upgraded from standard bike lanes, see Figure 5.1. We improve on the literature by addressing sorting. We do so by focusing on Manhattan, which offers the advantage that it allows for, as we will show, accurate traffic proxies by using yellow taxi trips. Furthermore, the grid structure of its streets facilitates clean identification of causal effects as one can derive likely routes (discussed below).[85]

[83]The authors find an increase in the *total number* of collisions on Cycle Superhighways, driven by increased mileage on those roads. This highlights that addressing sorting is essential in road safety analyses.

[84]This is in contrast to our results, which seems to suggest that most safety benefits are away from junctions, see Section 5.5.

[85]Mangrum and Molnar (2018) analyse how much of the overall slowdown in NYC's traffic can be attributed to taxis. In their study, bike lanes are used as a control variable, but their sample includes only a few bike lanes. Hence they cannot provide (average) causal effects of bike lanes on congestion. Furthermore, they do not focus on road safety and do not address cyclist's sorting.

We focus specifically on congestion and safety effects of bike lanes to provide a policy-oriented overview of the impacts of bike lanes on traffic conditions. Because bike lanes induce rerouting of cars and cyclists, estimation at street-level is prone to sorting biases. We address sorting by aggregating street-level data to area-direction observations, where we exploit that north and south-bound trips have to traverse these areas. As a consequence, all streets affected by bike lanes, including those used for rerouting, are then covered by the estimated treatment effects. This approach not only avoids sorting bias but also provides us with an area-wide treatment effect, which is a relevant input for evaluating bike lane policies.

This paper proceeds as follows. Section 5.2 describes the data and presents descriptive statistics. Section 5.3 analyses the extent that one can use taxi trips as proxy for traffic flow conditions. Section 5.4 describes the empirical strategy. Section 5.5 discusses results. Section 5.6 concludes.

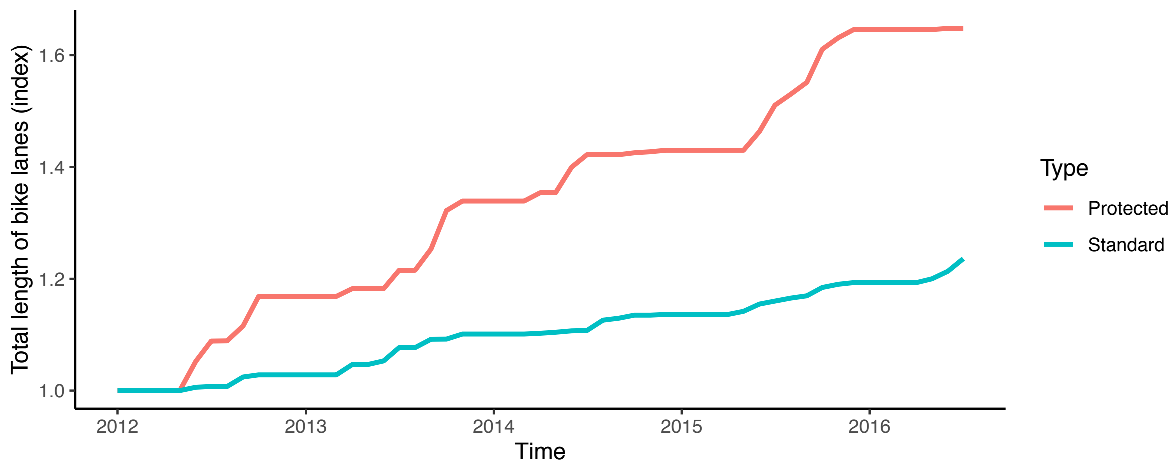


Figure 5.2: Growth of dedicated bike lanes in New York City over the sample period.

Notes: Protected bike path indicates a lane that is separated from other travel lanes, by parking bays, a barrier, or both. Standard bike lanes are marked on the road surface and adjacent to other travel lanes. Indices are calculated per month with January 2012 as base.

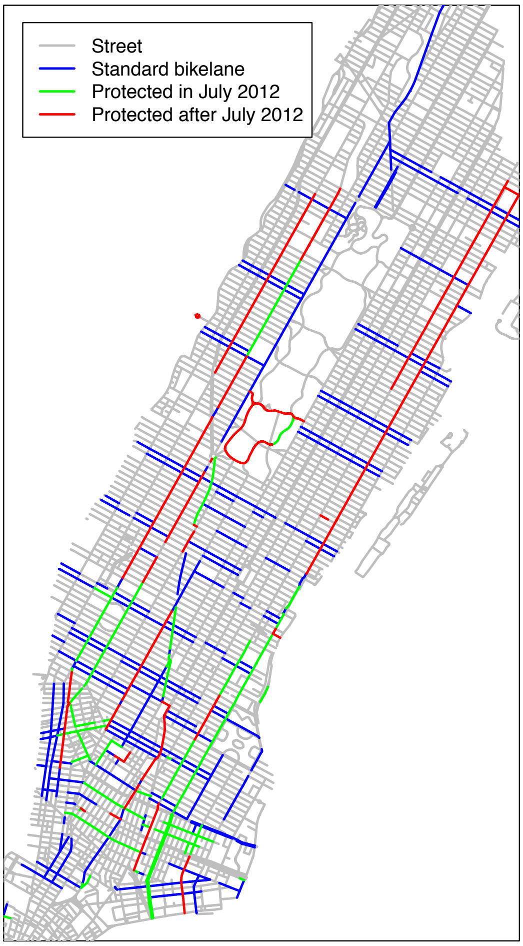
5.2 Data

5.2.1 Street network

We obtain information on the street network and bike lanes in NYC from City of New York (2020a,c). For comparability between treated streets (those with a protected bike lane) and control streets, we only focus on 'normal' streets and thereby exclude highways, tunnels, trails, etc. Information on the type and installation date of bike lanes is available for each street segment. Figure 5.2 shows the development of bike lanes in NYC during our sample period. The figure highlights that bicycle infrastructure grew substantially, with an increase of around 60% in the total length of protected bike paths.[86]

The map in Figure 5.3a shows the spatial distribution of existing (in July 2012) and newly installed (after July 2012 but before July 2016) protected bike lanes on Manhattan. The map highlights that within our sample period, several streets received a protected bike lane.

[86]The expansion of protected paths occurred in waves, which provides us with temporal variation that we exploit for identification.



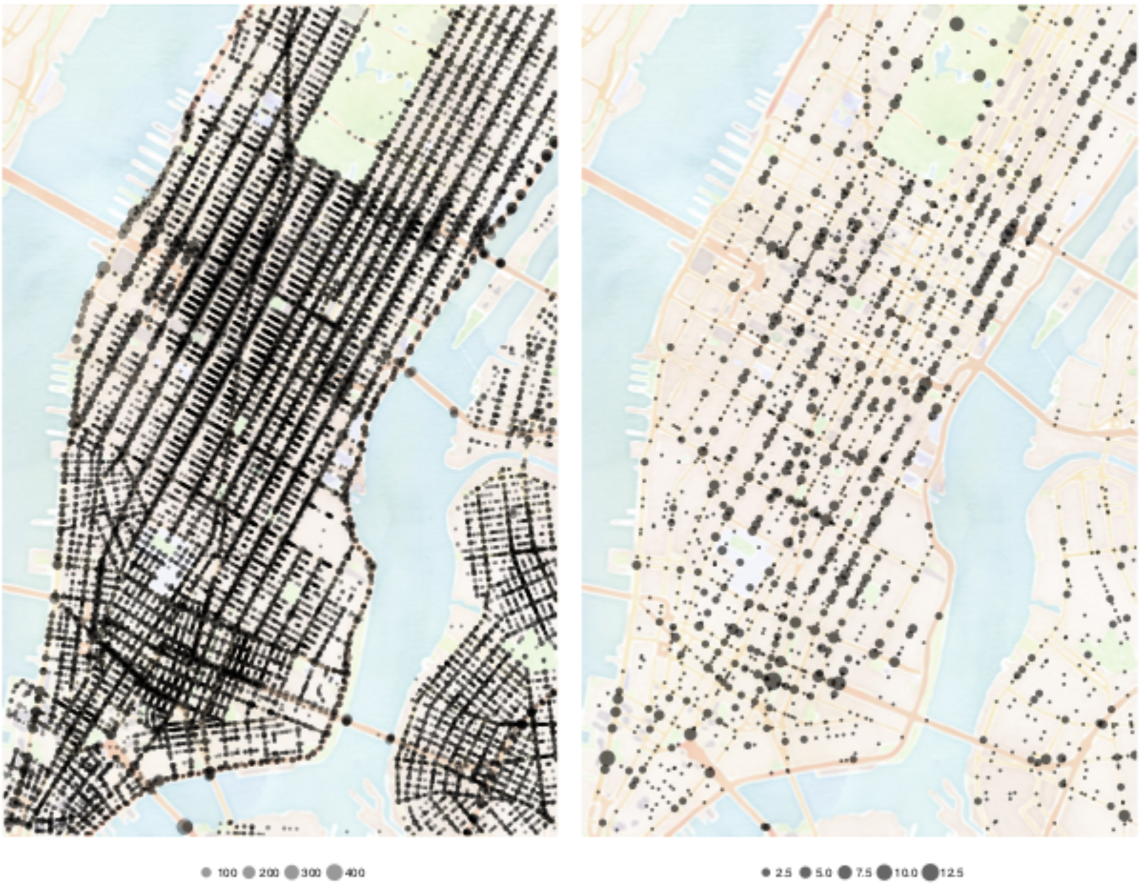
(a) Bike lanes on Manhattan.

(b) Manhattan areas.

Figure 5.3: Maps with the spatial distribution of bike lanes and defined aggregation areas.

We focus on Manhattan, for which nearly all protected bike lane installations are on the north-south axis. We use this feature for our identification by defining 35 areas as in Figure 5.3b. Because routes are either going north or south, we have 70 area-directions. Each area covers a whole slice of Manhattan spanning roughly eight east-west streets each. The idea here is to exploit that north-bound and south-bound trips have to traverse these areas.

As a consequence, any change in infrastructure within an area that induces rerouting must affect traffic flow on other streets *within the same area*. We can, therefore, address sorting (due to rerouting) by using aggregate data for area-directions. Hence, we treat the whole area-direction as treated in case a bike lane is installed in that area.



(a) All accidents. (b) Accidents with cyclists involved.

Figure 5.4: Map with spatial distribution road accidents in Manhattan between 2012 and 2020

Notes: Map tile by Stamen Design, under CC BY 3.0, base layer by OpenStreetMap, under CC BY SA.

5.2.2 Road accidents

We observe 0.73 million accidents (of which 3.1% with cyclists involved), as reported by New York Police Department (NYPD) between July 2012 and June 2016 (City of New York, 2020b). Based on the geographical location and a timestamp, we assign accidents to the closest street segment (with time-specific bike lane properties at the time of the accident). We further observe the number of vehicles involved and the severity of the accident, that we classify as material damage (more than \$1000 of damage, but no injuries), severe (at least one person injured) and fatal (one or more deaths). Figure 5.4 shows accident locations for Manhattan, and indicates that accidents (including those with bikes) are spread across most of Manhattan.

5.2.3 Car traffic

For 387 locations, we observe hourly traffic volumes based on on-site counters from City of New York (2020e). These data contain counts of all vehicles that pass a certain street segment. However, these data only cover short time windows, typically less than a week. Hence, we use these data to validate our approach to use taxi trips as a proxy for traffic flow and speed.[87]

We observe 480 million trips made in yellow taxis from July 2012 until June 2016, obtained from City of New York (2020d). For each trip, we observe the origin and destination, departure time, arrival time, and trip distance as measured by the taxi meter.[88] Figure 5.5 shows that most taxi trips cover less than 5km, take less than 20 minutes, and that travel speed is often below 20 km/h. Panel 5.5d shows that a substantial amount of trips (more than 20 million) have a trip distance equal to the euclidean distance between origin and destination, which implies that these trips did not take any turn, such that we can infer their *exact* route. We will use these trips to identify traffic conditions on Manhattan (see similarly Mangrum and Molnar, 2018).

We select the subsample of trips that start and end on the same road which we will refer to as within-road trips. We then use this subsample to infer a proxy for traffic speed and traffic flow as observed using on-site counters. In Section 5.3 below, we discuss this method in detail and assess the quality of the proxy for flow.

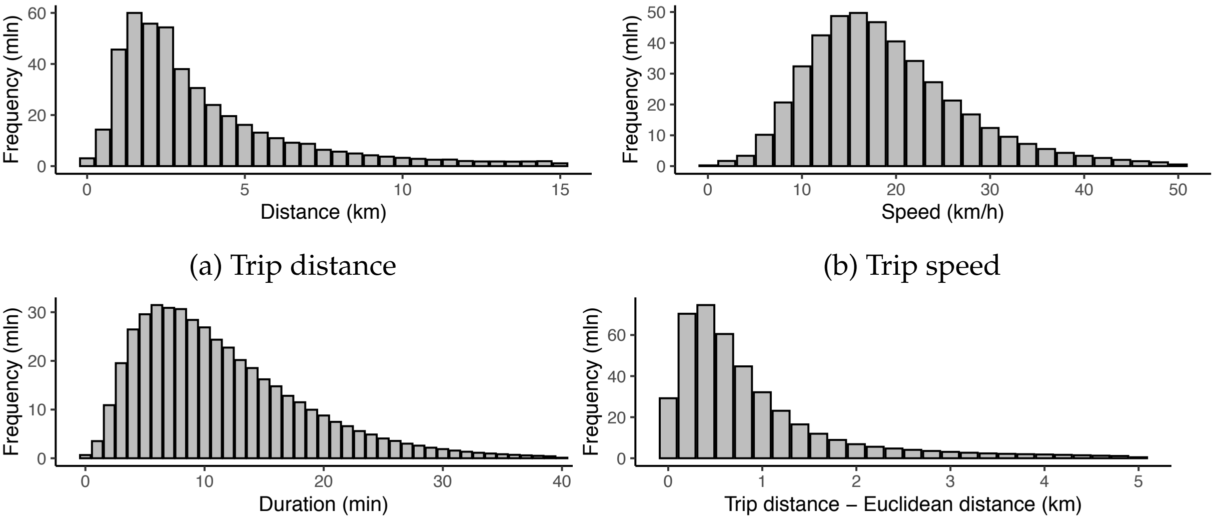
5.2.4 Descriptive statistics

We geographically assign all traffic and accident observations to the nearest street segment. We first determine hourly averages (for taxi traffic) and hourly counts (for accidents). We then aggregate to weeks, where we weight the speed for the number of taxi passings per hour. We thus get a volume-weighted weekly indicator for speed, so that this measure captures congestion weighted for the number of trips.

Next, we select streets that on Manhattan and have a north or south direction. The resulting data contains observations for 209 weeks on 1,786 street segments, with

[87]Our main assumption here is that taxi passings are proportional to total traffic flow, discussed in detail below in Section 5.3

[88]From July 2016 onwards City of New York (2020d) does not report the exact origin and destination pair. Therefore, we cannot precisely identify within-road trips.



(a) Trip distance

(b) Trip speed

(c) Trip duration

(d) Difference between euclidean and actual distance.

Figure 5.5: Descriptives taxi data.

provides us with an unbalanced panel with 299,376 street-level observations. Panel A inn Table 5.1 provides descriptive statistics.

We also aggregate our street-level data to the area-direction level, where we combine all north-bound or south-bound streets segments within an area as depicted on the map in Figure 5.3b. The gives 14,839 observations from 35 areas each with two directions. Panel B in Table 5.1 provides descriptive statistics at this level of aggregation.

5.3 Taxi data as proxy for traffic

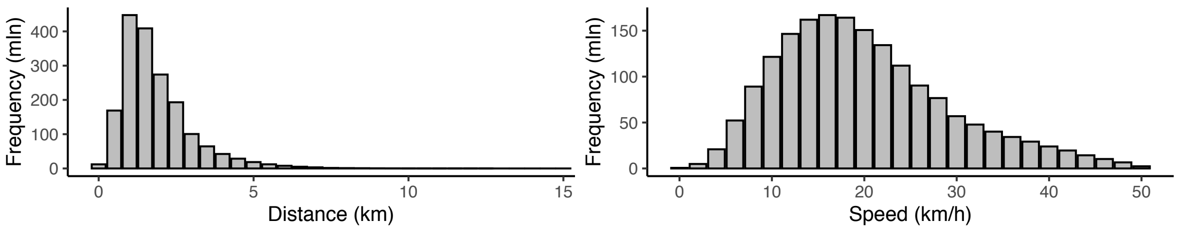
We use information on taxi trips to construct street-level indicators for speed, travel time and flow. For speed, we closely follow the strategy of Mangrum and Molnar (2018). First, we select a subsample of within-road trips, i.e. taxi trips that have their origin and destination on the same road (street or avenue). Next, we exclude observations for which the great circle distance between its OD pair is smaller than the trip distance as measured by the taxi meter.[89] The resulting sample contains 17.92 million trips for which we can be sure that no detour was taken (e.g. to avoid congestion). For this subsample, Figure 5.6 shows that, as expected, the trip distance is shorter compared to the full sample, but the speed is roughly similar.

[89] This can occur because of the error margin in the GPS meter (see Mangrum and Molnar, 2018).

Table 5.1: Descriptive statistics.

Statistic	Mean	St. Dev.	Min	Max
Panel A: Street level				
Speed (km/h)	22.143	6.116	0.102	93.590
Taxi volume	541.203	756.886	1	5,720
Accidents (all)	0.119	0.413	0	9
Accidents (at junctions)	0.115	0.407	0	8
Accidents (severe)	0.017	0.133	0	5
Accidents (material)	0.102	0.378	0	8
Accidents (with bikes)	0.005	0.071	0	2
Protected bike lane	0.119	0.324	0	1
Panel B: Area-direction level				
Speed (km/h)	22.732	5.363	0.453	64.213
Taxi volume	34,003.410	43,158.610	0	359,776
Accidents (all)	2.886	2.897	0	19
Accidents (at junctions)	2.775	2.862	0	18
Accidents (severe)	0.415	0.715	0	6
Accidents (material)	2.471	2.593	0	17
Accidents (with bikes)	0.118	0.361	0	4
Protected bike lane	0.623	0.485	0	1

Notes: We have 299,376 street-level observations and 14,839 area-direction observations.



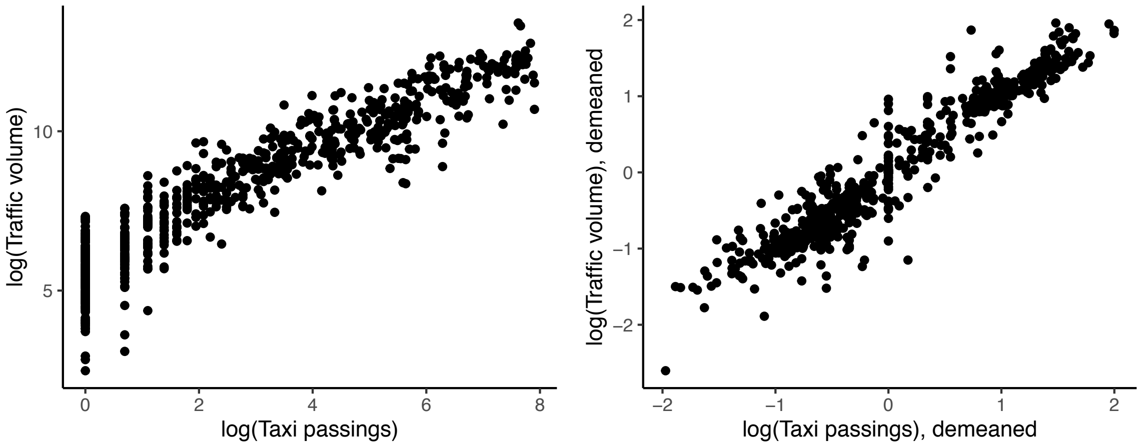
(a) Trip distance

(b) Trip speed

Figure 5.6: Distribution of trip distance and travel speed of within-road taxi trips.

We calculate the travel time (in hours per kilometre) for each week w and each link l (a street segment in-between two junctions) as the weighted average of all the trips that fully traversed that link.[90] Further, we calculate the taxi flow per link as the sum

[90]Mangrum and Molnar (2018) also explore two alternative methods to infer the speed of segments based on taxi trips, but note that the unconditional mean of taxis passings, as used here, performs best in terms of bias-variance trade-off. We focus on parts of the city that are less crowded than their study area, so in our setting it is even more important to have less variance.



(a) In logarithms. (b) Demeaned per location.

Figure 5.7: Traffic flow and taxi passings at weekly level.

of all within-trips that pass by a certain segment. We use these counts to proxy for total traffic flow. Figure 5.7 shows that weekly taxi flow and total flow are strongly correlated, especially when demeaned with a link-specific constant as in panel (b). In the next section, we further analyse this relationship.

5.3.1 Predicting total flow with taxi flow

We analyse the quality of our proxy for flow in Table 5.2. Column (1) shows that in a bivariate linear regression, taxi flow has a strong correlation with total traffic flow, but also indicates that their relationship is not proportional and that the predictive power is modest (as indicated by the R^2 of around 0.56). In contrast, column (2) shows that adding a link fixed effect yields a precisely estimated coefficient of unity, which implies that up to a street-specific (log) constant, taxi passings are exactly proportional to the total traffic flow at a weekly level. More importantly, the within-R^2 is close to unity (0.96) such that we can accurately predict weekly traffic flow using taxi passing counts.

Columns (3)-(4) show that focusing on the number of drop offs or pick ups on a link seems to be a less precise predictor of total traffic flow, although proportionality is still within a 95% confidence interval. Column (5) shows that the sum of all taxi counts yields similar results. Column (6) highlights that pick ups are stronger correlated to total traffic flow than drop offs. This makes sense, especially in NYC, as taxis cruise to find demand in busy streets.

The results of our hourly (Table 5.B.1) and daily (Table 5.B.2) analysis in Appendix

B indicates that at these lower levels of temporal aggregation, the within-R^2 is much lower, such that taxi trips do not accurately capture total traffic flow. We, therefore, conclude that the finest temporal resolution for which taxi passings can proxy total traffic flow is at the weekly level and that street-segment fixed effects are essential to obtain an accurate proxy for traffic volumes.

Table 5.2: Estimation results testing predictive power of taxi trips on traffic volume.

	log(Total traffic flow)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(Taxi passings)	0.602***	1.007***				0.599***
	(0.036)	(0.024)				(0.118)
log(Taxi pick ups)			0.987***			0.321**
			(0.034)			(0.147)
log(Taxi drop offs)				0.963***		0.088
				(0.037)		(0.084)
log(Taxi sum)					1.031***	
					(0.023)	
Street segment FE		Yes	Yes	Yes	Yes	Yes
Within R^2	0.565	0.958	0.949	0.922	0.964	0.965
Observations	212	212	206	209	212	204
R^2	0.565	0.981	0.976	0.964	0.984	0.983

Notes: Robust standard errors in parentheses are clustered at the level of a street segment (i.e. counter location).***, **, * indicate significance at 1%, 5%, and 10%.

5.4 Identification strategy

We aim to identify the causal effect of bike lane installation on road safety, car traffic speed and car traffic flow. Four main statistical challenges arise in our setting. First, bike lanes are not installed at random as they are targeted at unsafe streets and areas (City of New York, 2014). We address this issue by including location fixed-effects such that we exploit within-location variation that stems from bike lane installation.

Second, there are plausibly general trends in transport and road safety that correlate with the bike lane installation program in NYC. Therefore, we include time fixed effects to absorb time trends that correlate with bike lane installation. We thus apply a two-way fixed effects approach, to estimate our effects based on deviations from city-wide time trends *and* from location-specific characteristics.

Third, one expects travellers to reroute after the installation of a bike lane. For instance, cyclists may reroute from streets parallel to benefit from the new bike lane, and cars may avoid roads with bike lanes. This implies that a street-level analysis is prone to biases due to treatment-induced rerouting (or violation of the stable unit treatment assumption). To address this issue, we use aggregated observations of streets in the same direction within areas. Thereby we exploit Manhattan's elongated shape and grid-structured street network.

Fourth, roadworks required for the installation of bike lanes are likely to disrupt traffic and thereby affecting congestion and road safety. To avoid such effects to influence our results, we exclude observations three months prior to a change in infrastructure. In addition, we exclude the first month after the bike lane has been installed to avoid any post-roadwork adjustments that may affect our results.

5.4.1 Main specification

To control for sorting due to rerouting, we combine all streets in an area and travel direction into direction-area-week observations. We classify an area-direction as treated when it is possible to traverse an area north bound or south bound fully on a protected bike lane. This approach not only assures causality by avoiding violation of the stable unit treatment value assumption, but also yields policy-relevant estimates at the area level. We consider variants of the following general specification:

$$Y_{zt} = \phi_z + \kappa_t + \beta \cdot \text{bikelane}_{zt} + \gamma \cdot X_{zt} + \epsilon_{zt}, \tag{5.1}$$

where z denotes area-direction and t time in weeks. We estimate specifications with Y denoting log of speed (inverse travel time), log of taxi traffic flow, and accidents counts. For the latter, we use Poisson regression.[91] We include two types of fixed effects. First, ϕ_z, is an area-direction fixed effect, which absorbs any unobserved time-invariant characteristic of the area, including the unobserved constant that assures that taxi flow accurately describes the total traffic flow. Second, κ_t is a time fixed effect, that controls for city-wide trends and seasonality. Further, X denotes traffic controls, which, depending on the specification includes log of taxi traffic flow, log of speed, or both. Finally, ϵ_{zt} denotes the error term that we cluster at the level of an area to adjust the standard errors for serial correlation.

[91] We will use the algorithm as developed by Correia et al. (2019) to avoid the computational burden of estimating parameters for the fixed effects.

5.4.2 Estimating sorting effects

To get a sense of rerouting and sorting, we also perform a descriptive analysis at the level of a street segment. We use a similar two-way fixed effects approach as above, to control for city-wide trends and street-segment specific unobservables, but in contrast to the causal model in (5.1) we estimate the treatment parameter on variation across streets *within* each area-direction. We consider a similar specification as above:

$$Y_{zst} = \tilde{\phi}_{zs} + \tilde{\theta}_{zt} + \tilde{\beta} \cdot \text{bikelane}_{zst} + \tilde{\delta} \cdot X_{zst} + \tilde{\epsilon}_{zst}, \tag{5.2}$$

where z denotes an area, s a street segment, and t time in weeks. We now include fixed effects for each street segment within an area ($\tilde{\phi}_{zs}$), and also one for each week interacted with area-direction ($\tilde{\theta}_{zt}$). The latter assures that we identify changes in the traffic *within* areas. Therefore, $\tilde{\beta}$ will capture both the causal effect, but also sorting effects due to rerouting. Hence, comparing estimates from (5.1) and (5.2) sheds light on the extend to which travellers reroute following bike lane installation.

Table 5.3: Effect of protected bike lanes on traffic accidents at area-direction level.

	(1) Accidents	(2) Accidents	(3) Accidents	(4) Accidents	(5) Bike acc.	(6) No bikes acc.
Bike lane	0.685***	0.707***	0.223*	-0.029	-0.421***	-0.013
	(0.162)	(0.162)	(0.131)	(0.033)	(0.141)	(0.034)
log(Traf. volume)			0.242***	0.164***	0.453*	0.155***
			(0.029)	(0.049)	(0.248)	(0.051)
Area-direction FE				Yes	Yes	Yes
Time FE		Yes	Yes	Yes	Yes	Yes
Observations	12,362	12,362	12,362	12,362	11,900	12,362

Notes: Coefficients in columns (1)-(4) are estimated using Poisson regression. All dependent variables are in logs. Robust standard errors in parentheses are clustered at the level of an area.***, **, * indicate significance at 1%, 5%, and 10%.

5.5 Results

5.5.1 Main road safety results

Table 5.3 shows the main regression results. Column (1) shows the unconditional correlation between bike lanes and accidents. We find a positive correlation: bike lanes are installed in areas with around 98% more accidents compared to areas without a bike lane. We stress that this is not a causal effect since we do not control for area characteristics. In column (2), we include time-fixed effects, which hardly affects the point estimate compared to the column (1). This indicates that bike lanes do no correlate with seasonal trends in accidents. In column (3), we still find a positive (but smaller) effect on accidents when controlling for traffic volume. For a Poisson model, when controlling for the log traffic volume, the effect of bike lanes can be interpreted as the effect on the accidents rates, i.e. the number of accidents per traffic volume. Hence, the result here implies that areas with bike lanes tend to have higher accident *rates* than those without a bike lane.

Column (4) depicts our preferred specification for overall accident levels. Here, we obtain a causal estimate by absorbing unobserved area characteristics using an area-direction fixed effect. The results highlight that we do not find a statistically significant effect of bike lanes on traffic accidents. We thus conclude that installing a bike lane in an area does not seem to affect overall accident levels.

Table 5.4: Effect of protected bike lanes split by location.

	(1) All	(2) Not at junction	(3) At junction	(4) At junction with bikes
Bike lane	-0.029	-0.894***	0.010	-0.394**
	(0.031)	(0.252)	(0.032)	(0.154)
log(Traf. volume)	0.164***	0.491**	0.113**	0.363
	(0.044)	(0.235)	(0.045)	(0.255)
Area-direction FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	12,362	10,077	12,362	11,836

Notes: Coefficients are estimated using Poisson regression. Robust standard errors in parentheses are clustered at the level of an area.***, **, * indicate significance at 1%, 5%, and 10%.

In the next analysis, we distinguish between accidents involving bicycles and those without bicycles. We then find, as shown in column (5), evidence for a causal 34% reduction in accidents with cyclists involved due to bike lane installation in an area. In contrast, we do *not* find such an effect on accidents without bicycles involved, as shown in column (6). The latter makes sense because we control for any change in volume induced by the bike lane.

5.5.2 Effect at and away from junctions

In Table 5.4, we use subsamples to analyse how bike lanes have differential effects at junctions and away from junctions. Column (1) is a copy of our preferred specification as a reference. In column (2) where we focus on accidents away from junctions, we find a statistically significant coefficient, indicating a 59% reduction in the number accidents involving all modes, due to bike lane installation. Column (3) highlights that focusing only on accidents at junctions still leads to a null result.[92]

In column (4) we focus on accidents involving cyclists. Here, we find a negative effect that is statistically significant, but not statistically different from our main estimate for accidents with bikes involved. This implies that for cyclists, bike lanes lead to similar safety improvements both at junctions and on streets.

[92] Naturally, the point estimate is higher than the one from column (1), since the latter is a weighted average of treatment effects at junctions and away from junctions. Because the vast majority of accidents occur at junctions, it makes sense that we obtain null results both in column (1) and (3), while still finding a negative and statistically significant effect in column (2).

Table 5.5: Effect of protected bike lanes on traffic indicators.

	(1) log(Traf. speed)	(2) log(Traf. vol.)	(3) log(Traf. speed)	(4) log(Traf. vol.)
Bike lane	-0.013	0.007	-0.013***	0.023***
	(0.011)	(0.028)	(0.003)	(0.008)
Aggregation level	Area	Area	Street	Street
Time FE	Yes	Yes	Yes	Yes
Area-direction FE	Yes	Yes		
Time × Area-direction FE			Yes	Yes
Street-segment FE			Yes	Yes
R^2	0.803	0.995	0.894	0.986
Observations	12,362	12,362	244,189	244,189

Notes: Robust standard errors in parentheses are clustered at the level of an area.***, **, * indicate significance at 1%, 5%, and 10%.

We also explored analysing bike accidents away from junctions. However, given the few accidents away from junctions, this analysis yields data that is too sparse to be able analyse with a Poison fixed effects approach, that we need to assure causality. Because the point estimate for bikes at junctions (column (4) in Table 5.4), is smaller than the one for all bike accidents (column (5) in Table 5.3), it is very plausible that bike accidents away from junctions are also reduced due to bike lanes.

5.5.3 Further sensitivity analyses

In Appendix B we report further sensitivity analyses. Table 5.B.3 shows that the results are not affected by the inclusion of traffic speed as control. Further, Table 5.B.4 highlights that do not find statistically significant coefficients when focusing on severe accidents (with at least one injury or a fatality) or accidents with only material damage. This suggests that bike lanes do not differentially affect light or more severe accidents.

5.5.4 Traffic and sorting

In Table 5.5 we analyse the effects on traffic. In columns (1) and (2) we use observations at the area-direction level, such that we can obtain a causal effect. In column (1), we find a negative, but not statistically significant, effect of bike lanes on the speed of car traffic for the whole area. This suggests that bike lanes have no differential effects on traffic speeds within the same area-direction. Column (2) shows that we cannot identify any effect on (area-direction wide) traffic flow. This suggests that weekly demand is not affected by a bike lane. Note that we control for city-wide trends, so our results here are abstracted from overall modal shifts (e.g. from cars to bikes). Table 5.B.3 in Appendix B shows that controlling adding additional controls for speed or traffic volume does not change the estimates.

In columns (3) and (4) we estimate the same specifications, but now on street-level observations. These results are descriptive in the sense that they both capture a causal effect of bike lanes, as well as any effects from rerouting. In column (3) we find that bike lanes face traffic speeds that are 1.3% lower compared to other streets in that area-direction. In column (4) we find that streets with bike lanes have a 2.3% higher weekly flow compared to other streets in the same area-direction. This suggests that some cars reroute *towards* roads with bike lanes, e.g. because of safety improvements.

5.6 Conclusion

Governments increasingly invest in cycling infrastructure. In this paper, we focus on New York City and estimate the causal effect of protected bike lanes on traffic speed and flow, and road safety *at the area level*. Because a street-level analysis is prone to biases due to treatment-induced rerouting, we use aggregated observations of streets in the same direction within narrowly defined areas on Manhattan. Thereby we exploit Manhattan's elongated shape and grid-structured street network to assure causality, but also to obtain policy-relevant area-level estimates.

We find that bike lanes improve safety for cyclists both away from junctions and at junctions. Protected bike lanes reduce by 34%. Our results further indicate that bike lanes have no statistically significant effect on overall road safety at junctions. However, we find that a bike lane in an area reduces accidents away from junctions by 59% for *all modes* combined.

Using taxi trips as an accurate proxy for traffic indicators, we find no statistically significant evidence for an effect of bike lanes on traffic speed or traffic flow (at the area level). However, we find that the traffic speed on streets with bike lanes is 1.3% lower compared to streets in the same direction within the same area. At the same time, our results indicate that streets with a bike lane accommodate a 2.3% higher throughput. If we assume that streets on Manhattan often operate at their maximum capacity, this increase in traffic flow is likely possible due to the reduction in speed, such that traffic density is increased.

In our setting, we have to be agnostic about underlying mechanisms. For instance, we cannot distinguish between a behavioural effect (e.g. cyclists and car driver may be more or less cautious due to a bike line), and a mechanical effect (e.g. even with the same driving behaviour the number accidents may be reduced). For a more detailed understanding, future research could disentangle these mechanisms by collecting higher resolution data on driving behaviour from motorised vehicles and bicycles. Ultimately, one could think of empirically estimating a model with endogenous speed choice as a function of road safety (such as the one developed by Verhoef and Rouwendal, 2004). Thereby, such a design could isolate mechanical and behavioural effects of bike lanes on road safety.

Appendix 5.A Additional descriptives

5.A.1 Additional maps

Figure 5.A.1: Map with locations of traffic counters.

Appendix 5.B Additional results

5.B.1 Further taxi proxy analysis

Table 5.B.1: Estimation results testing predictive power of taxi trips at hourly level.

	log(Total traffic flow)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(Taxi passings)	0.324***	0.262***				0.098***
	(0.033)	(0.013)				(0.007)
log(Taxi pick ups)			0.186***			0.098***
			(0.010)			(0.011)
log(Taxi drop offs)				0.158***		0.062***
				(0.008)		(0.008)
log(Taxi sum)					0.323***	
					(0.013)	
Street segment FE		Yes	Yes	Yes	Yes	Yes
Within R^2	0.210	0.201	0.142	0.099	0.307	0.177
Observations	20,416	20,416	15,639	14,986	20,416	13,277
R^2	0.210	0.746	0.765	0.760	0.780	0.803

Notes: Standard errors clustered at street segment.***, **, * indicate significance at 1%, 5%, and 10%.

Table 5.B.2: Estimation results testing predictive power of taxi trips at daily level.

	log(Total traffic flow)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(Taxi passings)	0.800***	0.852***				0.568***
	(0.021)	(0.019)				(0.037)
log(Taxi pick ups)			0.668***			0.200***
			(0.021)			(0.028)
log(Taxi drop offs)				0.643***		0.058**
				(0.024)		(0.026)
log(Taxi sum)					0.772***	
					(0.018)	
Street segment FE		Yes	Yes	Yes	Yes	Yes
Within R^2	0.822	0.704	0.660	0.595	0.672	0.837
Observations	1,991	1,991	1,477	1,512	1,991	1,373
R^2	0.822	0.965	0.949	0.941	0.961	0.973

Notes: Standard errors clustered at street segment.***, **, * indicate significance at 1%, 5%, and 10%.

5.B.2 Robustness checks

Table 5.B.3: Effect of protected bike lanes on traffic accidents at area-direction level.

	(1) Accidents	(2) Accidents	(3) log(Speed)	(4) log(Speed)	(5) log(Volume)	(6) log(Volume)
Bike lane	-0.029	-0.032	-0.013	-0.012	0.007	0.005
	(0.033)	(0.033)	(0.011)	(0.011)	(0.028)	(0.028)
log(Traf. volume)	0.164***	0.149***		-0.050**		
	(0.049)	(0.049)		(0.022)		
log(Traf. speed)		-0.298***				-0.183***
		(0.065)				(0.066)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Area-direction FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2			0.803	0.804	0.995	0.995
Observations	12,362	12,362	12,362	12,362	12,362	12,362

Notes: Standard errors clustered at area level.***, **, * indicate significance at 1%, 5%, and 10%.

Table 5.B.4: Effect of protected bike lanes on accident severity.

	(1) All	(2) Severe	(3) Severe	(4) Material	(5) Material
Bike lane	-0.032	0.007	0.003	-0.032	-0.038
	(0.031)	(0.076)	(0.076)	(0.033)	(0.033)
log(Traf. volume)	0.149***		0.016		0.179***
	(0.044)		(0.113)		(0.047)
log(Traf. speed)	-0.298***		-0.227		-0.306***
	(0.072)		(0.178)		(0.077)
Area-direction FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	12,362	12,362	12,362	12,362	12,362

Notes: Standard errors clustered at area level.***, **, * indicate significance at 1%, 5%, and 10%.

6

Hands on the wheel, eyes on the phone: the effect of smartphone usage fees on road safety

6.1 Introduction

Traffic accidents are an important loss to society. In the European Union (EU), for example, about 25,000 road users lost their lives due to traffic accidents in 2018. For every death on European roads, there are an additional 50 injuries of which 8 are severe and 4 cause permanent disability (European Commission, 2019b). Next to this physical harm, accidents also cause psychological suffering to those directly involved and to friends and relatives of the victims. Traffic accidents also lead to monetary losses due to damages to private and public property and are a major cause of traffic congestion. The total costs of traffic accidents in the EU are estimated to be about €280 billion, or 2% of GDP, which makes it the most important external cost of transportation (European Commission, 2019a). Similar numbers can be found for the United

This chapter is based on Brands et al. (2020), the authors would like to thank Jos van Ommeren, Erik Verhoef, Hans Koster, Paul Koster, Niels Bos, Jiska Klein, Dan Graham, Csaba Pogonyi, Laila Ait Bihi Ouali, Niek Mouter, Hendrik Wolff and conference and seminar participants in Amsterdam (VU), Paris (ITEA), London (Imperial College), Budapest (hEART), Toulouse (SBCA), Rotterdam (EEA), Berkeley (UEA), and Jakarta (Universitas Indonesia). We also would like to thank Rijkswaterstaat Netherlands for granting us access to the data.

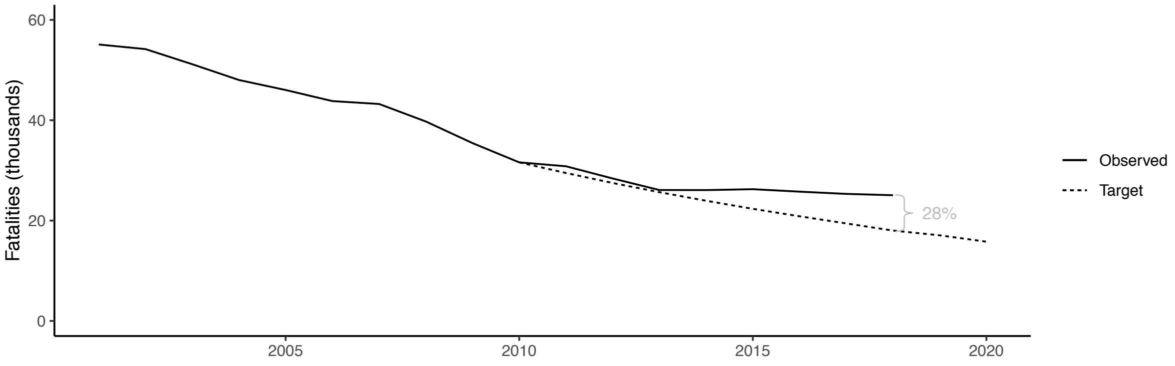


Figure 6.1: Road fatalities in the EU and 2020 policy target (European Commission, 2019b)

States and other countries (Blincoe et al., 2015).

These high costs explain the vast body of scientific literature on traffic accidents that exists today, including important contributions from the field of economics on related topics such as the risk of drunk driving (Levitt and Porter, 2001a), the size of the accident externality caused by one typical additional driver (Edlin and Karaca-Mandic, 2006), and the effect of mandatory seatbelt laws on traffic fatalities (Cohen and Einav, 2003).[93] The substantial costs of accidents also provide governments with a strong rationale to prioritise safety in road design, and in traffic and vehicle-related regulation. Safety concerns in this respect largely shape policy decisions on aspects such as speed limits, road geometry, obligatory usage of seatbelts, and factors that affect the ability of road users to maintain attention on the driving task. This includes prohibiting the use of alcohol and cell phones by drivers. Figure 6.1 indicates that stricter safety regulations over the past two decades have had a promising impact on the number of road fatalities in the EU. However, progress in terms of reductions in road fatalities, as compared to the EU policy target formulated by the European Commission, began to diverge and stagnate in 2013, even after accounting for vehicle kilometres travelled.[94]

Despite regulations that forbid car drivers from using mobile phones while driving, effective regulation has proved to be difficult, and technological progress in recent years has transformed cell phones into omnipresent devices that can be seen as a major cause of distraction in traffic. Smartphones stand out as a major culprit, as they have enabled various novel distractions, including sending and receiving messages

[93]Other notable contributions include: Levitt and Porter (2001b), Adams and Cotti (2008), Jacobsen (2011), and DeAngelo and Hansen (2014).

[94]Data on vehicle kilometres travelled for all EU countries does not span back until 2000, so we plot fatality rates per million passenger-km for four major EU countries in Figure 6.A.1 of Appendix A, which shows a similar trend.

via numerous applications, news updates, video calling, and receiving notifications from social media platforms. In experimental settings, this has been shown to cause visual, cognitive, and physical distractions which result in longer reaction times, less awareness, and various other deficiencies which restrict full control of the vehicle (Zhao et al., 2013, Young et al., 2014, Haque and Washington, 2015).

Findings from the lab are generally corroborated by observational studies in naturalistic settings and crash-based studies (see e.g. Dingus et al., 2016, Redelmeier and Tibshirani, 1997, and McEvoy et al., 2005). However, various studies using field data fail to conclusively prove this relation.[95] In the first large scale field study of its kind, Bhargava and Pathania (2013) estimate the effect of mobile calls on accidents using a discontinuity in the price scheme at 9 pm between 2002 and 2005. They find a 7.2 percent increase in call likelihood after the price drop but no corresponding increase in the number of accidents at the 9 pm threshold. Further research on the effect of statewide mobile phone bans in the US indicates that the effects are short-lived, if detectable at all (Abouk and Adams, 2013; Burger et al., 2014).

The most recent studies that focus on smartphones find more conclusive negative safety effects. Hersh et al. (2019) exploit temporal variation in 3G coverage in California between 2001 and 2013 to study the effect of gaining access to mobile data on vehicle accidents. After controlling for vehicle kilometres travelled and road segment fixed effects, the authors find that crash rates increase by 1.1 percentage points when roads receive 3G coverage. Furthermore, Faccio and McConnell (2020) find that locations with a lot of activity of Pokémon Go (a popular video game app on the smartphone at the time) faced more vehicle accidents after the introduction of the game, suggesting that 136 of the total 2850 nation wide crashes (approximately 5%) in the five months after the introduction of the game could be attributed to it.

Although numerous studies have investigated the link between phone use and accidents, a substantial research gap prevails.[96] Most existing estimates are dated, while mobile phone use has dramatically changed since the turn of the century in terms of adoption, exposure and capabilities.[97] For example, in the much-cited study by Redelmeier and Tibshirani (1997), only 18% of drivers owned mobile phones which had limited capabilities, while in more recent studies, Bhargava and Pathania (2013) only focus on mobile calling and Hersh et al. (2019) end their study in 2013. Furthermore, studies that do address the interaction between modern smartphones, with data usage, and accidents, either focus on very specific non-generalisable phone-use (Poké-

[95] Drivers may also be able to navigate streets more easily using navigation applications, hence the effect of phone use on traffic accidents is not per se negative.

[96] See e.g. reviews by WHO (2011), Oviedo-Trespalacios et al., 2016, and Lipovac et al. (2017).

[97] Mobile phone subscriptions per capita have been above one in the world since 2016 (World Bank, 2019) and in 2018 smartphone penetration was above 70% in many developed nations (Newzoo, 2018).

mon Go in Faccio and McConnell, 2020), or only focus on highways (Hersh et al., 2019). In addition, most studies do not account for unobserved factors that may be correlated to both phone use and accident likelihood, such as risk preferences at the individual level and demand factors at the aggregate level. Finally, as sample sizes were often small in experimental and crash-based studies, generalisation to aggregate effects is often problematic. Therefore, an important and ongoing research question is to what extent does smartphone use while driving affect the number and likelihood of traffic accidents.

We propose a novel approach based on field data and a natural experiment induced by a change in EU roaming regulations. The specific policy, imposed in June 2017, mandated mobile phone operators to abolish all roaming surcharges for EU customers travelling outside their home country network within the EU. The policy, dubbed *Roam Like at Home* (RLAH), implied that people travelling abroad within the EU now face their home fee, which is substantially lower than pre-policy charges. As a consequence, growth in roaming cellular traffic increased sharply after the policy. Mobile data use while roaming grew by over 200 percentage points, whereas local usage was not affected by the policy and faced stable growth rates.[98] We hypothesise that, as of June 2017, EU citizens driving abroad are more likely to be distracted by their phone, while nothing changed for local usage.

We use microdata on all police-reported road accidents in the Netherlands from 2014 until 2018. We then use vehicle registration information to classify which (foreign) drivers are plausibly treated by the RLAH policy. The causal effect of phone use on road accidents is then estimated using a difference-in-differences (DiD) approach, where we use the RLAH policy as treatment, and local users as control group. This allows us to overcome endogeneity issues from earlier studies due to measurement error in phone use and omitted variables. Our key identification assumption is that in the absence of the policy, the number of vehicle accidents by roaming users should follow similar trends to local drivers, for which we provide evidence in our parallel trends plot.

Our findings imply that the increase in phone use due to the policy causes the number of accidents to increase by around 10%. Under plausible assumptions, this implies a crash risk odds ratio of around 3.8. Under the assumption that this mechanism also carries over to local drivers and holds for other EU countries, our results then imply that each year as many as 2,500 road fatalities in the EU can be attributed to phone use. This suggests that about one-third of the gap between the EU target and the observed number of fatalities shown in Figure 6.1 could be reduced by successfully banning mobile phone use of drivers.

[98] Growth rates have been calculated using information from the International Roaming BEREC Benchmark Data Reports (for roaming) and the Dutch Authority for Consumers and Markets (for locals).

This study contributes to the existing literature in five ways. First, our results provide a causal estimate of phone use on road safety based on a novel method. Second, because our identifying variation comes from a very recent policy intervention, our estimates take into account modern distractions of smartphones, and particularly changes in mobile data use. Third, because our analysis is based on revealed and non-experimental field data of all registered accidents in the Netherlands, we are able to estimate an aggregate effect. This is especially relevant given the urgency of road safety issues and the rapid growth in cellular traffic. Fourth, with our approach, we can estimate how smartphone distractions affect accidents for different severity levels and on different road types. We show that phone distractions increase accident risk predominantly on local urban roads, which highlights that studies focusing solely on highways underestimate the total effect. In addition, our results indicate that both light accidents, as well as fatal accidents, increase due to smartphone use. Fifth, we introduce an identification strategy that is directly applicable to *all* other countries in the European Union, allowing for convenient cross-validation of our results using data from other countries in future research.

The rest of this paper is structured as follows. Section 6.3 describes the policy context, Section 6.4 explains the methods employed, and Section 6.2 presents the data we use. Section 6.5 discusses our results, robustness checks, and implications. Finally, Section 6.6 concludes.

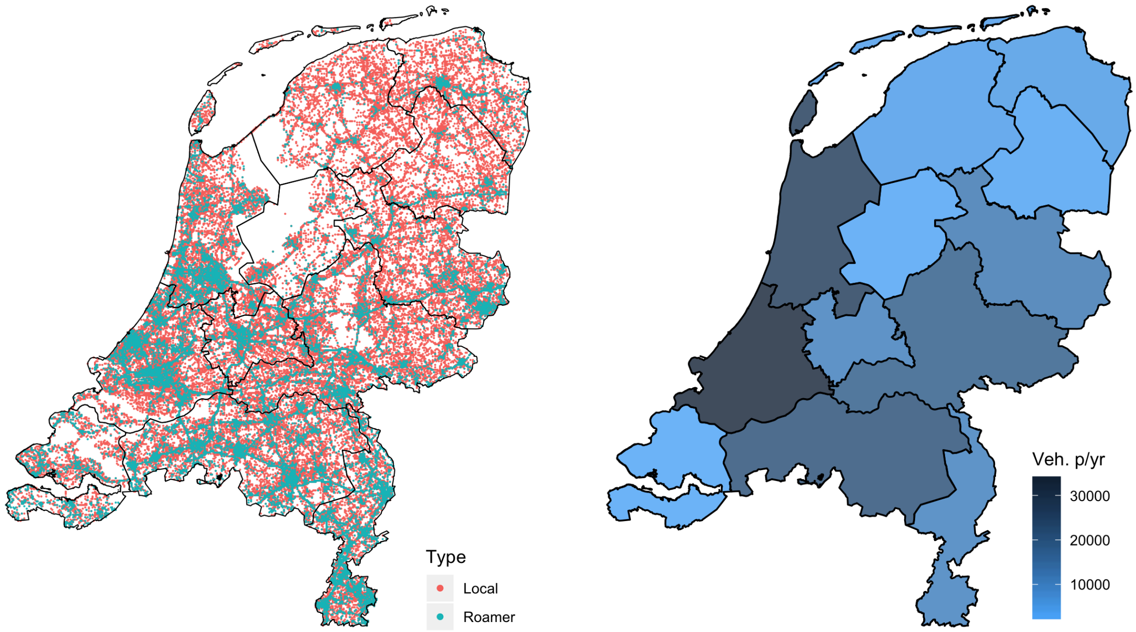
6.2 Data and context

6.2.1 Road safety data

We observe police reported accidents in the Netherlands as published by the Dutch Ministry of Infrastructure and Water Management (specifically 'Rijkswaterstaat'). The maps in Figure 6.2 plot the locations and annual counts of vehicles involved in accidents per province. The maps highlight that accidents are spread across the country, but more concentrated around urban areas and highways.

Our data contains characteristics of road accidents and of the parties involved.[99] For each accident, we observe accident circumstances, such as day of the week, time of the day, road type, weather conditions, and road surface conditions. Furthermore, the dataset contains vehicle related characteristics, such as vehicle type, vehicle manoeuvre just before the crash, sex and age of the driver, and the country in which the

[99] We use the full dataset available to researchers as we require privacy sensitive information on vehicle registration nationality. A publicly available version of the data is available on `data.overheid.nl`, but does not contain all party characteristics.



(a) Accident locations.

(b) Vehicles in accidents per year per province.

Figure 6.2: Maps of the Netherlands with accident locations and counts per province.

vehicle is registered.[100] Finally, party related variables are also reported and provide information such as age and sex of involved parties, casualty severity and whether the casualty was a driver, passenger, cyclist, or pedestrian.

We directly observe the vehicles' country of registration. Drivers of cars registered in EU countries, but outside of the Netherlands are likely to reside in those EU countries. Therefore, vehicle registration is a good proxy of whether the driver incurs roaming costs (before RLAH) or uses the local network instead.[101] To abstract from long term trends, we use data for the years 2014 until 2018, which contains 0.76 million vehicles involved in 0.44 million accidents. Most accidents have more than one vehicle involved (78%), therefore we use information at the party level to avoid measurement error which may be present at the accident level, as police reports do not

[100] For our particular application we cannot use most of these characteristics as they are often missing for non-local cars. This is because these data stem from the car registry in the Netherlands, which is not connected to databases from other countries. The data does not contain information on whether a car is rented or leased.

[101] Dutch law requires that any vehicle staying in the Netherlands for more than six months must obtain a Dutch licence plate. Note that, due to our difference-in-difference method, misclassification can pose a problem for the efficiency of our estimator, but will not bias our estimates under the plausible assumption that misclassification is not correlated to the roaming regulation.

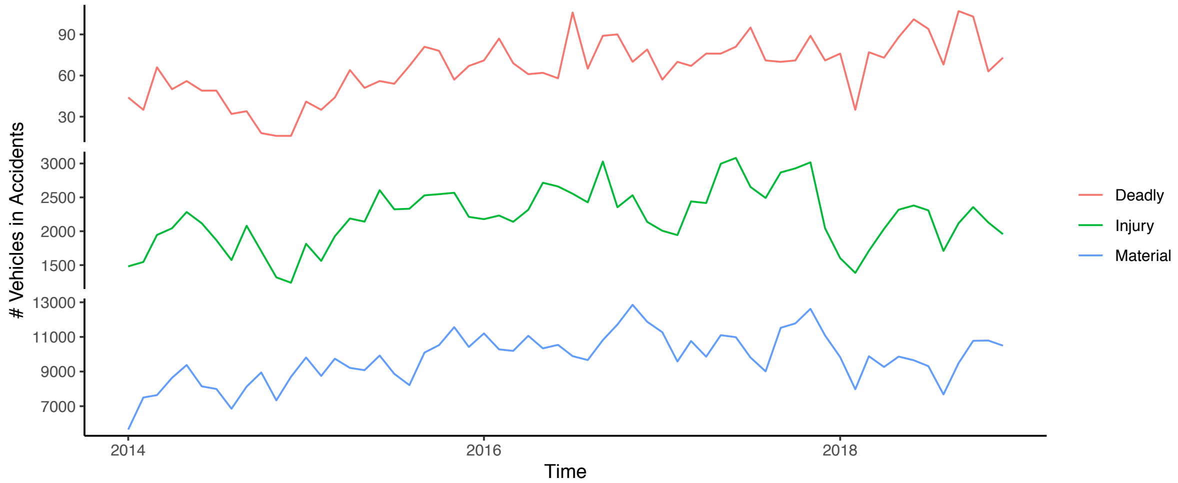


Figure 6.3: Number vehicles involved in accidents per month by severity.

indicate which party was at fault. We discuss this issue and how we deal with it in more detail in Section 6.4.2.

6.2.1.1 Trends in road safety

Figure 6.3 shows that there appears to be an increase in the number of vehicles involved in accidents over all levels of severity. Over the period of study, our data shows that the annual number of deadly accidents increased by around 20%, while the number of accidents involving injury and material damage increased by about 50%, with most of the change between 2014 and 2016. In an average month there are around 74 vehicles involved in deadly accidents, 2,381 vehicle accidents involving injury and 10,280 vehicle accidents involving material damage.[102]

6.2.1.2 Grouping roaming drivers

We combine observations in our sample into six country groups for our main analysis. The aim of this grouping is to strike a balance between, on the one hand, optimally controlling for unobserved heterogeneity per country of origin (by means of group fixed effects), and on the other hand, preserving statistical power by avoiding zero counts (which are omitted due to the log transformation of the dependent variable,

[102] We also checked whether the number of vehicles per accident is stable over time, which turns out to be the case, both for accidents with only locals as well as accidents with at least one roaming user involved.

see Section 6.4 for a discussion).

The first group contains vehicles with a Dutch registration and is our control group (95.12% of sample). Second and third, are the two adjacent countries, with 1.76% of German vehicles and 1.04% of Belgian vehicles, respectively. The fourth group contains other western European countries, which account for 0.42% of vehicles in accidents. Drivers from these countries often visit the Netherlands as tourists.[103] The fifth group contains Romanian, Polish, and Bulgarian vehicles (1.32%) which are relatively common on Dutch roads due to joint economic activity and labour migration. More than for other cases, drivers from these labour migration countries may have a Dutch phone subscription and thus might not be treated by the RLAH policy. Therefore, it is important to include a separate fixed effect for vehicles from these countries. It also allows us to run a robustness check where we exclude vehicles from these countries, which highlights that vehicles from these countries do *not* drive our overall results (see Section 6.5.3.1). The sixth group contains all remaining EU countries (0.33%).

6.2.2 Descriptive statistics

6.2.2.1 Vehicles involved in accidents

Around 5% of vehicles involved in accidents are from roaming users, 46% of drivers are female and the average age is 42 years old. Of the total number of accidents, 0.58% are deadly, 18.7% result in injury, and 80.72% cause material damage only.[104]

Local and roaming drivers involved in accidents are roughly comparable, but roaming users tend to be younger, male, and drive more on fast roads than local drivers.[105] In terms of the damage reported, the share of material damage is relatively large for roaming vehicles. This may be a reporting bias, as language barriers can make it more likely for the police to be called in these situations with only material damage, whereas locals may more easily settle without police present. Importantly, dissimilarities between local and roaming drivers do not threaten our identification under

[103] These are: France, Great-Britain, Denmark, Spain, Austria, Portugal, Luxembourg, Sweden, Italy, Ireland, Norway, and Finland.

[104] Table 6.A.1 in Appendix A presents the descriptive statistics for vehicles involved in accidents.

[105] Table 6.A.2 in Appendix A provides more detailed descriptives of vehicles involved in accidents by group.

Table 6.1: Descriptive statistics for province-month data

Statistic	N	Mean	St. Dev.	Min	Max
Panel A: Locals					
Vehicles in accidents	720	953.11	757.40	84	3,297
log(Vehicles in accidents)	720	6.52	0.86	4.43	8.10
No trucks	720	189.13	133.80	26	564
Single vehicle accidents (SV)	720	742.60	590.43	72	2,503
Hotel Nights ($\times$ 1000)	720	148.99	118.92	11	565
Panel B: Roamers					
Vehicles in accidents	3,600	9.78	13.41	0	92
log(Vehicles in accidents)	3,032	1.79	1.19	0.00	4.52
No trucks	3,600	1.68	2.53	0	20
Single vehicle accidents (SV)	3,600	6.16	9.54	0	71
Hotel Nights ($\times$ 1000)	3,600	22.21	72.31	0	707

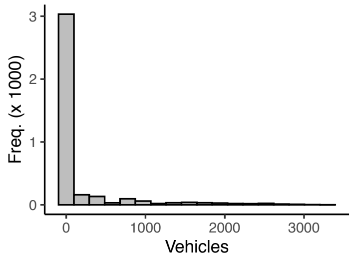
the plausible assumption that the RLAH policy does not induce sorting.[106] Dissimilarities become more relevant when generalizing estimated effects to the untreated population. We discuss the assumptions required to attribute the estimated effect to all drivers in Section 6.5.4.

6.2.2.2 Distribution of accidents

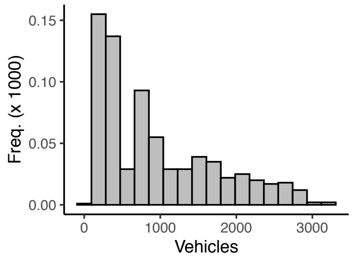
Our dependent variable is the number of vehicles involved in accidents, aggregated by province, month and country group. Table 6.1 presents descriptive statistics for various subsets. Naturally, the mean of the count of vehicles involved in accidents is in levels much larger for locals than for roaming users. In logs, however, the figures are more comparable and the standard deviation is in the same ballpark. Further we find that after controlling for the different mean levels —as we do by including country fixed effects—the treated and control appear to have similar distributions (discussed below).

Figure 6.4 shows histograms of the dependent variable after log transformation and after demeaning for fixed effects. Panels (a)–(c) indicate that these empirical distributions are left-skewed, as to be expected from count data. Similarly, panels (d)–(f) show

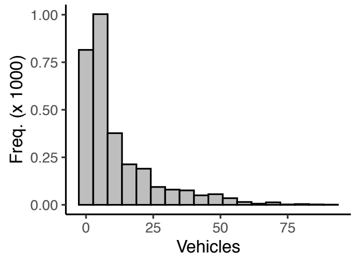
[106]Figure 6.A.3 in Appendix A shows that the age distribution of roaming users does not change considerably after the policy was implemented. We note, however, that even if we find a policy-induced sorting in the distribution of drivers in accidents, this does not necessarily bias our estimates, as it may be a result of the policy e.g. younger drivers may be more likely to use their phone and therefore be more represented in accidents, while the distribution of age groups in kilometres travelled may be the same.



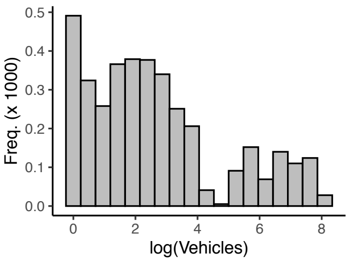
(a) In levels.



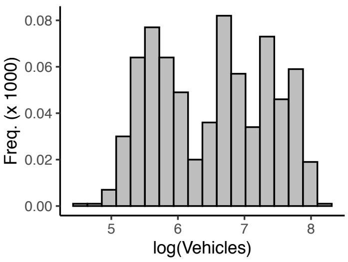
(b) In levels (locals).



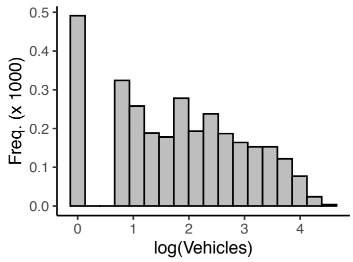
(c) In levels (roaming users).



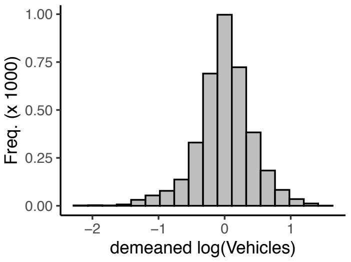
(d) In logs.



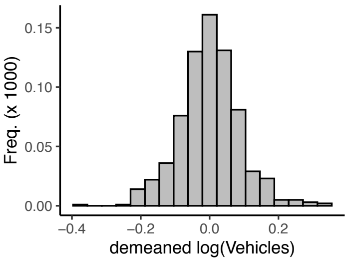
(e) In logs (locals).



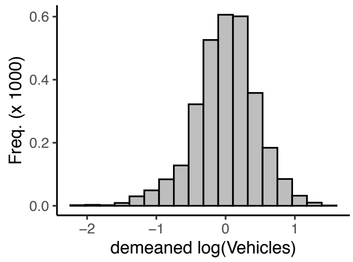
(f) In logs (roaming users).



(g) In logs demeaned.



(h) In logs demeaned (locals).



(i) In logs demeaned (roaming users).

Figure 6.4: Histograms of vehicles per month per province.

that after taking logs of these counts, distributions still seem to be slightly skewed to the left. However, if we demean by our panel and time fixed effects, as in panels (g)–(i), distributions seem quite symmetric, albeit with a larger variance for roaming compared to local users. This is non-problematic, however, when using standard errors that are robust to heteroskedasticity.

6.2.2.3 Hotel nights data as proxy for traffic intensity

An important concern with our approach may be that country specific trends in traffic intensity, or vehicle kilometres travelled (VKT), might drive our results. For example, an increase in tourism over time may result in relatively more VKT by roaming users

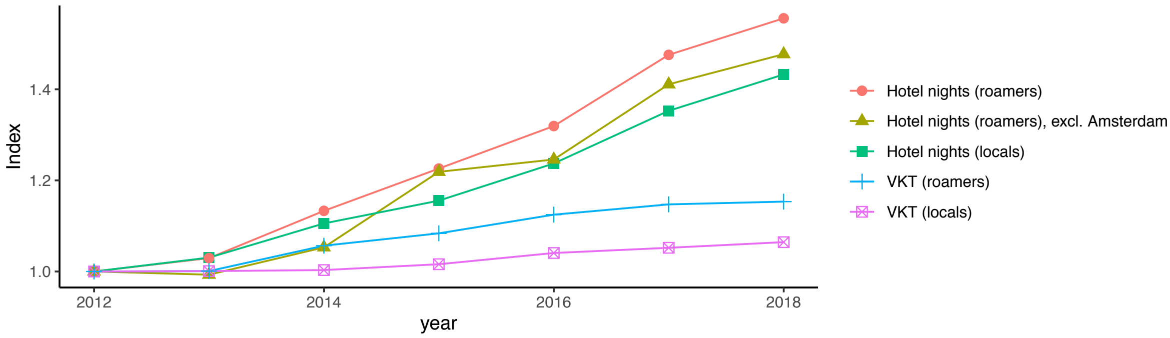


Figure 6.5: Trends in hotel nights and vehicle kilometers

and therefore increase the likelihood of a roaming accident after the introduction of RLAH. We do not observe VKT for each drivers' country at the required level of temporal (monthly) and spatial (province) disaggregation. Instead, we use overnight stays in hotels, obtained from Statistics Netherlands (2019a), to proxy for changes in tourism and thereby monthly traffic intensity. For each province, we observe the number of overnight stays per month, disaggregated into guests' country of origin. We assess the quality of this proxy in two ways.

First, we observe country wide VKT at the annual level for locals and non-locals. Figure 6.5 shows annual growth rates of hotel nights and VKT for local and roaming (non-local) drivers. The figure highlights that over the course of the five years prior to the treatment, VKT by roaming drivers grew more compared to local VKT. However, a similar, yet even stronger trend is visible for hotel nights. Even when we exclude the province containing Amsterdam, an obvious hot spot of growth in hotel nights, we see a similar pattern. This suggests that we can capture trends in VKT with hotel nights, albeit potentially overestimating changes in VKT as it increases less.

Second, we analyse how traffic intensity and the number of vehicles involved in accidents are related to hotel nights for Dutch drivers, for which we observe traffic intensities on highways at the province-month level (Statistics Netherlands, 2019b). Table 6.A.5 in Appendix A shows that, after controlling for time and panel (in this case simply province) fixed effects, there is no statistically significant effect of hotel nights for Dutch nationals with respect to traffic intensity, or number of vehicles involved in accidents. Importantly however, we do find a statistically significant and robust effect for the case of roaming drivers and the number of vehicles involved in accidents. This suggests that hotel nights are a good proxy for country specific changes in VKT from tourism and business related trips. Furthermore, the R^2 in column (2) is 0.99, which indicates that almost all of the variation in the traffic intensity can be explained by our fixed effects, suggesting that group specific changes in traffic intensity are unlikely to effect our estimates.[107]

6.3 The Roam Like at Home Policy

On 27 October 2015, the European Parliament adopted regulation No. 2015/2120 which prescribed that all roaming surcharges should be abolished within the EU.[108] Following a decade of EU roaming regulations which aimed to gradually reduce roaming fees within the EEA, the *Roam Like at Home* (RLAH) policy meant that, effective 15 June 2017, telecommunication network providers were required to abolish all roaming surcharges in addition to domestic retail prices for EU roaming customers.

The policy dramatically reduced the costs of phone use abroad, both compared to the gradual reductions prior to RLAH and compared to the pre-RLAH prices (BEREC, 2019). For example, leading mobile operators such as Vodafone Germany, offered daily roaming packages such as EasyTravel in early May of 2017 providing "phone calls, texting and surfing abroad [within the EU] just like at home" for a price of €2.99 per day. This equates to around €90 per month and is over four times more than standard domestic packages offering calls, texts and data at the time (Vodafone, 2017).[109] The special Eurobarameter (2018) survey, carried out one year after RLAH, suggests that awareness of RLAH was already high with 62% of Europeans that travelled in the previous 12 months being aware that roaming charges had been eliminated, and only 19% of travellers claiming to never use mobile data (down from 42%). Nevertheless, around 50% of the respondents still claim to restrictively use mobile data while abroad, suggesting that EU roaming users still use their mobile phones comparatively less than locals.

To evaluate the effect of RLAH, we collect data on mobile phone usage of roaming users in the EU from the International Roaming BEREC Benchmark Data Reports

[107] Note that we find a borderline significant (significant only at the 10% level) negative estimate for hotel nights of locals in column (8). This might be an indication that drivers who are staying in a hotel, are driving more safely because they are unfamiliar with the area. This would be in line with findings in observational studies. Another possible explanation could come from region specific holidays that vary in timing between years for given regions, and between regions for given years.

[108] Roaming refers to mobile phones connecting to a cellular network abroad. In the absence of regulation, mobile network operators generally charge additional fees for using this service.

[109] Regulated wholesale data rates were capped at €0.05 per MB or €50 per GB, so using data outside of a data bundle may have been restrictively expensive.

and local usage from the Dutch Authority for Consumers and Markets (ACM).[110] Figure 6.6 plots the average monthly data traffic in MB's per roaming user for each quarter between 2012 and 2018, with the shaded region representing when RLAH was active.[111] It indicates that since RLAH was introduced, roaming usage appears to catch up with developments in local cellular data traffic. Roaming usage is still about four times lower than local usage after the policy, but this is a result of the relatively short period of time European tourists spend outside their country of residence (e.g. the average trip duration was about 8.4 nights in 2017 (Eurostat, 2019)). It also shows that cellular roaming traffic exhibits a strong upward growth trend for both groups and demonstrates a high degree of seasonal variation for roaming users. This is not surprising as technological advancements (e.g. introduction of 4G-network) and the increased adoption of smartphones has resulted in higher speeds, lower prices, and more demand, while tourism, and therefore roaming usage, tends to be seasonal. It is therefore useful, when comparing the annual growth rates of cellular traffic, to compare each quarter with the same quarter in the previous year.

Figure 6.7 illustrates that the RLAH policy resulted in a very large increase in the growth rate of phone use of roaming users one year after the policy while having *no discernible effect* on locals. Table 6.A.4 in Appendix A documents the average annual growth rate before and after the policy for roaming users as compared to locals. It indicates a substantial increase in the growth rate of roaming data usage by 200 percentage points, while texts and calls also increased by around 20 to 80 percentage points, relative to locals.[112] This further demonstrates that the policy had large effects on the overall phone use of roaming users, while especially effecting data usage.

6.4 Empirical methods

Our aim is to estimate how smartphone usage affects road safety. Because data on smartphone usage of drivers is privacy sensitive and not made available for research

[110] BEREC only includes information on the number of active roaming users, referred to as roaming subscribers in the BEREC reports, since the second quarter of 2016. BEREC considers a subscriber to be a roaming subscriber if roaming services were active *at least once* in the concerned period. In order to calculate the average monthly usage before this period, we predict the number of subscribers using a log-linear model with a time trend and quarter dummies. Using total data usage gives almost identical results (available from BEREC upon request). We document this in Appendix B.

[111] Note that the second quarter of 2017 already contains 15 days during which the policy was active, namely the second half of June. Furthermore anticipating RLAH, several large network providers dropped roaming charges earlier in the year, such as Vodaphone UK in April (CNET, 2017).

[112] RLAH was not the only roaming policy introduced during the period of study. Other regulations, notably price caps, also resulted in moderate growth in roaming usage, which may explain the increased growth in data around the end of 2014 (BEREC, 2016).

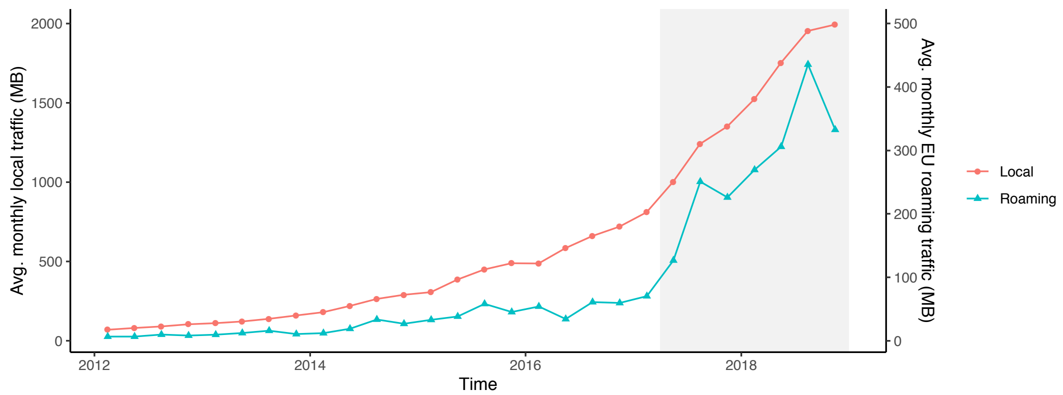


Figure 6.6: Average monthly data traffic per quarter.

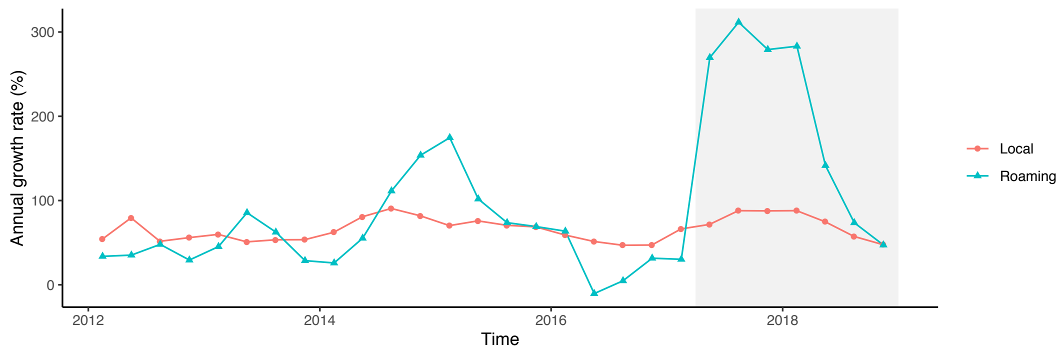


Figure 6.7: Annual growth rates in cellular data traffic per quarter.

purposes, we use the implementation of the RLAH policy as a source of exogenous variation. We hypothesise that a substantial reduction in phone usage fees induces more phone use while driving, which in turn leads to an increase in accidents due to driver distraction. Unique for the RLAH price change, and essential for our identification strategy, is that fees for domestic phone use (i.e. within the home country) are not affected by the policy. This allows us to define a control group, in our case drivers with a Dutch phone subscription, and a treatment group, drivers with a phone subscription from any other EU country. As a consequence, we can employ a difference-in-differences (DiD) approach to estimate the effect of the policy-induced increase in phone use on road safety. Below, we first introduce the general statistical model and subsequently discuss how we deal with the statistical challenges that arise in our setting.

6.4.1 Statistical model

We use a standard DiD approach, where we estimate how the RLAH policy affects the number of vehicles involved in road accidents. We define V_{it} as the number of vehicles involved in accidents for each country-group in each province, indexed by i, at time t.[113] Countries are grouped to strike a balance between optimally controlling for unobserved heterogeneity per country of origin, and preserving statistical power by avoiding zero counts (see Section 6.2.1.2 for more details). We consider the general following model:

$$\log(V)_{it} = \beta T_{it} + \gamma \mathbf{H}_{it} + \delta \mathbf{W}_{it} + \phi_i + \kappa_t + \epsilon_{it}, \tag{6.1}$$

where $\log$ denotes the natural logarithm. The treatment effect, denoted by T_{it}, is a dummy equal to one after the policy was introduced for vehicles from roaming countries. We proxy for traffic intensity using vector $\mathbf{H}_{it}$, which contains separate control variables, in logs, for the number of hotel nights of locals and roaming users, and a dummy in case we observe a zero.[114] Further, vector $\mathbf{W}_{it}$ contains weather controls, that we include to improve the efficiency of the estimator.[115]

Finally, we include panel and time fixed effects. Time-invariant characteristics of drivers and the area in which they drive, such as the road network, attractiveness to tourists, and number of car users, are captured by a country-group province fixed effect, ϕ_i, which represents the panel element in our analysis. We also control for any unobserved time trends affecting all drivers, for instance due to road maintenance or infrastructure improvements, by including a time fixed effect, κ_t, for each year-month.

We note that using a smartphone was rather costly for roaming users before the policy. It might therefore be useful to assume that before the policy roaming drivers did *not* use their phone at all while driving. However, if roaming users *did* use their phones while driving prior to new roaming regulations, we still accurately estimate the effect of the price drop, but underestimate the total effect of phone use. Our es-

[113] Because we essentially have a count model, our temporal and spatial resolutions are arbitrary. We aim for the most fine-grained resolution to maximally use variation over time and space. We are in this respect, constrained by the resolution of the essential control variables. We aggregate at the province-month level because this is the most fine-grained resolution for which we can control for country-specific VKT.

[114] We obtain hotel nights per province per country of origin from Statistics Netherlands (2019a), which is measured in thousands. In case of a zero, which we only observe for roaming countries, we set the value to one (so that the log is zero) and use a dummy to control for these cases separately. This means that we correct for any bias due to inflation at when a zero is reported.

[115] These include for each province and month the average temperature, average rainfall, and number of days with temperatures below 0 °C.

timates should therefore be considered as a lower bound of the total effect of smartphone distractions on road accidents.

6.4.2 Measurement error

Measurement error poses a statistical challenge in our setting, because we do not directly observe within-vehicle phone use, nor the type of phone subscription drivers have. Below, we identify three implications of this challenge and discuss how we deal with them.

First, for multi-vehicle accidents, we cannot identify which driver caused the accident, if any at all. This means that we have measurement error in the dependent variable, which makes our estimates potentially imprecise, albeit still unbiased if the measurement error is random. We address this issue by focussing on vehicles rather than accidents because multi-vehicle accidents might include both treated and control-group drivers. In addition, we also perform a robustness check where we consider a subsample with single-vehicle accidents (e.g. a car crashing into a tree). This approach rules out measurement error of this sort but comes at the cost of having less statistical power, as only a small fraction of the accidents in the data are single-vehicle accidents (17.58%). As it is a priori not possible to decide which is the preferred approach, we report results for both estimation strategies.[116]

Second, some drivers of vehicles that are registered abroad might still have a Dutch phone subscription. For instance, drivers that live in bordering regions in Belgium or Germany and often work in the Netherlands. These drivers will be erroneously classified as treated, and will bias our estimates downwards.[117] To address this issue, we will run a robustness check where we exclude all border provinces, as it is likely that this measurement problem is most pronounced in those regions.

Third, some roaming users may not have to pay smartphone charges themselves. One can think of unlimited subscriptions paid by drivers' employers or having a Dutch subscription while living just across the border. This insensitivity to the price

[116] Another related issue which is solved by taking single-vehicle accidents is that roaming accidents may result in more multi-vehicle accidents. This would violate the SUTVA, but it is unlikely to be problematic in this setting due to the size of the control group; around 95% of vehicles in our accidents sample are part of the control group. Also, we also checked whether the number of vehicles per accidents changes over time, which is not the case, both for accidents with only locals as well as accidents with at least one roaming user involved.

[117] Additionally, some roaming users might be driving a Dutch car, for instance, a rental car, and will hence be erroneously designated as untreated. This may lead to a small downward bias, however, due to a large number of accidents in the control group (local users) it is unlikely to have a substantial effect.

would also result in a downward bias of the estimate. We address this concern in two ways. Firstly, we re-estimate our main model on a sub-sample where we exclude trucks and vans, assuming that drivers of these vehicles are most likely to have such arrangements with their employer, and secondly, on a sub-sample without bordering countries or typical labour migration countries.

6.4.3 Trends in vehicle kilometres travelled

A potential confounding factor is vehicle kilometres travelled (VKT) by roaming drivers. For instance, because countries and provinces vary in their popularity as a holiday destination over time (Taylor and Ortiz, 2009), there may be more roaming accidents due to increased tourism rather than due to increased phone distractions. Another potential reason for temporal variation in VKT by roaming drivers could come from changes in trade and business trips as a result of ongoing globalisation. Because these trends affect treated drivers (e.g. tourists) but not local drivers, it poses a potential threat to our identification strategy and may lead to overstating the effect of phone distraction on road safety.

Ideally, one would want to directly control for VKT to avoid any bias from traffic intensity, but this information is not available.[118] Instead, we show that the number of hotel nights per country of origin is a good proxy for both tourism and business-related traffic (see Section 6.2.2.3 for an extensive discussion on the quality of this proxy). This implies that, if the relation between traffic and hotel nights is stable over time, then controlling for hotel nights will absorb a bias that stems from VKT trends of roaming drivers.[119] Nevertheless, we also perform two additional robustness checks. Firstly, we include a roamer-specific linear time trend which captures nationwide trends in accidents of all roaming users combined. This approach then estimates the policy effect *conditional on* a roaming-user-specific trend in accidents, which provides a lower bound for the estimated effect. This time trend does, of course, also absorbs part of the treatment effect, such that this estimation is only useful to assess our estimates' sensitivity to trends. Secondly, we re-estimate our models using only two years of observations between July 2016 and July 2018 (i.e. one year before and one year after the policy), for which it is implausible that there are major trends in tourism transport modes conditional on hotel nights.

[118]For non-Dutch vehicles, Statistics Netherlands only provides imputed annual figures of VKT for the whole country. For all traffic combined, there are intensity measures available at the province-month level. These will be used to validate our VKT proxy (hotel nights).

[119]This is a reasonable assumption for our five-year study period, but may not hold in the long run (e.g. if cheap flights and high-speed trains make cars a less attractive mode).

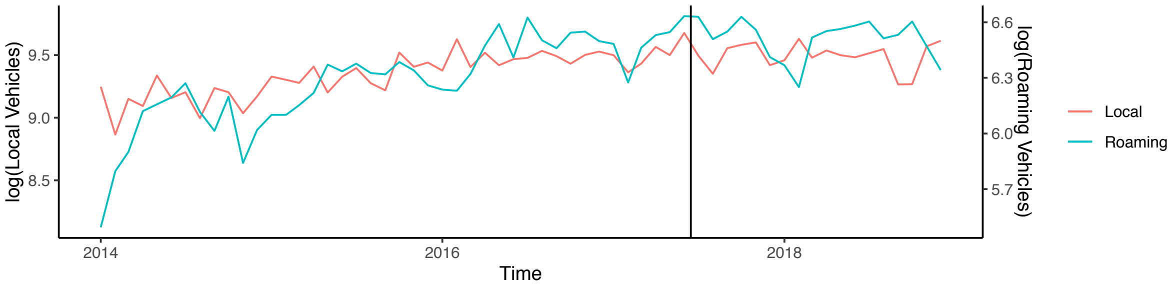


Figure 6.8: Roaming and local user vehicles in accidents.

Figure 6.9: Graphical representation of common trends in aggregated province-month data.

6.4.4 Standard errors

In our setting, the number of observations depends on an arbitrary temporal and spatial resolution. We aggregate vehicle data to province-month observations, to align the resolution with our control variables. However, if accidents are serially correlated, ordinary least squares (OLS) standard errors may be too small (Bertrand et al., 2004). To address this issue, we cluster our standard errors at the time-invariant level of a province and country-group, which leaves us with $12 \times 6 = 72$ clusters (12 provinces and 6 country groups). In addition, we run a robustness check where we ignore all time-series variation and aggregate our data into two periods, one before and one after the policy. This rules out any autocorrelation in error terms, and the outcome highlights that our results and standard errors are hardly affected by serial correlation.

6.5 Results

6.5.1 Parallel trends

We first examine overall trends of local vehicles (control group) and roaming vehicles (treated group) involved in accidents. Figure 6.8 shows that nationwide accident counts for these groups follow similar trends.[120] The figure also highlights that these measures are quite noisy and that no clear jump is observable around the policy introduction in 2017.

[120] In January 2014 there are fewer roaming user accidents, this seems to be a reporting-issue in the data source. Our robustness check where we focus only on one year before and after the policy indicates that this issue does not affect our conclusions.

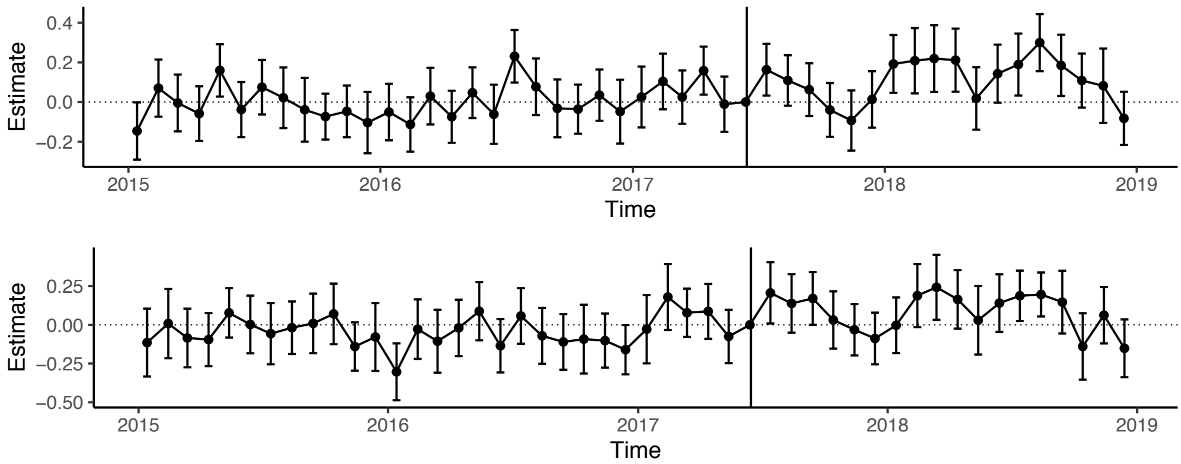


Figure 6.10: Treatment effect per month for full sample (top) and single-vehicles (bottom).

For a more rigorous analysis of a common trend, in Figure 6.10 we plot estimates of a monthly treatment effect, while including all controls and fixed effects as in our preferred specification in (6.1). Here, the coefficients are estimated using an indicator for whether the province-month count of vehicle accidents are for roaming users, interacted with year-month dummies.[121]

The results in Figure 6.10 indicate that no clear pre-trend exists and that local drivers are a suitable control group for roaming drivers, conditional on controls and fixed effects. Furthermore, after the policy, there is a clear positive impact on accidents as indicated by the increased proportion of positive and statistically significant estimates.[122] This pattern is even more pronounced in the bottom panel of the plot, where we focus specifically on single-vehicle accidents.

[121] Specifically, the figure plots the β_τ coefficients from estimating:

$$\log(V_{it}) = \sum_{\tau=-41}^{60} \beta_\tau R_{i,t-\tau} + \gamma \log(H_{it}) + \phi_i + \kappa_t + \epsilon_{it}, \tag{6.2}$$

where $R_{i,t-\tau}$ is an indicator variable for whether the vehicle count is for roaming users or not, interacted with a year-month dummy, and β_τ is the effect of the policy for each year-month t. To be able to include the seasonality fixed effect κ_t in this setting, we omit the treated$\times$year-month dummies for the first full year; otherwise perfect multicollinearity emerges. The error bars represent robust 95% confidence intervals for each monthly point estimate.

[122] 44% of the coefficients are positive and statistically significant post-policy as compared to only 10% pre-policy.

Table 6.2: Main regression results

	log(# Vehicles in Accidents)				
	(1)	(2)	(3)	(4)	(5)
Treatment effect	0.124***	0.125***	0.176***	0.089***	0.094***
	(0.034)	(0.035)	(0.025)	(0.030)	(0.031)
Roamer × trend				0.003**	
				(0.001)	
log(Hotel nights roamers)					0.298***
					(0.076)
log(Hotel nights locals)					-0.089
					(0.063)
Temperature		0.057***	-0.005*	-0.005*	-0.004*
		(0.016)	(0.003)	(0.003)	(0.002)
Rain		0.061	-0.001	-0.001	0.003
		(0.061)	(0.012)	(0.012)	(0.013)
# Frost days		0.157***	0.025*	0.025*	0.019
		(0.040)	(0.013)	(0.013)	(0.012)
Time FE		Yes	Yes	Yes	Yes
Panel FE			Yes	Yes	Yes
Clusters	72	72	72	72	72
Local vehicles	686k	686k	686k	686k	686k
Roaming vehicles	35k	35k	35k	35k	35k
Observations	3,688	3,688	3,688	3,688	3,688
R^2	0.729	0.748	0.965	0.965	0.967

Notes: Column (1) is a basic DiD regression which includes a dummy for roaming user, policy and the interaction between roaming user and policy (denoted treatment effect). Robust standard errors in parentheses are clustered at the province and country-group level. Hotel nights are split into two orthogonal variables for local and roaming users. An additional dummy is included when hotel nights were inflated (only occurs for roaming users). ***, **, * indicate significance at 1%, 5%, and 10%.

6.5.2 Estimation results

Table 6.2 shows the estimation results with incremental levels of controls and fixed effects. Column (1) shows that with only the minimal DiD controls, we find a statistically significant effect of over 12%.[123] Column (2) shows that overall time trends (captured by year×month fixed effects), and weather controls hardly change the estimated treatment effect. In column (3) we add panel fixed effects, where our panel identifier is a province-country group. This increases the point estimates and lowers the standard errors, indicating that these fixed effects improve the efficiency of the estimator and suggests that accident counts are heterogeneous across provinces

[123]Here we run the most simple DiD regression, which includes a dummy for the RLAH policy, a dummy for whether the country group consists of roaming users, and the treatment effect is the interaction between these two dummy variables.

and country-groups. Column (4) shows that the estimated treatment effect declines significantly when we add a linear roaming-specific monthly time trend. This is potentially a bad control that can also pick up part of the treatment effect, but the results here imply that any major nationwide trends in accidents of roaming users only partially affect the results.

Our preferred specification is the one used in column (5), in which we include controls for hotel nights as a proxy for traffic intensity. We find a point estimate of 0.094 with a standard error of 0.031. This implies that the policy-induced increase in phone use leads to an increase in the number of vehicles involved in accidents of 9.91%, with a 95% confidence interval of $3\% - 17\%$. The point estimate declines as compared to (3) and the hotel nights elasticity of roaming users has the expected sign. It indicates that a 1% increase in hotel nights for roaming users is associated with an increase of around 0.3% in the number of vehicles involved in accidents. The hotel nights effect is insignificant for locals, conditional on our set of fixed effects. This makes sense as traffic intensity for roaming users is likely to follow seasonal tourist trends while most local traffic is generated by work commutes and other daily activities. Importantly, fixed effects already absorb overall trends in VKT, heterogeneity across provinces, and heterogeneity across vehicle countries. Therefore, the statistical significance of the hotel nights elasticity, and the fact that the point estimate of the treatment effect is smaller when we include hotel nights, highlights that we indeed capture country-specific long term trends in VKT.

6.5.3 Robustness checks

In this section we perform a vast range of robustness checks. Tables with results are available in Appendix C.

6.5.3.1 Measurement error and endogeneity

One type of measurement error arises because we do not accurately observe which vehicle potentially caused the accident. Table 6.C.1 in Appendix C shows estimation results using different subsets of accident types and vehicle involvement. Columns (1–2) show that focusing on different types of accidents yields very similar results. Excluding trucks and focusing on single-vehicle accidents leads to very similar or only slightly stronger point estimates. Focusing on single-vehicle accidents may suggest we reduce measurement error slightly, but again the point estimates are not statistically different from the main estimate.

As discussed before, our analysis may suffer from measurement error in the treatment assignment, for instance by having a Dutch phone subscription while still driving a non-Dutch car or vice versa. It is likely that measurement error is most pronounced in bordering provinces and for drivers with a close connection to The Netherlands. This can either be due to proximity (like bordering regions or countries) or due to strong economic links (e.g. labour migration). If we exclude bordering countries, we find somewhat larger effects while if we remove bordering provinces or drivers from labour migration countries, we find only slightly smaller effects. Excluding border provinces also mitigates potential concerns that border provinces face more VKT due to the policy, e.g. if people are more likely to go shopping across the border because phone usage is cheaper. Such an endogenous response might induce sorting and thereby poses a threat to our identification strategy. These results indicate that our results do not suffer from a severe downward bias from measurement error.

6.5.3.2 Accounting for VKT trends

So far, we have assumed that country-of-origin specific trends in VKT are well-captured by our hotel nights proxy. Results from Section 6.2 suggest that this is a plausible assumption. Nevertheless, to further rule out any issue with long-term trends in non-local road traffic as a potential confounder, in columns (1–2) of Table 6.C.2, we restrict our sample to one year before and one year after the policy (i.e. from June 2016 to July 2018). This approach yields an estimate of 6.8% for all vehicles and 14% for single-vehicle accidents which are very comparable to our main results. This highlights that long term trends in VKT cannot explain the observed increase in vehicles involved in accidents.

6.5.3.3 Accounting for auto-correlation in error structure

In our main analysis, we use the number of vehicles involved in accidents per province per month as the observational unit. If there is strong serial correlation, then OLS standard errors may be incorrect, even when clustering at a time-invariant level as we do (Bertrand et al., 2004). To deal with this issue in the most conservative way, we re-estimate our main models on data aggregated to pre and post-policy averages.[124] Columns (3–4) in Table 6.C.2 show that the statistical significance is only slightly lower as compared to our main analysis (the t-statistic = 2.1 as compared to 3.1 in our preferred specification). This provides strong evidence that serial correlation does not pose a threat to our statistical inference.

[124] After aggregating, the data represents the log number of vehicles, hotel nights, and weather conditions, by country group and province, for an average month in the pre and post data.

6.5.3.4 Weighting

Our aim is to approximately recover the phone-use effect *per driver*, rather than at a province level. This suggests that we should use sample weights for VKT at the individual level.[125] Because these data are not available on the vehicle accident level, we test the robustness of our results to four weighting schemes that are closely related to VKT.[126] As regions differ in the total number of roaming drivers involved in accidents, this also allows assigning higher weights to provinces that tend to have relatively more roaming drivers and therefore may be more informative. Table 6.C.3 shows that our main results hardly change if we use weights based on 1) roaming accident numbers, 2) total accident numbers, 3) traffic intensity, and 4) hotel nights. This suggests that our fixed effects and log-level specification already sufficiently account for differences in VKT between regions.

6.5.3.5 Accounting for zero counts

In our main analyses, we use a log-linear specification, which performs well with a sufficient number of accidents. However, during some months, for some country-groups, we observe few or even zero vehicles in accidents (14.02% pre and 4.95% post policy). These cases are naturally excluded from our log-linear regressions. However, they might be less likely to occur after the policy due to policy-induced phone distractions. As a consequence, our estimations might suffer from a slight downward bias by excluding more zero counts before than after the RLAH policy introduction for treated vehicles. To test if such a bias exists, we re-estimate our main specification as in (6.1) using a Poisson pseudo-maximum likelihood count model. Table 6.C.4 presents the results from this re-estimation, which allows us to include all province-month observations.[127] The coefficients are remarkably similar and in column (5), our preferred specification with hotel nights, the results indicate that the policy caused 9.4% more accidents and is statistically significant at the 1% level.[128]

[125] Note however that weighting might lead to erroneously small standard errors when there is clustering in the disturbances (Solon et al., 2015). Therefore, as the latter is likely to be the case in our setting, we are cautious with weights and report the more conservative estimates (without weighting) as main results.

[126] Note that for accident numbers we use the time-invariant pre-policy number of roaming and total accidents.

[127] This means we have 4,248 province-month observations as compared to 3,688 in column (5) of Table 6.2.

[128] Column (4) of this specification indicates that it indeed appears that the roamer specific time trend is a bad control, as could be expected.

6.5.3.6 Heterogeneous effects

In addition to the average treatment effect that we estimate in our main analysis, we test for measurable heterogeneity in the effect of phone use, for various subgroups of drivers and road characteristics.

We first test whether the effect size varies by age group. Table 6.C.5 in Appendix C suggests that our main effect predominantly applies to drivers in the age group between 30 and 50. We find statistically insignificant effects for age groups below 30 and above 50. However, as the 95% confidence intervals overlap, we cannot conclusively determine that the effects are statistically different, which might be due to less precision. Lab-based studies also tend to be inconclusive on the performance differences of distracted driving across age groups. Oviedo-Trespalacios et al. (2016) synthesise the most recent literature, and find that although “older drivers tend to engage less in a secondary task like using mobile phones while driving [...], the performance of younger drivers, who are inclined to use a mobile phone while driving, has been reported to be less affected by mobile phone tasks than older drivers” (p. 369). It is therefore not surprising that many studies report a negligible effect of age differences.

We also investigate the treatment effect on different road types. Phone distractions may disproportionately impact the likelihood of causing an accident in more challenging road conditions, such as in urban areas and on local roads where drivers often share the road with other vehicles and modes (e.g. pedestrians and cyclists). To test this hypothesis, we split the sample into three road types based on the speed limit. To assure sufficient statistical power, we define the following three road classes with roughly equal numbers of accidents: below 50 km/h, between 50 km/h and 100 km/h, and above 100 km/h. These groups roughly represent local roads in urban areas, local roads in rural areas, and highways. Similarly, we test whether our estimates are different for vehicles involved in more severe accidents (fatal or injury) versus accidents with only material damage. Results of these estimations are presented in Table 6.3.

Columns (1–3) indicate that most of the estimated effect comes from local roads, and we do not find evidence of a reduction in road safety on highways. This suggests that phone distractions are either more risky on local roads (e.g. due to crossings and traffic lights), or that drivers use their phone less frequently on highways (e.g. because it is perceived as more dangerous).[129]

[129] We cannot fully isolate the effect of phone usage from that of increased car navigation, but the fact that we find only an effect on urban roads may indicate that car navigation does not increase safety in urbanised areas.

Table 6.3: Estimation results using subsamples of road types and severity.

	log(# Vehicles in Accidents)				
	$<$ 50km/h	50km/h - 100km/h	>100km/h	Fatal/Injury	Material
	(1)	(2)	(3)	(4)	(5)
Treatment effect	0.098**	0.050	-0.065	0.115**	0.083**
	(0.038)	(0.040)	(0.049)	(0.054)	(0.033)
log(Hotel nights roamers)	0.178***	0.260***	0.220***	0.130**	0.294***
	(0.058)	(0.078)	(0.045)	(0.051)	(0.075)
log(Hotel nights locals)	-0.031	0.004	-0.125	0.214***	-0.144**
	(0.054)	(0.065)	(0.116)	(0.077)	(0.067)
Weather controls	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes	Yes
Clusters	72	72	72	70	72
Local vehicles	368k	135k	101k	133k	554k
Roaming vehicles	14k	8k	9k	3k	32k
Observations	3,083	2,796	2,818	2,136	3,636
R^2	0.964	0.955	0.934	0.961	0.965

Notes: Robust standard errors in parentheses are clustered at the province and country-group level.***, **, * indicate significance at 1%, 5%, and 10%.

Finally, columns (4–5) indicate that the main result holds, regardless of accident severity, suggesting that mobile phone distractions play an important role in accidents with varying degrees of severity. Our results do not support the hypotheses that phone distractions predominantly increase accidents with material damage, for instance, if people mostly use phones in low-speed, low-risk, situations like traffic jams.

6.5.4 Implications

Our robustness checks indicate that the effect of phone use generally falls within the 95% confidence interval $3\% - 17\%$ of our main estimate. Furthermore, 9.91% is likely to be a conservative estimate of the total effect of phone use because we only estimate the effect induced by the price drop, while roaming users were likely to use their phones, albeit infrequently, prior to the policy. In this section, we calculate the total number of accidents and the relative risk of phone use implied by our main estimate.

6.5.4.1 Total number of accidents caused by phone use

To calculate the number of accidents associated with phone use, we compare the observed number with a counterfactual situation where all drivers face phone usage fees equal to the pre-policy roaming charges. In other words, we consider how many accidents could be avoided if all drivers faced higher phone usage costs and thereby used their phones less. This is a policy-relevant variable because governments can directly affect these costs by, for example, imposing more stringent regulation which increases the costs of being caught using a mobile phone while driving. Importantly, the RLAH policy abolished additional roaming surcharges, such that after the policy, roaming and local users face the same phone use costs and accident risk.[130]

We effectively estimate a local average treatment effect (LATE) of smartphone use while driving, using the RLAH policy as a shifter. The effect is local because some roaming drivers may not comply with the new policy in the sense that they may not increase smartphone usage. If we assume that non-compliance is the same in the treatment and control group, and that the treatment and control group are sufficiently comparable, then we can generalise our LATE to an average treatment effect (ATE) for all road users.

Based on observable driver characteristics, roaming users tend to be younger and drive on faster roads than local drivers (see discussion in Section 6.2.2). Nevertheless, our analysis on heterogeneous effects across age groups suggests that differences in driver age lead to similar results, while highways tend to be safer than local roads with respect to the accident risk of phone distractions (see Table 6.A.3 in Appendix A). Therefore, based on observable characteristics, our estimates may underestimate the ATE because roaming users are more likely to drive on highways.

One remaining concern might be that unobservable driver characteristics, such as familiarity with roads and other infrastructure, make roaming and local drivers not comparable. For instance, if driving on unfamiliar roads increases accidents risk, then this may be further exacerbated by phone distraction. However, Intini et al. (2018) find no clear evidence for increased accident risk due to unfamiliarity with the road network. On the contrary, they find that familiarity is associated with increased accident risk. More research is required to understand the interaction between driver distractions and road familiarity, but at this stage there seems to be no clear indication that our results overestimate the ATE due to road familiarity.

[130] In other words, the RLAH policy caused roaming drivers to 'catch up' with local drivers' smartphone usage and the distractions and associated accident risk. There may still be variations across mobile phone plans and across countries, but these no longer depend on roaming or local use. In addition, these differences are most likely fairly constant over time in the short run and are more related to local demand and supply conditions than to the RLAH policy.

In sum, it seems plausible that our estimate is roughly similar to the ATE. Our results then imply that phone use causes 13,563 additional accidents annually in the Netherlands, of which about 2,536 result in injury and 79 are fatal. Furthermore, if the ATE is applicable to other EU countries, this would imply that around 2,500 road fatalities in the EU in 2018 may be attributable to phone use.[131] As shown in Figure 6.1, the gap between the EU 2020 target and actual fatalities was 28% (7,044 cases). Our results then suggest that around one-third of this gap could be closed by successfully banning mobile phone use while driving.

6.5.4.2 External effect

We do a back-of-the-envelope calculation to determine the share of drivers that got involved in accidents without being distracted themselves. This can be loosely interpreted as a smartphone-induced increased in the external effect of car use. Let us assume that in each accident, just *one* driver was potentially causing the accident due to phone distraction. Then, out of 764k drivers involved in accidents in our data, 334.89k (43.8%) of them were involved in a crash without contributing to the cause of the accident themselves. If we focus on local roads—where we find the strongest effect of distraction—we find a similar figure of 43.9%.

We use these figures to calculate a simple smartphone-induced increase in the external safety effect of car use, expressed in terms of vehicles involved in accidents. Starting with our main estimate of a 9.91% increase in vehicles involved in accidents due to phone distractions, we calculate that in all accidents, on average about 4.1% of vehicles were affected due to distraction of *other* drivers. Note that this calculation crucially hinges on the assumption that in each phone-induced accident, only *one* driver was distracted. This may seem plausible but may be violated in rare cases.

6.5.4.3 Crash risk odds ratio

We follow Bhargava and Pathania (2013) and translate our estimate for the effect of the change in mobile phone use, due to the RLAH policy, on the number of vehicles involved in accidents to the crash risk odds ratio (or 'relative risk') which allows us to compare our results to the existing literature. This requires two key parameters, the percentage of roaming users that are on their phone while driving or the 'baseline prevalence', and the change in phone use due to the policy, denoted by b and c respectively.

[131] This can be calculated by multiplying our main estimate by the total number of fatal vehicle accidents in 2018, so 9.91% $\times$ 25,058 = 2,470.

Observational studies, based on roadside surveys, indicate that average phone use in the car ranges between $1 - 11\%$ (European Road Safety Observatory, 2015).[132] These field studies do not distinguish between roaming and local drivers. However, there is a good reason to expect that the baseline prevalence is overestimated for roaming users because roaming was very costly before RLAH. Therefore we consider a range of b $\in$ $[0.01, 0.10]$, in the sensitivity analysis, but note that lower values are more likely.

As for the increase in phone use due to the policy, Table 6.A.4 suggests that RLAH induced an increase in the annual growth rate of mobile data of around 200 percentage points, and calls and texts of around 80 and 20 percentage points, respectively. We assume that aggregate changes in roaming use also apply to drivers visiting the Netherlands and consider a range of c $\in$ $[0.5, 2]$. It is possible that most of this 200 percentage points increase comes from watching videos and playing songs, which may not (fully) translate to an equivalent increase in distractions while driving. This would imply that the lower values in the specified range for c are more relevant and more applicable to our setting.

Using these parameters, we can calculate a range of possible relative risk factors, denoted by RR, implied by our preferred estimate, $\hat{\beta}$, using the formulation:

$$\hat{\beta}[1 \times (1 - b) + RR \times b] = RR \times bc - bc. \tag{6.3}$$

To reflect the uncertainty of these assumptions, Table 6.4 illustrates how our key parameters influence the implied RR estimates. It indicates that RR is decreasing in the baseline prevalence and in the change in phone use due to the policy. In other words, if the policy had a small impact on phone use and roaming drivers used their phone very little prior to the policy, our estimate implies larger risks associated with phone use while driving.

That said, we take a conservative estimate for the baseline prevalence of 3% and the change in phone use due to the policy of 100%. This would imply a relative risk of phone use of 3.8.[133] We consider this to be a conservative estimate as it is unlikely that roaming drivers used their phones as intensively as local drivers due to the high pre-policy roaming costs.

Comparing these estimates to the existing literature suggests that our conservative estimate of the crash risk associated with modern smartphone usage is similar to

[132] Based on a naturalistic driving setting between 2012 and 2015, Dingus et al. (2016) observe handheld cell phone prevalence in the US to be about 6.3%. There is no reason to expect that prevalence is substantially different in the Netherlands, and therefore, we expect that the findings in European Road Safety Observatory (2015) capture a meaningful range for our study.

[133] Re-arranging terms, we can find $RR = \frac{\hat{\beta} - \hat{\beta}b + bc}{b(\hat{\beta} + c)}$. Plugging in $b = 0.03$ and $c = 1$ gives: $RR = 3.8$.

Table 6.4: Sensitivity of implied accident risk.

	Baseline prevalence, b				
Δ phone use due to RLAH, c	1%	2%	3%	5%	10%
50%	17.10	8.90	6.20	4.00	2.30
80%	11.70	6.30	4.40	3.00	1.90
100%	9.80	5.30	3.80	2.60	1.70
150%	7.00	4.00	2.90	2.10	1.50
200%	5.60	3.30	2.50	1.80	1.40

Notes: This table presents the relative accident risk implied by our baseline estimate from column (5) in Table 6.2. The relative risk is calculated by re-arranging equation (6.3) such that: $RR = \frac{\hat{\beta}-\hat{\beta}b+bc}{b(\hat{\beta}+c)}$. Baseline prevalence reflects the percentage of time roaming drivers spend on the phone while in the car. An illustration is outlined in the text.

earlier crash-based studies, but are significantly larger than recent field studies.[134] This suggests that the crash risks of phone use are slightly lower in magnitude than those found for positive levels of blood alcohol.[135] As mentioned earlier, previous research focuses mainly on the effects of calling, or focuses on specific road types and phone use, however, modern smartphones offer substantially more usability and potential for distraction, and our findings suggest that these effects are more likely to be present on local urban roads. Our estimates for the change in mobile phone use due to the RLAH policy suggest that we mainly pick up an effect from using more mobile data (increase in the growth rate of about 200 percentage points as compared to local drivers) which may explain why we find larger implied relative risk estimates than some earlier field studies.

6.6 Conclusion

In this study, we provide novel evidence on the effect of cell phone use on car accidents. We exploit variation in the cell phone usage fees in the Netherlands following the *Roam Like at Home* (RLAH) policy introduced by the European Union (EU) in 2017. This intervention is used as a treatment, and applies to roaming users—non-Dutch drivers from the EU—, which allows us to employ a difference-in-differences approach.

[134] Redelmeier and Tibshirani (1997) find a RR of about 4.3, Dingus et al. (2016) find the RR of cell phone use to be 3.6, and Bhargava and Pathania (2013) do not find any effect. Hersh et al. (2019) do not calculate the RR, however their main estimate of 1.1% is far lower than our main estimate of 9.91%.

[135] Levitt and Porter (2001a) finds a crash risk of 7 and 13 for positive levels of blood alcohol and illegal levels respectively.

We show that the growth rate of mobile calls, texts, and particularly data usage increased substantially after the change in roaming regulations, making roaming phone use more in line with usage in home countries. While we do not directly observe actual phone use of drivers, the observed increase in usage is likely to (partly) carry over to phone use while driving. We estimate that decreased smartphone usage fees lead to an increase in the number of vehicles involved in accidents of 9.91% (95% confidence interval $3\% - 17\%$). This is likely to be an underestimate of the *total effect* of phone use while driving, as our estimates capture the effect of an *increase* in smartphone use, which was not fully absent before the policy.

Under the assumption that the identified mechanism carries over to all EU drivers, our estimate implies that, in 2018, around 2,500 road fatalities in the EU could be attributed to phone use. Our results then suggest that around one-third of the gap between realised safety improvements on roads and the EU 2020 target can be attributed to mobile phone use.[136]

Our findings indicate that the existing literature may underestimate the risks associated with modern smartphone usage while driving. Our main result implies a crash risk odds ratio associated with mobile phone use of around 3.8, which is likely to be a conservative estimate. All in all, our results suggest that smartphones are making roads less safe, and this has important implications for road safety policies.

Our paper provides an estimate of the *average* effect of smartphone usage on the number of vehicles involved in traffic accidents, which may conceal considerable differences between specific groups of drivers. We look into heterogeneous effects by estimating models for different sub-samples (e.g. for different age groups, or excluding trucks). Future research could delve into this further, by estimating propensities of specific groups of drivers to use their phone while driving. Ride-hailing drivers, for example, may have a relatively high propensity to be distracted by their phone, which might be an important factor in explaining the results of Barrios et al. (2020), who find that ride-hailing services increased the number of traffic accidents in the US. Such evidence could provide valuable input for related regulation and policies.

[136]In 2018, the EU was 28% away from their 2020 target (see Figure 6.1).

Appendix 6.A Additional descriptives

Table 6.A.1: Descriptive statistics: Vehicles in accidents.

Statistic	N	Mean	St. Dev.	Min	Max
Roaming	764,065	0.046	0.210	0	1
Age	561,136	42.488	15.015	0.000	110.000
Female	764,065	0.455	0.707	0	10
Maximum speed (km)	653,055	63.726	26.823	15.000	130.000
Deadly	764,065	0.006	0.076	0	1
Injury	764,065	0.187	0.390	0	1
Material	764,065	0.807	0.394	0	1

Table 6.A.2: Descriptive statistics by group: Vehicles in accidents.

Variable	Roaming	Local	Diff	Tstat
Age	40.903	42.566	-1.663	18.998
Female	0.383	0.459	-0.075	21.007
Maximum speed (km)	74.511	63.200	11.312	-62.898
Deadly	0.006	0.006	-0.000	0.503
Injury	0.088	0.192	-0.103	65.301
Material	0.906	0.802	0.104	-63.736

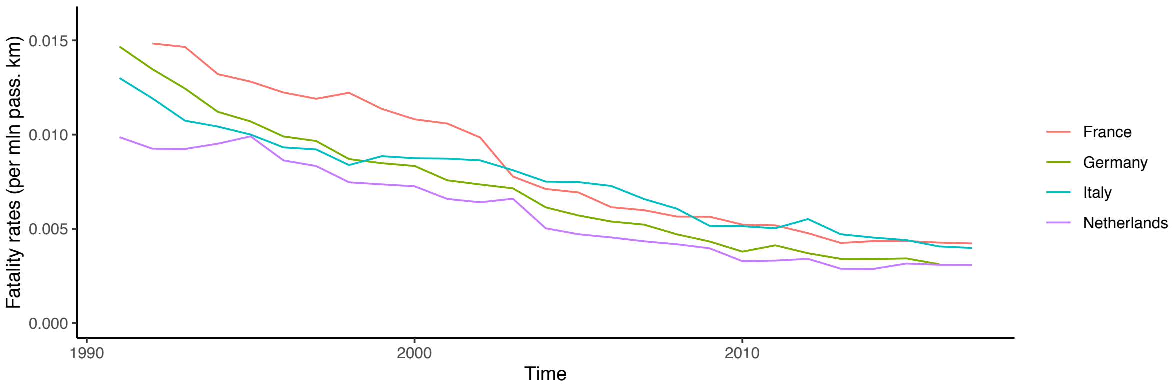


Figure 6.A.1: Fatality rates in road accidents over time in major EU countries and the Netherlands

Table 6.A.3: Relative frequencies of road types by severity

	Local roads	Major roads	Highways
Fatal/injury	14.1 %	4.3 %	2.0 %
Material damage	46.8 %	17.9 %	14.8 %
Total	60.9 %	22.2 %	16.9 %

Table 6.A.4: Difference-in-differences in annual growth of phone use.

		Annual growth rate (%)		Δ Annual growth rate (p.p.)	
Usage	User	Pre	Post	Diff	DiD
Calls	Local	4.29	-2.49	-6.78	
Calls	Roaming	-1.06	71.16	72.21	78.99
Data	Local	67.05	83.82	16.76	
Data	Roaming	68.09	285.89	217.80	201.04
Texts	Local	-18.24	-1.80	16.45	
Texts	Roaming	-22.01	18.37	40.38	23.93

Notes: Pre-policy refers to the the average annual growth rates of cellular traffic comparing each quarter with the same quarter in the previous year, over three years (Q1 2014 − Q1 2017) prior to the implementation of RLAH. Post-policy is one year, Q2 2017 − Q1 2018, after RLAH.

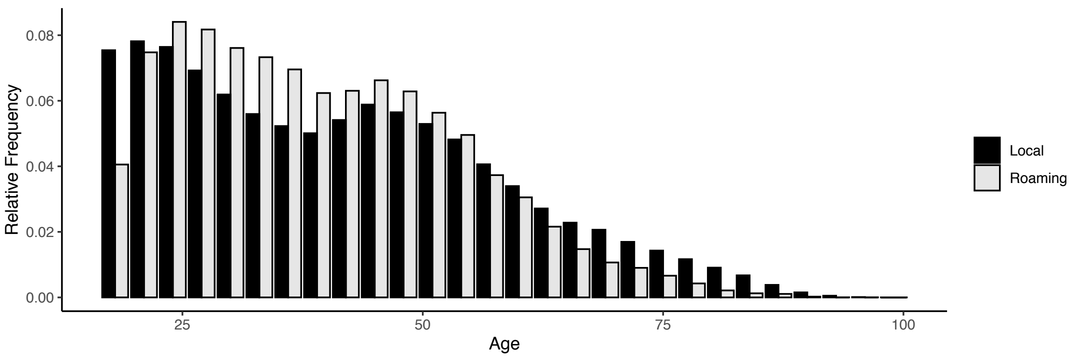


Figure 6.A.2: Age of local and roaming users.

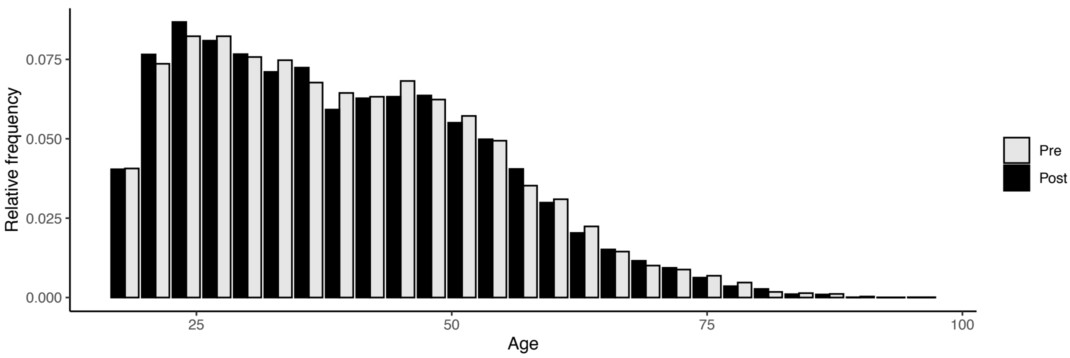


Figure 6.A.3: Age of roaming users pre and post policy.

6.A.1 Analysis of hotel nights as proxy for vehicle kilometres travelled

Table 6.A.5: Regression results for analysing traffic and hotel nights for Dutch drivers.

	log(Traffic intensity)		log(# Vehicles in accidents)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Hot. loc.)	0.241**	0.043	0.787***	0.114			0.629***	-0.132*
	(0.108)	(0.065)	(0.117)	(0.084)			(0.070)	(0.075)
log(Hot. roam.)					0.288***	0.322***	0.303***	0.311***
					(0.086)	(0.077)	(0.082)	(0.078)
Time FE		Yes		Yes		Yes		Yes
Panel FE		Yes		Yes		Yes		Yes
Subsample	Loc.	Loc.	Loc.	Loc.	Roam.	Roam.	All	All
Within R2	0.233	0.017	0.660	0.025	0.241	0.046	0.806	0.047
Observations	564	564	564	564	2,980	3,688	3,688	3,688
R^2	0.233	0.992	0.660	0.990	0.241	0.966	0.806	0.966

Notes: Robust standard errors in parentheses are clustered at the province and country-group level.***, **, * indicate significance at 1%, 5%, and 10%.

Appendix 6.B Predicting smartphone use per subscriber pre-2016

We obtain roaming usage data from the EU Body of European Regulators for Electronic Communications (BEREC). Their reports include the time series "EEA average consumption per month per total number of roaming subscribers (in GB)" from the second quarter of 2016 onwards. Therefore, in order to get a better picture of the long term changes in roaming data usage, we use data on the total "EEA Retail data traffic (millions of GB)" (available as of 2007) and predict the number of subscribers in earlier periods using a simple model. The advantage of this approach is that the number of subscribers appears to follow a rather simple dynamic process and means that we only need to predict the denominator. We can then also compare the growth in our metric to the total growth in mobile data use which gives us more confidence that the predictions are as close as possible to actual figures.

We observe quarterly data on the number of roaming subscribers from the second quarter of 2016 until the first quarter of 2019. The top panel in Figure 6.B.1 indicates that the number of subscribers appears to follow a somewhat log-linear growth trend with a strong seasonal pattern which is likely related to summer tourism. We therefore estimate the number of subscribers using the following regression equation:

$$\log(S_t) = \gamma \text{Trend}_t + \phi_{q(t)} + \epsilon_t, \tag{6.B.1}$$

where $\log(S_t)$ is the natural logarithm of the number of subscribers, Trend_t is a linear time trend capturing the growth over time, and $\phi_{q(t)}$ are quarter dummies that capture seasonal variations. The resulting model has an $R^2 = 0.92$, which suggests that it captures the vast share of roaming subscriber dynamics. This is further confirmed by the bottom panel of Figure 6.B.1 which compares the actual and predicted number of subscribers and the resulting calculation of data roaming per subscriber. Finally, Figure 6.B.2 compares the difference between growth in roaming data per subscriber and the total roaming data use. While the trends are almost identical, it indicates a larger growth in total data use which is likely a result of capturing overall trends in growth in subscribers (which is relatively constant) and may also be a result of the RLAH policy that causes the number of people actively using roaming while travelling to increase. Overall, it suggests that the predicted change in data usage is a conservative estimate of the effect of the policy.

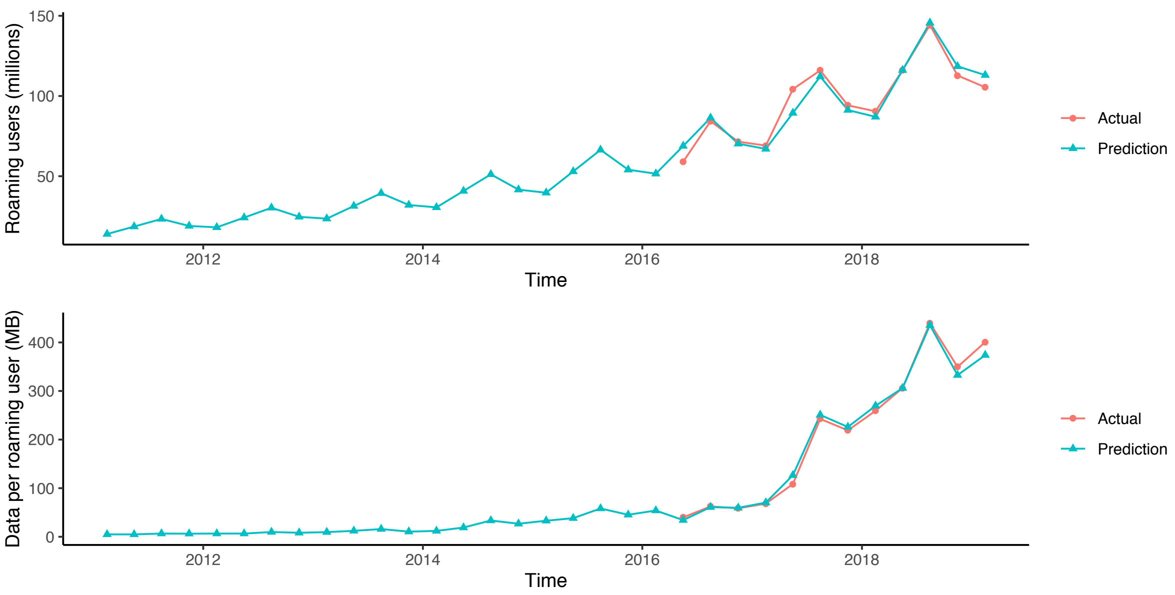


Figure 6.B.1: Predicting number of EU roaming subscribers and data consumption

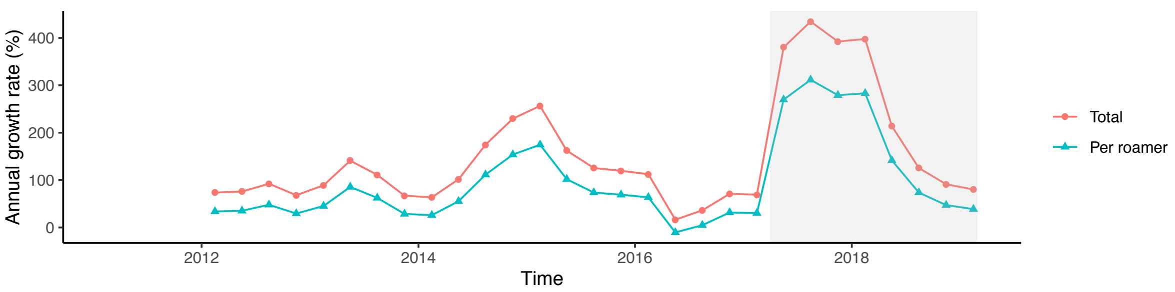


Figure 6.B.2: Growth in roaming data usage per subscriber as compared to the total

Appendix 6.C Robustness checks and sensitivity analyses

Table 6.C.1: Results correcting for sources of measurement error

	log(# Vehicles in Accidents)				
	No trucks	SV	No border prov.	No border countr.	No BG/PL/RO
	(1)	(2)	(3)	(4)	(5)
Treatment effect	0.086**	0.096***	0.072*	0.134***	0.077**
	(0.034)	(0.033)	(0.037)	(0.030)	(0.034)
log(Hotel roam.)	0.297***	0.147**	0.424***	0.132***	0.335***
	(0.093)	(0.056)	(0.121)	(0.037)	(0.084)
log(Hotel loc.)	-0.117*	-0.167***	0.087	-0.038	-0.071
	(0.061)	(0.054)	(0.082)	(0.072)	(0.064)
Weather controls	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes	Yes
Clusters	72	72	30	48	60
Local vehicles	535k	136k	356k	686k	686k
Roaming vehicles	22k	6k	12k	15k	26k
Observations	3,303	2,658	1,523	2,458	3,026
R^2	0.966	0.962	0.968	0.977	0.972

*Notes:*Robust standard errors in parentheses are clustered at the province and country-group level.***, **, * indicate significance at 1%, 5%, and 10%.

Table 6.C.2: Results using only one year pre/post (1–2), and data aggregated to two periods (3–4).

	log(# Vehicles in accidents)			
	All	Single vehicle	All	Single vehicle
	(1)	(2)	(3)	(4)
Treatment effect	0.065**	0.131***	0.148**	0.162*
	(0.029)	(0.036)	(0.070)	(0.091)
log(Hotel nights roamers)	0.316***	0.192***	0.114	-0.082
	(0.094)	(0.064)	(0.094)	(0.109)
log(Hotel nights locals)	0.036	-0.053	0.148	-0.037
	(0.069)	(0.078)	(0.424)	(0.409)
Weather controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes
Clusters	72	72	72	72
Local vehicles	319k	59k	23k	4k
Roaming vehicles	18k	3k	1k	0k
Observations	1,593	1,162	144	143
R^2	0.969	0.962	0.999	0.998

Notes: Robust standard errors in parentheses are clustered at the province and country-group level. Columns (1–2) are obtained using data from June 2016 until July 2018. Columns (3–4) are obtained after aggregating the data into two periods, one before the policy and one after. For interpretation purposes, after aggregation, variables are then rescaled to their initial units (e.g. monthly averages).***, **, * indicate significance at 1%, 5%, and 10%.

Table 6.C.3: Regression results using weighted least squares.

	log(# Vehicles in Accidents)			
	(1)	(2)	(3)	(4)
Treatment effect	0.111***	0.128***	0.110***	0.101***
	(0.030)	(0.033)	(0.028)	(0.037)
log(Hotel nights roamers)	0.196***	0.185***	0.255***	0.256***
	(0.061)	(0.057)	(0.059)	(0.092)
log(Hotel nights locals)	-0.212***	-0.173*	-0.083	-0.221*
	(0.078)	(0.094)	(0.066)	(0.118)
Weather controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes
Weights	Total veh.	Roaming veh.	Avg traf. intens.	Avg hotel nights
Clusters	72	72	72	72
Local vehicles	686k	686k	686k	686k
Roaming vehicles	35k	35k	35k	35k
Observations	3,688	3,688	3,688	3,688
R^2	0.970	0.969	0.966	0.968

Notes: Estimated using weighted least squares, with pre-policy total number of (roaming) vehicles as weights. Robust standard errors in parentheses are clustered at the province and country-group level.***, **, * indicate significance at 1%, 5%, and 10%.

Table 6.C.4: Estimation results using Poisson regression.

	# Vehicles in Accidents				
	(1)	(2)	(3)	(4)	(5)
Treatment effect	0.186***	0.186***	0.187***	0.0381	0.0899***
	(0.0234)	(0.0210)	(0.0223)	(0.0262)	(0.0298)
Roamer × trend				0.00520***	
				(0.000921)	
log(Hotel nights roamers)					0.335***
					(0.0709)
log(Hotel nights locals)					-0.000314
					(0.0689)
Temperature		0.0366	-0.00100	-0.00101	-0.00108
		(0.0264)	(0.000852)	(0.000851)	(0.000850)
Rain		0.198**	0.0134***	0.0132***	0.0139***
		(0.0788)	(0.00447)	(0.00443)	(0.00452)
# Frost days		0.0571	-0.000380	-0.000526	-0.000896
		(0.0631)	(0.00301)	(0.00297)	(0.00352)
Time FE	No	Yes	Yes	Yes	Yes
Panel FE	No	No	Yes	Yes	Yes
Clusters	72	72	72	72	72
Local vehicles	686	686	686	686	686
Roaming vehicles	35	35	35	35	35
Observations	4,248	4,248	4,248	4,248	4,248

Notes: Robust standard errors in parentheses are clustered at the province and country-group level.***, **, * indicate significance at 1%, 5%, and 10%.

Table 6.C.5: Estimation results for subsamples with different age groups.

	log(# Vehicles in Accidents)					
	All	Age ≤ 30	$30 <$Age< 50	Age≥ 50	Age≥ 65	Age unknown
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment effect	0.094***	0.039	0.099**	0.058	0.055	0.193**
	(0.031)	(0.038)	(0.039)	(0.035)	(0.051)	(0.087)
log(Hotel nights roamers)	0.298***	0.201***	0.200***	0.204***	0.027	0.181**
	(0.076)	(0.039)	(0.050)	(0.064)	(0.049)	(0.072)
log(Hotel nights locals)	-0.089	0.058	-0.046	0.016	0.063	-0.238
	(0.063)	(0.070)	(0.080)	(0.069)	(0.096)	(0.166)
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	72	72	72	72	62	72
Local vehicles	686k	189k	221k	172k	59k	104k
Roaming vehicles	35k	7k	12k	6k	1k	9k
Observations	3,688	2,638	3,072	2,572	1,422	2,822
R^2	0.967	0.959	0.954	0.959	0.970	0.924

Notes: Robust standard errors in parentheses are clustered at the province and country-group level.***, **, * indicate significance at 1%, 5%, and 10%.

7
Summary and conclusions

This thesis provides a collection of empirical studies in environmental and transport economics. It contains five chapters in which natural experiments are employed to quantify mechanisms in the realm of air pollution, road safety, congestion, and substitution in urban transport networks.

Chapter 2 tested the hypothesis that particulate matter has a direct effect on human decision making. As a natural experiment, this study focused on whether PM increases the probability of drawing in chess games using information from the Dutch club competition. The results provide evidence for a reasonably strong effect: A 10μg increase in PM_{10} (33.6% of mean concentration) leads to a 5.8% increase in draws. This chapter has demonstrated that air pollution causes individuals to take less risk.

Chapter 3 estimated how air pollution in general, and ambient ozone in particular, affects human physical activity through impaired lung functioning. The study provides evidence of the immediate impact of air pollution on time delays in urban outdoor activities. This effect is estimated on cycling speeds in London using several estimation strategies. The results show that ozone reduces speed for concentrations above 20 ppb, which is far below the minimum threshold suggested by other studies. A 10 ppb increase in ozone concentration leads to a 0.3-0.4% reduction in cycling speed, despite that most cycling trips are short so that exposure to ozone tends to be short. It seems plausible that ozone induces time losses of similar magnitude of other outdoor activities, such as walking.

Chapter 4 studied how public transport and cycling are related in a dense urban network. Specifically, it focused on how demand for public rental bicycles is affected by local and temporary metro interruptions in Paris. A unique dataset is constructed by linking metro interruptions announced in Twitter communication by the Parisian metro operator to usage data on the Vélib' PBS. The results show that, as a direct consequence of a metro interruption, the consumption of bicycles within 100 metres of metro stations increases by 0.72 bicycles per hour per docking station on average, and with 1.54 bicycles per hour per docking station during the first 20 minutes; an increase of approximately 11% and 22% respectively. Due to their effects on demand, metro interruptions increase the probability of empty stocks at docking stations with 15%. The findings highlight that cycling is a local net substitute for metro service, and that public rental bicycles can alleviate time losses stemming from interruptions in public transport.

Chapter 5 focused on New York City and estimates the causal effect of protected bike lanes on traffic speed and flow, and road safety. Because a street-level analysis is prone to biases due to treatment-induced rerouting, aggregated observations of streets in the same direction within narrowly defined areas on Manhattan are used. Thereby, Manhattan's elongated shape and grid-structured street network are exploited to assure causality, but also to obtain policy-relevant area-level estimates. Bike lanes are found to improve safety for cyclists both on streets and at junctions. Once an area can be completely traversed on a protected bike lane, accidents with cyclists involved are estimated to be reduced by 34%. The results further indicate that bike lanes have no statistically significant effect on overall road safety at junctions, but reduce accidents away from junctions by 59% for *all modes* in the whole area.

Using taxi trips as an accurate proxy for traffic indicators, bike lanes appear to have no statistically significant effect on traffic speed of traffic flow at the area-direction level. However, traffic speed on streets with bike lanes is 1.3% lower compared to streets in the same direction in the same area. At the same time, the results indicate that streets with a bike lane accommodate a 2.3% higher throughput.

Chapter 6 investigates to what extent smartphones play a role in the number of road accidents. The study exploits variation in phone usage fees in the Netherlands following the European Union (EU) roaming regulations in 2017, which abolished all roaming surcharges for EU residents. This change is used to estimate a difference-in-differences model where non-Dutch drivers from the EU are treated, while Dutch drivers serve as a control group. Phone use patterns show that the growth rate of mobile calls, texts, and particularly data usage increased substantially after the change in roaming regulations, making roaming phone use more in line with usage in home countries. While actual phone use of drivers is not directly observed, the in overall phone use is likely to (partly) carry over to phone use while driving. The results then suggest that 10% of accidents can be explained by the use of smartphones. The findings further indicate that phone distraction increases accidents of *all* severity levels by a similar magnitude, and that phone-related accidents mainly happen on local urban roads.

In addition to each chapter's estimation results, this thesis has demonstrated how one can use natural experiments in cases where traditional experiments are infeasible. For instance, it would be unethical to run an experiment where cyclists or chess players are exposed to air pollution by a researcher. At the same time, the vast majority of the world population is exposed to air pollution, such that it is highly relevant to understand its impact on health and behaviour. Furthermore, the behavioural effects of air pollution are often modest. Therefore, even *if* society would consider (lab) experiments with air pollution acceptable, then the subtlety of the impact of air pollution requires a large study population. Running such large scale experiments would make traditional experiments very costly, such that natural experiments are an attractive method to improve our understanding of the impact of air pollution.

To study road safety, similar ethical concerns as above make it difficult to run (field) experiments. Especially in the case of smartphone distraction of drivers, traditional experiments are infeasible. This is because ethical concerns (researchers would put participants in danger) are further exacerbated by the fact that the treatment (driving while holding a smartphone) is prohibited. Therefore, most researchers in this field use simulations to study the effects of driver distraction. However, with simulation approaches, one cannot estimate the full extent of the problem. Hence, a natural experiment is the only feasible method for analysing the aggregate effect of smartphone distraction of drivers.

Finally, because new policies often induce variation, they reveal underlying mechanisms and can be exploited as natural experiments. The installation of bike lanes in New York City is a prime example of such a policy. In that case, analysing observational traffic data not only allows inferring causal effects, but it allows for an evaluation of the policy itself. In a world where evidence-based policy-making is increasingly important, this thesis, therefore, also serves as an example of how policies can be evaluated and used as natural experiments.

Overall, this thesis provides a collection of studies based on natural experiments. It demonstrates that one can address important research questions in environmental and transport economics using observational data. There are many more applications possible, and data availability and computing power will arguably further increase. Therefore, natural experiments are likely to be an essential method for future research in a broad range of scientific fields.

Bibliography

Abouk, R. and S. Adams (2013). Texting bans and fatal accidents on roadways: do they work? Or do drivers just react to announcements of bans? *American Economic Journal: Applied Economics 5*(2), 179–99.

Adams, S. and C. Cotti (2008). Drunk driving after the passage of smoking bans in bars. *Journal of Public Economics 92*(5-6), 1288–1305.

Adams, W. C., K. A. Brookes, and E. S. Schelegle (1987). Effects of NO_2 alone and in combination with O_3 on young men and women. *Journal of Applied Physiology 62*(4), 1698–1704.

Adham, M. T. and P. J. Bentley (2015). An ecosystem algorithm for the dynamic redistribution of bicycles in London. In *International Conference on Information Processing in Cells and Tissues*, pp. 39–51. Springer.

Adler, M. W. and J. N. van Ommeren (2016). Does public transit reduce car travel externalities? Quasi-natural experiments' evidence from transit strikes. *Journal of Urban Economics 92*, 106–119.

Ahlfeldt, G. M. and E. Pietrostefani (2019). The economic effects of density: A synthesis. *Journal of Urban Economics 111*, 93–107.

Anenberg, S. C., J. Miller, R. Minjares, L. Du, D. K. Henze, F. Lacey, C. S. Malley, L. Emberson, V. Franco, Z. Klimont, et al. (2017). Impacts and mitigation of excess diesel-related NO_x emissions in 11 major vehicle markets. *Nature 545*(7655), 467.

Aragón, F. M., J. J. Miranda, and P. Oliva (2017). Particulate matter and labor supply: The role of caregiving and non-linearities. *Journal of Environmental Economics and Management 86*, 295–309.

Arceo, E., R. Hanna, and P. Oliva (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City. *The Economic Journal 126*(591), 257–280.

Archsmith, J., A. Heyes, and S. Saberian (2018). Air quality and error quantity: Pollution and performance in a high-skilled, quality-focused occupation. *Journal of the Association of Environmental and Resource Economists 5*(4), 827–863.

Barbosa, C. M. G., M. Terra-Filho, A. L. P. de Albuquerque, D. Di Giorgi, C. Grupi, C. E. Negrao, M. U. P. B. Rondon, D. G. Martinez, T. Marcourakis, F. A. dos Santos, et al. (2012). Burnt sugarcane harvesting–cardiovascular effects on a group of healthy workers, Brazil. *PloS one 7*(9).

Barrios, J. M., Y. Hochberg, and H. Yi (2020). The cost of convenience: Ridehailing and traffic fatalities. Technical report, National Bureau of Economic Research.

Beach, B. and W. W. Hanlon (2018). Coal smoke and mortality in an early industrial economy. *The Economic Journal 128*(615), 2652–2675.

Berdica, K. (2002). An introduction to road vulnerability: what has been done, is done and should be done. *Transport Policy 9*(2), 117–127.

BEREC (2016). International Roaming BEREC Benchmark Data Report April – September 2015. Body of European Regulators for Electronic Communications. Retrieved from `https://berec.europa.eu/files/document_register_store/2016/3/BoR(16)28-1_Roaming_Benchmark_Data_Report_04-09_2015.pdf`.

BEREC (2019). Roam Like at Home One year on. Body of European Regulators for Electronic Communications. Retrieved from `https://www.europarl.europa.eu/cmsdata/157554/Jeremy%20Godfrey%20-%20BEREC-%20Roam%20Like%20at%20Home%20-%20One%20year%20on.pdf`.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

Bhargava, S. and V. S. Pathania (2013). Driving under the (cellular) influence. *American Economic Journal: Economic Policy 5*(3), 92–125.

Blincoe, L., T. R. Miller, E. Zaloshnja, and B. A. Lawrence (2015). The economic and societal impact of motor vehicle crashes, 2010 (revised). National Highway Traffic Safety Administration, Washington, DC (2015). Publication No. DOT HS 812 013. Available at `http://www-nrd.nhtsa.dot.gov/Pubs/812013.pdf`.

Bondy, M., S. Roth, and L. Sager (2019). Crime is in the air: The contemporaneous relationship between air pollution and crime. *Journal of the Association of Environmental and Resource Economists*.

Borck, R. (2019). Public transport and urban pollution. *Regional Science and Urban Economics 77*, 356–366.

Borck, R. and P. Schrauth (2019). Population density and urban air quality. Technical report, CESifo Group Munich.

Borck, R. and T. Tabuchi (2018). Pollution and city size: can cities be too small? *Journal of Economic Geography 19*(5), 995–1020.

Bordagaray, M., L. dell'Olio, A. Fonzone, and Á. Ibeas (2016). Capturing the conditions that introduce systematic variation in bike-sharing travel behavior using data mining techniques. *Transportation Research Part C: Emerging Technologies 71*, 231–248.

Börjesson, M. and J. Eliasson (2012). The value of time and external benefits in bicycle appraisal. *Transportation Research Part A: Policy and Practice 46*(4), 673–683.

Brands, D., J. Klingen, and F. Ostermeijer (2020). Hands on the wheel, eyes on the phone: the effect of smart phone usage fees on road safety. *Tinbergen Institute Discussion Paper 2020-024/VIII*.

Bullock, C., F. Brereton, and S. Bailey (2017). The economic contribution of public bike-share to the sustainability and efficient functioning of cities. *Sustainable Cities and Society 28*, 76–87.

Burger, N. E., D. T. Kaffine, and B. Yu (2014). Did california's hand-held cell phone ban reduce accidents? *Transportation Research Part A: Policy and Practice 66*, 162–172.

Burkhardt, J., J. Bayham, A. Wilson, E. Carter, J. D. Berman, K. O'Dell, B. Ford, E. V. Fischer, and J. R. Pierce (2019). The effect of pollution on crime: Evidence from data on particulate matter and ozone. *Journal of Environmental Economics and Management 98*, 102267.

Cakmak, S., R. Dales, J. Leech, and L. Liu (2011). The influence of air pollution on cardiovascular and pulmonary function and exercise capacity: Canadian health measures survey (chms). *Environmental Research 111*(8), 1309–1312.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: methods and applications*. Cambridge University Press.

Carozzi, F. and S. Roth (2020). Dirty density: Air quality and the density of american cities. Technical report, IZA Discussion Papers.

Carslaw, D. C. and S. D. Beevers (2004). Investigating the potential importance of primary NO_2 emissions in a street canyon. *Atmospheric Environment 38*(22), 3585–3594.

Celis-Morales, C. A., D. M. Lyall, P. Welsh, J. Anderson, L. Steell, Y. Guo, R. Maldonado, D. F. Mackay, J. P. Pell, N. Sattar, et al. (2017). Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study. *BMJ 357*, j1456.

Chang, T., J. Graff Zivin, T. Gross, and M. Neidell (2016). Particulate pollution and the productivity of pear packers. *American Economic Journal: Economic Policy 8*(3), 141–69.

Chang, T. Y., J. Graff Zivin, T. Gross, and M. Neidell (2019a). The effect of pollution on worker productivity: evidence from call center workers in china. *American Economic Journal: Applied Economics 11*(1), 151–72.

Chang, T. Y., J. Graff Zivin, T. Gross, and M. Neidell (2019b). The effect of pollution on worker productivity: evidence from call center workers in china. *American Economic Journal: Applied Economics 11*(1), 151–72.

Chang, T. Y., W. Huang, and Y. Wang (2018). Something in the air: Pollution and the demand for health insurance. *The Review of Economic Studies 85*(3), 1609–1634.

Chappie, M. and L. Lave (1982). The health effects of air pollution: a reanalysis. *Journal of Urban Economics 12*(3), 346–376.

Chew, S. H., W. Huang, and X. Li (2019). Does haze cloud decision making? a natural laboratory experiment. *A Natural Laboratory Experiment (July 20, 2019)*.

City of New York (2014). City of New York. Vision Zero Action Plan. Available at: `http://www.nyc.gov/html/visionzero/pdf/nyc-vision-zero-action-plan.pdf`.

City of New York (2020a). Bicycle routes . Available at: `https://data.cityofnewyork.us/Transportation/Bicycle-Routes/7vsa-caz7`.

City of New York (2020b). Motor vehicle collisions - crashes. Available at: `https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95`.

City of New York (2020c). Nyc street centerline (cscl) . Available at:

https://data.cityofnewyork.us/City-Government/NYC-Street-Centerline-CSCL-/exjm-f27b.

City of New York (2020d). TLC Trip Record Data . Available at: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

City of New York (2020e). Traffic volume counts. Available at: https://data.cityofnewyork.us/browse?q=Traffic+Volume+Counts.

CNET (2017, April). Vodafone roaming now free in 40 countries. Retrieved from https://www.cnet.com/news/vodafone-roaming-now-free-in-40-countries/ on January 17, 2020.

Cohen, A. and L. Einav (2003). The effects of mandatory seat belt laws on driving behavior and traffic fatalities. *Review of Economics and Statistics 85*(4), 828–843.

Côme, E. and L. Oukhellou (2014). Model-based count series clustering for bike sharing system usage mining: a case study with the vélib'system of paris. *ACM Transactions on Intelligent Systems and Technology (TIST) 5*(3), 39.

Correia, S., P. Guimarães, and T. Zylkin (2019). Ppmlhdfe: Fast poisson estimation with high-dimensional fixed effects. *arXiv preprint arXiv:1903.01690*.

Couture, V., G. Duranton, and M. A. Turner (2018). Speed. *Review of Economics and Statistics 100*(4), 725–739.

Cox, A., F. Prager, and A. Rose (2011). Transportation security and the role of resilience: A foundation for operational metrics. *Transport Policy 18*(2), 307–317.

de Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review forthcoming*.

De Hartog, J. J., H. Boogaard, H. Nijland, and G. Hoek (2010). Do the health benefits of cycling outweigh the risks? *Environmental Health Perspectives*, 1109–1116.

De Vos, J. (2020). The effect of covid-19 and subsequent social distancing on travel behavior. *Transportation Research Interdisciplinary Perspectives*, 100121.

DeAngelo, G. and B. Hansen (2014). Life and death in the fast lane: Police enforcement and traffic fatalities. *American Economic Journal: Economic Policy 6*(2), 231–57.

DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation 12*(4), 3.

Dingus, T. A., F. Guo, S. Lee, J. F. Antin, M. Perez, M. Buchanan-King, and J. Hankey (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences 113*(10), 2636–2641.

Duflo, E. and A. Banerjee (2011). *Poor Economics: A radical rethinking of the way to fight global poverty*. Public Affairs.

Ebenstein, A., V. Lavy, and S. Roth (2016). The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics 8*(4), 36–65.

Edlin, A. S. and P. Karaca-Mandic (2006). The accident externality from driving. *Journal of Political Economy 114*(5), 931–955.

Eurobarameter (May 2018). The end of roaming charges one year later. European Comission. Retrieved from https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Survey/getSurveyDetail/instruments/FLASH/search/roaming%20charges/surveyKy/2192.

European Commission (2019a, January). Handbook on the External Costs of Transport. Retreived from https://ec.europa.eu/transport/themes/sustainable/studies/sustainable_en.

European Commission (2019b, November 20). Road Safety Facts & Figures. Retreived from https://ec.europa.eu/transport/road_safety/road-safety-facts-figures-1_en.

European Environmental Agency (2019). Air quality in Europe. https://www.eea.europa.eu/ds_resolveuid/E1MB45UNT6.

European Road Safety Observatory (2015). Cell phone use while driving. Retrieved from https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/erso-synthesis-2015-cellphone-detail_en.pdf.

Eurostat (2019, May). Tourism trips - introduction and key figures. Retreived on 16 January 2020 from https://ec.europa.eu/eurostat/statistics-explained/index.php/Tourism_trips_-_introduction_and_key_figures.

Ewing, J. (2018). 10 Monkeys and a Beetle: Inside VW's campaign for 'clean diesel'. *New York Times January 25, 2018*.

Faccio, M. and J. J. McConnell (2020). Death by pokémon go: The economic and human cost of using apps while driving. *Journal of Risk and Insurance 87*(3), 815–849.

Fann, N., A. D. Lamson, S. C. Anenberg, K. Wesson, D. Risley, and B. J. Hubbell (2012). Estimating the national public health burden associated with exposure to ambient PM2. 5 and ozone. *Risk Analysis 32*(1), 81–95.

Farias, F. and H. ApSimon (2006). Relative contributions from traffic and aircraft NO_x emissions to exposure in West London. *Environmental Modelling & Software 21*(4), 477–485.

Folinsbee, L. J., W. F. McDonnell, and D. H. Horstman (1988). Pulmonary function and symptom responses after 6.6-hour exposure to 0.12 ppm ozone with moderate exercise. *Japca 38*(1), 28–35.

Fosgerau, M. (2005). Speed and income. *Journal of Transport Economics and Policy 39*(2), 225–240.

Freedman, D. (1999). From association to causation: some remarks on the history of statistics. *Journal de la société française de statistique 140*(3), 5–32.

Goldberg, M. S., R. T. Burnett, J. Brook, J. C. Bailar III, M.-F. Valois, and R. Vincent (2001). Associations between daily cause-specific mortality and concentrations of ground-level ozone in Montreal, Quebec. *American Journal of Epidemiology 154*(9), 817–826.

Gong, H. J., P. W. Bradley, M. S. Simmons, and D. P. Tashkin (1986). Impaired exercise performance and pulmonary function in elite cyclists during low-level ozone exposure in a hot environment. *American Review of Respiratory Disease 134*(4), 726–733.

Gössling, S. and A. S. Choi (2015). Transport transitions in copenhagen: Comparing the cost of cars and bicycles. *Ecological Economics 113*, 106–113.

Graff Zivin, J. and M. Neidell (2013). Environment, health, and human capital. *Journal of Economic Literature 51*(3), 689–730.

Graff Zivin, J. and M. Neidell (2018). Air pollution's hidden impacts. *Science 359*(6371), 39–40.

Gu, J., B. Mohit, and P. A. Muennig (2017). The cost-effectiveness of bike lanes in New York City. *Injury prevention 23*(4), 239–243.

Hamanaka, R. B. and G. M. Mutlu (2018). Particulate matter air pollution: effects on the cardiovascular system. *Frontiers in Endocrinology 9*, 680.

Hamilton, T. L. and C. J. Wichman (2018). Bicycle infrastructure and traffic congestion: Evidence from DC's Capital Bikeshare. *Journal of Environmental Economics and Management 87*, 72–93.

Hamra, G. B., F. Laden, A. J. Cohen, O. Raaschou-Nielsen, M. Brauer, and D. Loomis (2015). Lung cancer and exposure to nitrogen dioxide and traffic: a systematic review and meta-analysis. *Environmental Health Perspectives 123*(11), 1107.

Han, Y., E. Côme, and L. Oukhellou (2014). Towards bicycle demand prediction of large-scale bicycle sharing system. *Transportation Research Board 93rd Annual Meeting* (14-2637).

Haque, M. M. and S. Washington (2015). The impact of mobile phone distraction on the braking behaviour of young drivers: a hazard-based duration model. *Transportation Research Part C: Emerging Technologies 50*, 13–27.

He, J., N. Gouveia, and A. Salvo (2019). External effects of diesel trucks circulating inside the Sao Paulo megacity. *Journal of the European Economic Association 17*(3), 947–989.

Herrnstadt, E., A. Heyes, E. Muehlegger, and S. Saberian (2016). Air pollution as a cause of violent crime: Evidence from Los Angeles and Chicago. *Manuscript in preparation*.

Hersh, J. S., B. J. Lang, and M. Lang (2019). Digitally distracted at the wheel: Car accidents and smartphone coverage. *Available at SSRN 3469824*.

Heyes, A., M. Neidell, and S. Saberian (2016). The effect of air pollution on investor behavior: Evidence from the s&p 500. Technical report, National Bureau of Economic Research.

Heyes, A. and S. Saberian (2019). Temperature and decisions: evidence from 207,000 court cases. *American Economic Journal: Applied Economics 11*(2), 238–65.

Honey-Roses, J., I. Anguelovski, J. Bohigas, V. Chireh, C. Daher, C. Konijnendijk,

J. Litt, V. Mawani, M. McCall, A. Orellana, et al. (2020). The impact of covid-19 on public space: a review of the emerging questions.

Huang, J., N. Xu, and H. Yu (2017). Pollution and performance: Do investors make worse trades on hazy days? *Available at SSRN 2846165*.

IFSTTAR (2017). Vélos en libre service (vls) and stats. `http://vlsstats.ifsttar.fr/`. Accessed -Januari-2017.

Intini, P., N. Berloco, P. Colonna, V. Ranieri, and E. Ryeng (2018). Exploring the relationships between drivers' familiarity and two-lane rural road accidents. A multi-level study. *Accident Analysis & Prevention 111*, 280–296.

Jacobsen, M. R. (2011). Fuel economy, car class mix, and safety. *American Economic Review 101*(3), 105–09.

Jans, J., P. Johansson, and J. P. Nilsson (2018). Economic status, air quality, and child health: Evidence from inversion episodes. *Journal of Health Economics 61*, 220–232.

JCDecaux (2008). Jcdecaux schemes: Vélib paris, sevici seville, (fr / es). Technical report.

Jensen, P., J.-B. Rouquier, N. Ovtracht, and C. Robardet (2010). Characterizing the speed and paths of shared bicycle use in Lyon. *Transportation Research Part D: transport and environment 15*(8), 522–524.

Kager, R., L. Bertolini, and M. Te Brömmelstroet (2016). Characterisation of and reflections on the synergy of bicycles and public transport. *Transportation Research Part A: Policy and Practice 85*, 208–219.

Kahn, M. E. and P. Li (2019). The effect of pollution and heat on high skill public sector worker productivity in China. Technical report, National Bureau of Economic Research.

Kelly, F., H. Anderson, B. Armstrong, R. Atkinson, B. Barratt, S. Beevers, D. Derwent, D. Green, I. Mudway, and P. Wilkinson (2011). The impact of the congestion charging scheme on air quality in London. part 1. emissions modeling and analysis of air pollution measurements. *Research Report (Health Effects Institute) 155*, 5–71.

Klingen, J. (2019). Do metro interruptions increase the demand for public rental bicycles? Evidence from Paris. *Transportation Research part A: policy and practice 123*, 216–228.

Klingen, J. and J. van Ommeren (2020a). Risk attitude and air pollution: Evidence from chess. Technical report, Tinbergen Institute.

Klingen, J. and J. van Ommeren (2020b). Urban air pollution and time losses: Evidence from cyclists in london. *Regional Science and Urban Economics 81*, 103504.

Kondo, M. C., C. Morrison, E. Guerra, E. J. Kaufman, and D. J. Wiebe (2018). Where do bike lanes work best? a bayesian spatial model of bicycle lanes and bicycle crashes. *Safety science 103*, 225–233.

Koster, H. R., I. Pasidis, and J. van Ommeren (2019). Shopping externalities and retail concentration: Evidence from Dutch shopping streets. *Journal of Urban Economics* (Forthcoming).

Krupa, S. V. and W. J. Manning (1988). Atmospheric ozone: formation and effects on vegetation. *Environmental Pollution 50*(1-2), 101–137.

Künn, S., J. Palacios, and N. Pestel (2019). The impact of indoor climate on human cognition: Evidence from chess tournaments.

Lelieveld, J. and F. J. Dentener (2000). What controls tropospheric ozone? *Journal of Geophysical Research: Atmospheres 105*(D3), 3531–3551.

Levitt, S. D. and J. Porter (2001a). How dangerous are drinking drivers? *Journal of Political Economy 109*(6), 1198–1237.

Levitt, S. D. and J. Porter (2001b). Sample selection in the estimation of air bag and seat belt effectiveness. *Review of Economics and Statistics 83*(4), 603–615.

Levy, J. I., S. M. Chemerynski, and J. A. Sarnat (2005). Ozone exposure and mortality:" an empiric bayes metaregression analysis". *Epidemiology*, 458–468.

Li, H., J. Cai, R. Chen, Z. Zhao, Z. Ying, L. Wang, J. Chen, K. Hao, P. L. Kinney, H. Chen, et al. (2017). Particulate matter exposure and stress hormone levels: a randomized, double-blind, crossover trial of air purification. *Circulation 136*(7), 618–627.

Li, H., D. J. Graham, and P. Liu (2017). Safety effects of the London cycle superhighways on cycle collisions. *Accident Analysis & Prevention 99*, 90–101.

Lichter, A., N. Pestel, and E. Sommer (2017). Productivity effects of air pollution: Evidence from professional soccer. *Labour Economics 48*, 54–66.

Lipovac, K., M. Đerić, M. Tešić, Z. Andrić, and B. Marić (2017). Mobile phone use while driving-literary review. *Transportation Research Part F: Traffic Psychology and Behaviour 47*, 132–142.

London Air (2019). London air quality network - King's College London. Available at: "http://www. londonair.org.uk".

Lu, J. G. (2019). Air pollution: A systematic review of its psychological, economic, and social effects. *Current opinion in psychology*.

Mangrum, D. and A. Molnar (2018). The marginal congestion of a taxi in New York City. *Processed, Vamderbilt University*.

Marie de Paris (2016). Le bilan des deéplacements 2015 en à Paris. Technical report.

Marshall, W. E. and N. W. Garrick (2011). Evidence on why bike-friendly cities are safer for all road users. *Environmental Practice 13*(1), 16–27.

Martin, P. E., D. Rothstein, and D. D. Larish (1992). Effects of age and physical activity status on the speed-aerobic demand relationship of walking. *Journal of Applied Physiology 73*(1), 200–206.

Martin-Moral, E. and A. Fonzone (2017). Bike share usage characterization-initial results from a cluster analysis of London Santander cycles scheme data. Technical report.

Mattsson, L.-G. and E. Jenelius (2015). Vulnerability and resilience of transport systems–a discussion of recent research. *Transportation Research Part A: Policy and Practice 81*, 16–34.

McEvoy, S. P., M. R. Stevenson, A. T. McCartt, M. Woodward, C. Haworth, P. Palamara, and R. Cercarelli (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *British Medical Journal 331*(7514), 428.

Metro bike (2017). The bike-sharing world at the end of 2016. http://bike-sharing.blogspot.nl/2017/01/the-bike-sharing-world-at-end-of-2016.html. [Online; accessed -July-2017].

Neidell, M. (2017). Air pollution and worker productivity. *IZA World of Labor 363*, 1–10.

Netherlands National Institute for Public Health and the Environment (2019). Luch - gevalideerde data. Bilthoven, Netherlands. retrieved from `https://www.rivm.nl/lucht/gevalideerde-data`.

Newzoo (2018). Top countries by smartphone users. Global Mobile Market Report. Retrieved from `https://newzoo.com/insights/rankings/top-countries-by-smartphone-penetration-and-users/` on Jan 14, 2020.

Nilsson, P., H. Stigson, M. Ohlin, and J. Strandroth (2017). Modelling the effect on injuries and fatalities when changing mode of transport from car to bicycle. *Accident Analysis & Prevention 100*, 30–36.

Noland, R. B. and K. A. Small (1995). *Travel-time uncertainty, departure time choice, and the cost of the morning commute*. Institute of Transportation Studies, University of California, Irvine.

NYC DoT (2020). New York City - department of transportation Available at: `https://www1.nyc.gov/html/dot/html/home/home.shtml`.

Oviedo-Trespalacios, O., M. M. Haque, M. King, and S. Washington (2016). Understanding the impacts of mobile phone distraction on driving performance: A systematic review. *Transportation Research Part C: Emerging Technologies 72*, 360–380.

Padgham, M. (2019, 2). dodgr: An R package for network flow aggregation. *Transport Findings*.

Prud'homme, R., M. Koning, L. Lenormand, and A. Fehr (2012). Public transport congestion costs: The case of the Paris subway. *Transport Policy 21*, 101–109.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Randriamanamihaga, A. N., E. Côme, L. Oukhellou, and G. Govaert (2014). Clustering the vélib dynamic origin/destination flows using a family of poisson mixture models. *Neurocomputing 141*, 124–138.

RATP (2017). Trafic annuel entrant par station du réseau ferré 2016. Technical report.

Redelmeier, D. A. and R. J. Tibshirani (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine 336*(7), 453–458.

Regan, K. W. and G. M. Haworth (2011). Intrinsic chess ratings. *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 834–839.

Rice, M. B., P. L. Ljungman, E. H. Wilker, K. S. Dorans, D. R. Gold, J. Schwartz, P. Koutrakis, G. R. Washko, G. T. O'Connor, and M. A. Mittleman (2015). Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the Framingham heart study. *American Journal of Respiratory and Critical Care Medicine 191*(6), 656–664.

Rietveld, P. (2000). The accessibility of railway stations: the role of the bicycle in The Netherlands. *Transportation Research Part D: Transport and Environment 5*(1), 71–75.

Royal Netherlands Meteorological Institute (2019). Daggegevens van het weer in nederland. De Bilt, Netherlands. Retrieved from `https://www.knmi.nl/nederland-nu/klimatologie/daggegevens` on November 08, 2019.

Sager, L. (2019). Estimating the effect of air pollution on road safety using atmospheric temperature inversions. *Journal of Environmental Economics and Management 98*, 102250.

Schepers, J. and E. Heinen (2013). How does a modal shift from short car trips to cycling affect road safety? *Accident Analysis and Prevention 50*, 1118–1127.

Shaheen, S. A. (2012). Public bikesharing in North America: Early operator and user understanding, mti report 11-19.

Sillman, S. (1999). The relation between ozone, NO_x and hydrocarbons in urban and polluted rural environments. *Atmospheric Environment 33*(12), 1821–1845.

Snow, J. (1855). *On the mode of communication of cholera*. John Churchill, London.

Solon, G., S. J. Haider, and J. M. Wooldridge (2015). What are we weighting for? *Journal of Human resources 50*(2), 301–316.

Statistics Netherlands (2019a). Hotels; gasten, overnachtingen, woonland, regio. Den Haag, Netherlands. Retrieved from `https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82061NED/table?dl=2ABB8` on July 08, 2019.

Statistics Netherlands (2019b). Verkeersintensiteit; rijkswegen, regio's. Den Haag, Netherlands. Retrieved from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83712NED/table?dl=F974 on July 08, 2019.

Studenski, S., S. Perera, K. Patel, C. Rosano, K. Faulkner, M. Inzitari, J. Brach, J. Chandler, P. Cawthon, E. B. Connor, et al. (2011). Gait speed and survival in older adults. *Jama 305*(1), 50–58.

Taylor, T. and R. A. Ortiz (2009). Impacts of climate change on domestic tourism in the UK: A panel data estimation. *Tourism Economics 15*(4), 803–812.

Transport for London (2016). Travel in London report 9. content.tfl.gov.uk/travel-in-london-report-9.pdf.

Transport for London (2018a). Santander cycles customer satisfaction and usage survey casual users only: Quarter 2 2017/18. http://content.tfl.gov.uk/santander-cycles-casuals-css-q2-2017-18.pdf.

Transport for London (2018b). Santander cycles customer satisfaction and usage survey members only: Quarter 2 2017/18. http://content.tfl.gov.uk/santander-cycles-members-css-q2-2017-18.pdf.

Transport for London (2018c). Travel in London report 11. content.tfl.gov.uk/travel-in-london-report-11.pdf.

Transport for London (2019). Transport for London cycling data. Available at: http://cycling.data.tfl.gov.uk/.

Van Ommeren, J. and J. Dargay (2006). The optimal choice of commuting speed: consequences for commuting time, distance and costs. *Journal of Transport Economics and Policy 40*(2), 279–296.

Vélib' (2017). Paris á velib: Subscriptions and fees. http://en.velib.paris.fr/Subscriptions-and-fees. [Online; accessed -July-2017].

Verhoef, E. T. and J. Rouwendal (2004). A behavioural model of traffic congestion: endogenizing speed choice, traffic safety and time losses. *Journal of Urban Economics 56*(3), 408–434.

Verstraeten, W. W., J. L. Neu, J. E. Williams, K. W. Bowman, J. R. Worden, and K. F. Boersma (2015). Rapid increases in tropospheric ozone production and export from China. *Nature geoscience 8*(9), 690.

Vodafone (2017, May 1). Easytravel: Calls, surf & text abroad. Retreived from https://web.archive.org/web/20170501112657/http://www.vodafone.de/privat/service/easytravel.html#tab_2BE0EAD972634A95B7A45D491C093743.

von Ferber, C., B. Berche, T. Holovatch, and Y. Holovatch (2012). A tale of two cities. *Journal of Transportation Security 5*(3), 199–216.

Wall, S. P., D. C. Lee, S. G. Frangos, M. Sethi, J. H. Heyer, P. Ayoung-Chee, and C. J. DiMaggio (2016). The effect of sharrows, painted bicycle lanes and physically protected paths on the severity of bicycle injuries caused by motor vehicles. *Safety 2*(4), 26.

Wang, D., T. Feng, and C. Liang (2008). Research on bicycle conversion factors. *Transportation Research Part A: Policy and Practice 42*(8), 1129–1139.

Wang, T., L. Xue, P. Brimblecombe, Y. F. Lam, L. Li, and L. Zhang (2017). Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. *Science of the Total Environment 575*, 1582–1596.

Wang, X. (2017). An empirical study of the impacts of ambient temperature on risk taking. *Psychology 8*(07), 1053.

Wang, X., G. Lindsey, J. E. Schoner, and A. Harrison (2015). Modeling bike share station activity: effects of nearby businesses and jobs on trips to and from stations. *Journal of Urban Planning and Development 142*(1), 04015001.

WHO (2011). Mobile phone use: A growing problem of driver distraction.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Nelson Education.

World Bank (2019). Mobile cellular subscriptions (per 100 people). Data retrieved from World Telecommunication/ICT Development Report and database, https://data.worldbank.org/indicator/IT.CEL.SETS.P2.

World Health Organization (2006). *Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide*. World Health Organization.

Wuerzer, T. and S. G. Mason (2016). Retail gravitation and economic impact: A market-driven analytical framework for bike-share station location analysis in the united states. *International Journal of Sustainable Transportation 10*(3), 247–259.

Young, K. L., C. M. Rudin-Brown, C. Patten, R. Ceci, and M. G. Lenné (2014). Effects of phone type on driving and eye glance behaviour while text-messaging. *Safety science 68*, 47–54.

Zhao, N., B. Reimer, B. Mehler, L. A. D'Ambrosio, and J. F. Coughlin (2013). Self-reported and observed risky driving behaviors among frequent and infrequent cell phone users. *Accident Analysis & Prevention 61*, 71–77.

Zivin, J. G. and M. Neidell (2012). The impact of pollution on worker productivity. *The American Economic Review 102*(7), 3652–3673.

Samenvatting (Dutch summary)

Dit proefschrift is een collectie van empirische studies in het domein van milieueconomie en transporteconomie. In vijf hoofdstukken worden natuurlijke experimenten gebruikt om mechanismen te kwantificeren op het gebied van luchtverontreiniging, verkeersveiligheid, verkeersdoorstroming, en substitutie in stedelijke vervoersnetwerken.

Hoofdstuk 2 test de hypothese dat fijnstof een direct effect heeft op individuele besluitvorming. Als een natuurlijk experiment, richt deze studie zich op de vraag of fijnstof de kans op remise verhoogt in schaakpartijen, gebruik makend van informatie van de Nederlandse clubcompetitie. De resultaten tonen een redelijk sterk effect: een toename van 10μg fijnstof (met een diameter kleiner dan 10 micrometer) leidt tot een toename van 5.8% in remises. Dit hoofdstuk toont daarmee aan dat luchtverontreiniging ervoor kan zorgen dat mensen minder risico nemen.

Hoofdstuk 3 toont aan dat luchtverontreiniging (ozon) de fysieke capaciteit van mensen negatief beïnvloedt. Dit effect is gevonden door fietssnelheden van miljoenen deelfietsgebuikers in Londen te analyseren. Uit de resultaten blijkt dat ozon de snelheid van fietsers verlaagt voor concentraties vanaf 20 ppb. Dit is een stuk lager dan in eerdere studies is gevonden, en ook lager dan de huidige milieustandaarden. Een toename van de ozonconcentratie met 10 ppb leidt tot een afname van fietssnelheden met 0.3-0.4%. Het lijkt aannemelijk dat ozon bij andere buitenactiviteiten, zoals wandelen, tot tijdverlies van vergelijkbare omvang leidt.

Hoofdstuk 4 onderzoekt de samenhang tussen openbaar stadsvervoer en deelfietsen. Specifiek wordt onderzocht hoe de vraag naar deelfietsen wordt beïnvloed door tijdelijke storingen in het metronetwerk van Parijs. De studie koppelt Twitterberichten over metrostoringen aan gebruiksgegevens van Vélib' deelfietsen. De resultaten tonen aan dat als een metro (tijdelijk) uitvalt, het aantal gebruikte deelfietsen binnen 100 meter van metrostations toeneemt met 11%, en met 22% gedurende de eerste 20 minuten. De metrostoring verhoogt de kans op een leeg deelfietsstation met 15%. Deze bevindingen impliceren dat deelfietsen een netto substituut zijn voor metrodiensten, en dat deelfietsen het tijdverlies als gevolg van storingen in het openbaar vervoer kunnen verminderen.

Hoofdstuk 5 richt zich op New York City en schat het effect van afgescheiden fietspaden op verkeersveiligheid, snelheid, en volume. Voor verkeersveiligheid wordt het stratennetwerk gekoppeld aan meer dan een miljoen geregistreerde verkeersongevallen. Het effect op de volume van het verkeer is in kaart gebracht met de data van een half miljard taxiritten. Fietspaden blijken de veiligheid voor fietsers te verbeteren, zowel op straten als op kruispunten. Zodra een gebied volledig kan worden doorkruist op een beschermd fietspad, wordt het aantal ongevallen met fietsers naar schatting met 34% verminderd. De resultaten tonen bovendien aan dat de verkeersveiligheid ook verbetert voor alle andere weggebruikers, met uitzondering van de verkeersveiligheid op kruispunten.

De analyse van taxiritten laat zien dat fietspaden geen significant effect hebben op de verkeerssnelheid en verkeersvolume op wijkniveau, maar op straatniveau is er een effect voor straten met fietspaden. De verkeerssnelheid op straten met fietspaden ligt 1.3% lager dan op straten zonder fietspaden in dezelfde richting in dezelfde wijk. Tegelijkertijd tonen de resultaten dat straten met een fietspad gemiddeld 2.3% hogere verkeersvolumes hebben.

In hoofdstuk 6 wordt onderzocht in hoeverre smartphones een rol spelen bij het aantal verkeersongevallen. De studie maakt gebruik van variatie in telefoongebruikstarieven in Nederland als gevolg van de roaming-regelgeving van de Europese Unie (EU) in 2017. Hierbij zijn alle roaming-toeslagen voor EU-inwoners zijn afgeschaft. Deze verandering wordt gebruikt om een *difference-in-differences* model te schatten waarbij niet-Nederlandse automobilisten uit de EU worden vergeleken met een controle groep met Nederlandse automobilisten. De resultaten suggereren dat 10% van de ongevallen kan worden verklaard door het gebruik van smartphones. De bevindingen wijzen er verder op dat afleiding door smartphones zowel ernstige als lichte ongevallen in een vergelijkbare mate laat toenemen. Daarnaast blijkt dat telefoongerelateerde ongevallen vooral gebeuren op lokale wegen binnen de bebouwde kom.

Naast de kwantitatieve resultaten van elk hoofdstuk geeft dit proefschrift voorbeelden van natuurlijke experimenten in situaties waar gecontroleerde experimenten niet haalbaar zijn door (onder meer) ethische bezwaren. Het zou bijvoorbeeld onethisch zijn om een experiment uit te voeren waarbij fietsers of schakers door een onderzoeker worden blootgesteld aan luchtvervuiling. Natuurlijke experimenten zijn vrij van ethische bezwaren en bieden daarnaast veel statistische kracht.

De gedragseffecten van luchtvervuiling zijn vaak subtiel. Dat betekent dat zelfs als de samenleving (laboratorium)experimenten met luchtvervuiling aanvaardbaar zou achten, de subtiliteit van het effect van luchtvervuiling een grote onderzoekspopulatie vereist. Het uitvoeren van dergelijke grootschalige experimenten zou zeer kostbaar zijn. Natuurlijke experimenten zijn dus een aantrekkelijke methode om inzicht te krijgen in de gevolgen van luchtvervuiling.

Voor het bestuderen van de verkeersveiligheid gelden soortgelijke ethische bezwaren, die het moeilijk maken om (veld)experimenten uit te voeren. Vooral in het geval van smartphone-afleiding van bestuurders zijn gecontroleerde experimenten niet haalbaar. In dit geval zouden onderzoekers de proefpersonen niet alleen in gevaar brengen, maar ook een illegale activiteit laten uitvoeren. Tot nu toe maakten de meeste onderzoekers in dit domein daarom gebruik van simulaties om de effecten van afleiding te bestuderen. Met simulatiebenaderingen is het echter lastig de totaliteit van het probleem in kaart brengen.

Ten slotte vormt nieuw ingevoerd beleid ook vaak een goede basis voor een natuurlijk experiment. De aanleg van fietspaden in New York City is een duidelijk voorbeeld van zo'n beleid. In dit geval kunnen uit de analyse van waargenomen verkeersgegevens niet alleen causale effecten worden afgeleid, maar kan ook het beleid zelf worden geëvalueerd.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

728 I. SAKALAUSKAITE, *Essays on Malpractice in Finance*

729 M.M. GARDBERG, *Financial Integration and Global Imbalances.*

730 U. THŰMMEL, *Of Machines and Men: Optimal Redistributive Policies under Technological Change*

731 B.J.L. KEIJSERS, *Essays in Applied Time Series Analysis*

732 G. CIMINELLI, *Essays on Macroeconomic Policies after the Crisis*

733 Z.M. LI, *Econometric Analysis of High-frequency Market Microstructure*

734 C.M. OOSTERVEEN, *Education Design Matters*

735 S.C. BARENDSE, *In and Outside the Tails: Making and Evaluating Forecasts*

736 S. SÓVÁGÓ, *Where to Go Next? Essays on the Economics of School Choice*

737 M. HENNEQUIN, *Expectations and Bubbles in Asset Market Experiments*

738 M.W. ADLER, *The Economics of Roads: Congestion, Public Transit and Accident Management*

739 R.J. DÖTTLING, *Essays in Financial Economics*

740 E.S. ZWIERS, *About Family and Fate: Childhood Circumstances and Human Capital Formation*

741 Y.M. KUTLUAY, *The Value of (Avoiding) Malaria*

742 A. BOROWSKA, *Methods for Accurate and Efficient Bayesian Analysis of Time Series*

743 B. HU, *The Amazon Business Model, the Platform Economy and Executive Compensation: Three Essays in Search Theory*

744 R.C. SPERNA WEILAND, *Essays on Macro-Financial Risks*

745 P.M. GOLEC, *Essays in Financial Economics*

746 M.N. SOUVERIJN, *Incentives at work*

747 M.H. COVENEY, *Modern Imperatives: Essays on Education and Health Policy*

748 P. VAN BRUGGEN, *On Measuring Preferences*

749 M.H.C. NIENTKER, *On the Stability of Stochastic Dynamic Systems and their use in Econometrics*

750 S. GARCIA MANDICÓ, *Social Insurance, Labor Supply and Intra-Household Spillovers*

751 Y. SUN, *Consumer Search and Quality*

752 I. KERKEMEZOS, *On the Dynamics of (Anti) Competitive Behaviour in the Airline Industry*

753 G.W. GOY, *Modern Challenges to Monetary Policy*

754 A.C. VAN VLODROP, *Essays on Modeling Time-Varying Parameters*

755 J. SUN, *Tell Me How To Vote, Understanding the Role of Media in Modern Elections*

756 J.H. THIEL, *Competition, Dynamic Pricing and Advice in Frictional Markets: Theory and Evidence from the Dutch Market for Mortgages*

757 A. NEGRIU, *On the Economics of Institutions and Technology: a Computational Approach*

758 F. GRESNIGT, *Identifying and Predicting Financial Earth Quakes using Hawkes Processes*

759 A. EMIRMAHMUTOGLU, *Misperceptions of Uncertainty and Their Applications to Prevention*

760 A. RUSU, *Essays in Public Economics*

761 M.A. COTOFAN, *Essays in Applied Microeconomics: Non-Monetary Incentives, Skill Formation, and Work Preferences*

762 B.P.J. ANDRÉE, *Theory and Application of Dynamic Spatial Time Series Models*

763 P. PELZL, *Macro Questions, Micro Data: The Effects of External Shocks on Firms*

764 D.M. KUNST, *Essays on Technological Change, Skill Premia and Development*

765 A.J. HUMMEL, *Tax Policy in Imperfect Labor Markets*

766 T. KLEIN, *Essays in Competition Economics*

767 M. VIGH, *Climbing the Socioeconomic Ladder: Essays on Sanitation and Schooling*

768 YAN XU, *Eliciting Preferences and Private Information: Tell Me What You Like and What You Think*

769 S. RELLSTAB, *Balancing Paid Work and Unpaid Care over the Life-Cycle*

770 Z. DENG, *Empirical Studies in Health and Development Economics*

771 L. KONG, *Identification Robust Testing in Linear Factor Models*

772 I. NEAMŢU, *Unintended Consequences of Post-Crisis Banking Reforms*

773 B. KLEIN TEESELINK, *From Mice to Men: Field Studies in Behavioral Economics*

774 B. TEREICK, *Making Crowds Wiser: The Role of Incentives, Individual Biases, and Improved Aggregation*

775 A. CASTELEIN, *Models for Individual Responses*

776 D. KOLESNYK, *Consumer Disclosures on Social Media Platforms: A Global Investigation*

777 M.A. ROLA-JANICKA, *Essays on Financial Instability and Political Economy of Regulation*